Fall 2019

College Campus Sexual Assault and Retention Rates

Abigail R. Hauer
Gettysburg College

College Campus Sexual Assault and Retention Rates

Abstract

Increased media attention on college crime, specifically sexual assault, has led to greater prioritization of campus safety when deciding whether to continue attending a college. This, coupled with society's view of a four-year college education as a necessity to succeed in the labor market, creates a potential tradeoff between safety on campus and future employment success. To analyze such tradeoff, I use data from the US Department of Education from 2014 to 2017 to examine whether college campus sexual assault at four-year American institutions impacts retention rates. Such results have implications for college policies to combat sexual assault on campus not only to keep students safe, but to prevent students from transferring or dropping out which could curb institutional money flow. Using an OLS model that addresses typical difficulties associated with time series work, I find that college campus sexual assault decreases retention rates at a statistically significant level, implying that college students value their safety at school more than any potential change in their future job market success due to transferring or dropping out.

Keywords

Econometrics, sexual assault, retention rates, college crime, Clery Act

Disciplines

Educational Assessment, Evaluation, and Research | Feminist, Gender, and Sexuality Studies | Higher Education

Comments

Written for ECON 352: Advanced Econometrics.

College Campus Sexual Assault and Retention Rates

Abigail Hauer

Gettysburg College

haueab01@gettysburg.edu

December 6, 2019

Increased media attention on college crime, specifically sexual assault, has led to greater prioritization of campus safety when deciding whether to continue attending a college. This, coupled with society's view of a four-year college education as a necessity to succeed in the labor market, creates a potential tradeoff between safety on campus and future employment success. To analyze such tradeoff, I use data from the US Department of Education from 2014 to 2017 to examine whether college campus sexual assault at four-year American institutions impacts retention rates. Such results have implications for college policies to combat sexual assault on campus not only to keep students safe, but to prevent students from transferring or dropping out which could curb institutional money flow. Using an OLS model that addresses typical difficulties associated with time series work, I find that college campus sexual assault decreases retention rates at a statistically significant level, implying that college students value their safety at school more than any potential change in their future job market success due to transferring or dropping out.

I would like to thank those who gave helpful comments after my in-class presentation, as well as Professor Cadigan for his continuous help throughout the semester.

Section 1: Introduction

The Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics Act, known as the Clery Act, requires colleges receiving federal funding to report campus crime statistics to the US Department of Education (DOE) each year (Gregory and Janosik 2002). Since the implementation of Clery Act by Congress in 1990, over 78 million college students reported criminal victimization (National Crime Victims' Rights Week 2017). Of the 78 million students reporting, "a few violent campus incidents highlighted by the media have drawn a spotlight to college and university campuses, [creating] the impression that campuses are increasingly dangerous places" (Fisher 1995).

Such campus crimes garnering media attention tend to be sexual assault crimes, such as the 2016 Brock Turner rape case at Stanford University that "[reverberated] across the country" and forged an ongoing "watershed moment in public perception of campus sexual assaults" (Fuller 2016). While such sexual assault cases promote media portrayal of college campuses as increasingly unsafe, especially for certain populations more susceptible to sexual assault, society progressively views college education as a necessity to succeed in the American labor market (Morrall et al. 2010, Fisher et al. 2000). Thus, campus safety and the societal expectations for college degrees force students to decide whether to continue attending a relatively unsafe school, transfer to a relatively safer school, or drop out entirely. This paper will investigate the relationship between campus sexual assault crimes and retention rates at four-year colleges and universities in the United States. I hypothesize that predicted retention rates significantly decline as campus sexual assaults increase. Although students' forecasted success in the labor market may fall by transferring or dropping out of a relatively unsafe school, I hypothesize that students value their

safety over relatively insubstantial changes in such predicted labor market success and thus choose to attend a relatively safer school with less campus sexual assault.

In the next section, I will describe existing literature on college selection, college crime, sexual assault crime, and retention rates to determine control variables that I include in my model; my paper advances the econometric literature on higher education by being the first to look specifically at the relationship between one primary independent variable and dependent variable—campus sexual assault campus and retention rates. In Section 3, I will develop my model to accurately analyze the relationship between campus sexual assault and retention rates. In Section 4, I will discuss the institutional level data used to test my hypothesis, specifically DOE College Scorecard data on academia and Campus Safety and Security data on crime. In Section 5, I will use such data to test my hypothesis on campus sexual assault decreasing predicted retention rates. In Section 6, I will make concluding remarks.

Section 2: Literature Review

Since certain components that contribute to an individual's decision to begin attending a college also likely affect their decision to *continue* attending a college, literature on college selection supplies control variables to include in my models. To discover what college characteristics students look for when making decisions about applying to or attending a specific school, Pampaloni (2010) uses open-ended and Likert-scale survey questions with predominantly New Jersey high school seniors at college open houses. By calculating descriptive statistics for survey responses, as well as Pearson correlation coefficients for the responses to the different method of survey questions, Pampaloni (2010) finds that students prioritize academic programs, location, and cost of institutions when selecting a college to attend.

Considering that parents also influence a student's college selection decision, Warwick and Mansfield (2003) conduct a Likert-scale survey with high school seniors and some of their parents at eight private, religiously affiliated schools to determine differences between students and parents regarding perceived risk factors in college selection.[1] Using descriptive statistics and Chi-Square tests, Warwick and Mansfield (2003) find that parents place a statistically significantly higher level of importance on institutional security and safety (i.e. campus crime).

Using more comprehensive survey data of over 9,000 students at two and four-year and public and private colleges, Janosik and Gehring (2003) report that 78 percent of students do not know about the Clery Act and its crime report, and only 8 percent of students are influenced by Clery Act report data when selecting a college to attend. Although parents place concern on campus crime during college selection, students do not substantially consider campus crime when selecting a school to apply to or attend.

While college selection literature provides factors I control for in my models,[2] the studies do not use representative surveys or robust econometric methods, rather just descriptive statistics, correlation coefficients, and Chi-Square tests. However, literature specifying campus crime as a dependent variable uses combinations of survey and national data, as well as econometric techniques, to determine the effect of institutional and community factors on campus crime.

Using Uniform Crime Report (UCR) data and survey data sent to college research offices and campus police forces, Morriss (1993) uses a regression model and stepwise regression to discover which factors account for the greatest variance in the predicted value of campus crime.

[1] The factors include financial, physical, functional, social, and psychological components.
[2] I control for average cost of attendance to rid my primary independent variable (campus sexual assault crimes) of any potential bias.

Although wealthier campuses can provide more deterrents,[3] Morriss (1993) finds that campus wealth explains "more than 10 percent of the variance in campus crime rates," overall determining that wealthier campuses offer more opportunities and targets for individuals to commit crime.[4]

Also seeking to determine which factors account for the greatest variance in the predicted value of campus crime, Volkwein et al (1995) uses 23 independent variables from numerous national databases to determine the relationship between campus crime and campus and student characteristics.[5] Using hierarchical and stepwise regressions, Volkwein et al (1995) finds that student characteristics explain the greatest amount of variance in predicted campus violent crime,[6] and that the 23 independent variables explain 79 percent of the variance in predicted campus property crime.

Similarly using numerous independent variables, McPheters (1978) uses national and survey data to test whether crime on college campuses has a relationship with student body, institutional, and neighboring community. Using two-stage least squares regressions, McPheters (1978) finds that the percentage of students living on campus in dormitories, the unemployment rate in the nearest city, and security expenditures significantly increase the predicted value of the number of campus crimes per 10,000 students.

Overall, while the literature on campus crime provides more control variables for my models,[7] studies do not focus on analyzing the relationship between one specific independent

[3] Wealth is indicated by tuition cost, ratio of total university operating expenditures to campus population, percentage of campus applicants not admitted, percentage of faculty holding tenure, and ratio of students to faculty.
[4] I control for average cost of attendance as a proxy for wealth to rid my primary independent variable of any potential bias.
[5] Volkwein et al (1995) uses the UCR and Consortium for Higher Education Campus Crime Research (CHECCR) for crime data, the Integrated Post-Secondary Education Database System (IPEDS) and College Board Survey for campus and student body characteristics, and the FBI and US Census for community data.
[6] The specific student characteristics include percentages of applicants accepted, percentage of students receiving financial aid, percentage of males, percentage of African Americans, percentage of foreign students, and percentage of students living in residence halls.
[7] I control for average cost of attendance, percentage of female students, and percentage of African American students.

variable and campus crime, rather they focus on determining a specific set of independent variables to explain the variance in predicted campus crime. Instead of specifying overall campus crime as a dependent variable, some literature specifically designates sexual assault crimes as the dependent variable. As Fisher (1990) uses survey data of nearly 4,000 undergraduate women and finds that there are approximately 27.7 rapes per 1,000 female students, more studies examine the gender and alcohol dynamic of sexual assault crimes on college campuses.

Using survey data from sociology and psychology classes at a large public university in the Northeast, Menard et al (2003) finds via OLS and tobit regressions that men are two times as likely to be sexually harassing and 3.5 times more likely to be sexually coercive than women. Menard et al (2003) also finds for both men and women that increased alcohol expectancy increases the likelihood of sexual harassment, and that "those who are sexually harassing or coercive may use alcohol expectancies to cognitively restructure the situation to one that is favorable to sexual aggression."

Regarding alcohol use, Grossman (1999) finds via a Bureau of Justice Statistics survey of prison inmates that alcohol is more prevalent in violent crimes, such as sexual assault, than in other crimes.[8] Consistent with Grossman's (1999) findings, Ullman (1999) finds via hierarchical regressions using a national sample of over 3,000 college women, that "victim and offender alcohol use at the time of the [sexual] assault were related to greater sexual victimization severity." Regardless of the influence of alcohol,[9] women as victims and non-victims are most likely to

[8] Grossman goes into greater detail as to why alcohol is more prevalent in violent crimes, citing three main reasons: (1) alcohol can alter one's behavior by increasing excitability and/or boosting courage, (2) "people misinterpret social cues with the result of violent reactions," and (3) "drunkenness may give people an excuse for aberrant behavior, despite whether or not pharmacological effects exist."

[9] Although women may change their behavior in response to sexual assault crimes regardless of the influence of alcohol, I control for liquor law violations as a proxy for alcohol abuse to rid my primary independent variable (sexual assault crimes) of any potential bias.

change their behavior to avoid unsafe practices susceptible to crime, including sexual assault (Morrall et al, 2010).

One way students may avoid unsafe behavior on campus is by dropping out or transferring from the school where one was sexually assaulted, affecting retention rates. Using a convenience survey distributed to nearly 2,000 students at an upper Midwest university, Elliott and Healy (2001) report that students are least satisfied with campus safety and security regarding college recruitment and retention. While Elliott and Healy's (2001) study does not conduct econometric techniques, other literature on retention rates does not include campus crime as an independent variable. Using data from various college guides of different years in the 1980s, Marcus (1989) conducts three-stage least squares regressions to explain the variation in private colleges' predicted retention rates, but also private colleges' average SAT score and the percentage of students that enroll.[10] Similarly, Marsh (2014) uses hierarchical multiple regressions to explain the variation in public colleges' first-year retention rate, specifically looking at student input, bridge, institutional structural, institutional financial, and faculty interaction factors.[11] From the two studies' statistically significant results, I add controls for average SAT score, percentage of female students, and percentage of fulltime students to my models.[12]

Although containing poor methodology with nonrepresentative surveys, the literature on college selection, college crime, sexual assault crime, and retention rates supplies control variables that I include in my model. However, the literature does not specifically aim to analyze the

[10] Using retention rate and the natural log of retention rate, Marcus (1989) uses the percentage of students accepted, average SAT score, percentage of freshmen who enroll, percentage of freshmen from the top 10 percent of their high school, percentage of students from out-of-state, student to faculty ratio, percentage of females, number of majors, tuition, average need-based freshmen scholarships, percentage of students living on campus, percentage of black students, and public tuition at leading university in state.

[11] See Marsh (2014) for a description on the specific variables that make up each category.

[12] The first three controls (average SAT score, percentage of female students, and admission rate) were statistically significant at the 5 or 1 percent level in Marcus (1989). The third and fourth controls (admission rate and percentage of fulltime students) were statistically significant at the 1 percent level in Marsh (2014).

relationship between one main independent and dependent variable—for example, literature on college crime and retention rates includes many independent variables to best explain the variation in the dependent variable. Thus, my paper would advance the econometric literature on higher education by being the first to look specifically at the relationship between one primary independent variable and dependent variable—sexual assault campus crimes and retention rates.

Section 3: Modeling

To investigate the relationship between sexual assault crimes and retention rates at four-year colleges and universities in the United States, I need data on both sexual assault crimes and retention rates at numerous four-year institutions over a substantial time period. I also need data on the controls suggested in the literature, specifically average cost of attendance, share of African American students, share of female students, share of fulltime students, average SAT score, and liquor law violations. As these variables likely correlate with campus sexual assault and affect retention rates, excluding them may lead to endogeneity in the error term of regressions. In order to avoid biasing coefficient estimates because of violating the error term zero conditional mean OLS assumption, I need to include such controls in my regressions.

Aside from those provided in the literature, I also desire data on controls on crime variables, specifically campus crime overall and public property crime. Campus crime overall[13] may correlate with campus sexual assault as crimes such as assault may precede sexual assaults. Campus crime overall likely impacts retention rates as such crimes may also function as a proxy for campus safety which students value in a college or university.[14] Public property crime accounts

[13] Crime consists of murders, negligent manslaughters, rapes, fondling instances, statutory rapes, robberies, aggravated assaults, burglaries, motor vehicle thefts, and arsons.

[14] Per Elliott and Healy's (2001) results in which students are least satisfied with campus safety and security regarding college retention, very low satisfaction with safety may drive a student to drop out or transfer from their school, thus decreasing retention.

for crime "immediately adjacent to, within, or surrounding on-campus geography," such as public streets and public sidewalks near campus (Clery, 1990). While public property crimes in and of themselves may also correlate with campus sexual assault as such property crimes may occur before a sexual assault,[15] property crimes also function as a proxy for the influence of crime in the near off-campus area that may percolate onto campus, which also may correlate with campus sexual assault.[16] Whether on its own merit or as a proxy, public property crime also likely impacts retention rates as such crimes may deter students from staying at a school due to fear of one's safety when traveling to areas adjacent to campus. Again to avoid biasing coefficient estimates because of violating the error term zero conditional mean OLS assumption, I must include campus crime overall and public property crime to rid the error term of endogeneity.

Per the desired independent and dependent variables, as well as controls derived from the literature as well as additional crime controls, I develop the following model to effectively analyze whether campus sexual assault significantly decreases retention rates:

$$\log(retention\ rate)_{it} = \beta_{0\,it} + \beta_1 sexual\ assault_{it} + \theta X_{it} + \gamma Z_{it} + \varepsilon_{it}$$

In such model, the individual subscript is at the institutional level and the time subscript is per year. B_0 is the intercept of the model, depicting what the retention rate would be if all subsequent variables were 0.[17] β_1 is the coefficient estimate for campus sexual assault— the sign, magnitude, and statistical significance in difference from 0 will be analyzed to determine if campus sexual assault impacts retention rates. Retention rate is a log variable to better analyze how increased campus sexual assault impacts retention— now, the coefficients are interpreted as how a 1 unit

[15] For instance, if a student is assaulted on a public street and then taken into a campus building to be sexually assaulted.
[16] As McPheters (1978) finds that the unemployment rate in the nearest city increases the predicted value of the number of campus crimes per 10,000 students, I predict that public property crime as a proxy for crime in the area adjacent to campus may also increase the number of campus crimes, including campus sexual assault.
[17] The intercept of the model is nonsensical and is not necessary to analyze for the premise of the question at hand.

increase in an independent variable impacts predicted retention rates by (100*coefficient) percentage points instead of percent. θX and γZ are the controls from literature and outside literature, respectively.

$$\theta X_{it} = share\ black_{it} + share\ female_{it} + share\ fulltime_{it} + \log(average\ cost)_{it}$$
$$+ \log(average\ SAT)_{it} + alcohol\ violation_{it}$$
$$\gamma Z_{it} = campus\ crime_{it} + public\ property\ crime_{it}$$

I include average SAT and average cost as log variables because the two variables are better analyzed as percentage changes instead of level changes.[18] I also demean each crime variable as based upon the share of fulltime students; for example, the coefficient on campus sexual assault would represent how an increase in 1 campus sexual assault *per fulltime student* impacts predicted retention rates.[19]

To prevent any unobservable characteristics that do not vary over time from biasing the coefficient estimates, I also include institution-level fixed effects. Such fixed effects may include the culture/atmosphere of the institution and the share of Greek life on campus.[20]

$$\log(retention\ rate)_{it} = \beta_{0\,it} + \beta_1 sexual\ assault_{it} + \theta X_{it} + \gamma Z_{it} + \alpha_i + \varepsilon_{it}$$

I also difference and detrend each variable, as well as include a year variable, to combat any potential functional form misspecification, spurious correlation, and/or unit root behavior.

$$\Delta\log(retention\ rate)_{it} = \beta_{0\,it} + \beta_1 \Delta sexual\ assault_{it} + \theta \Delta X_{it} + \gamma \Delta Z_{it} + \alpha_i + year + \varepsilon_{it}$$

[18] A better interpretation of either coefficient would be how a 1 percent increase impacts retention rates instead of a 1-point increase on the SAT or a 1 dollar increase in cost, as such change is miniscule.

[19] I detail each crime variable (campus sexual assault, campus crime overall, public policy crime, and campus alcohol violation) as per fulltime student so as account for the size of the fulltime student body, allowing me to use colleges and universities of varying sizes.

[20] I did not discover any literature on whether the share of Greek life on campus varies over time, nor any data on Greek life at each institution. Per my knowledge of Gettysburg College Greek life and its relatively consistent share of the student population, I make the assumption that Greek life is an unobservable characteristic that varies across institution but not over time, thus it is included in the fixed effect.

Additionally, I cluster based upon each institution to counter any potential serial correlation, per suggestion from Wooldridge (2011). I also use heteroskedasticity-robust standard errors to prevent any potential heteroskedasticity in my model.

Although I have made numerous attempts to mitigate any potential biasing of the coefficient estimates due to endogeneity in the error term, omitted variable bias may still exist in my model. Specifically, I do not have data on the share of Greek students on campus, nor if such Greeks have access to Greek-only housing on or near campus (i.e. fraternity and/or sorority housing). Although the share of Greek students may be capsulated within the fixed effects if the share does not vary over time, I still desire specific data for each institution as I hypothesize that the share of Greek life correlates with campus sexual assault as Greek life supplies unsafe scenarios in which students are more susceptible to being sexually assaulted.[21] Greek life also likely impacts retention rates, thus its exclusion from the model may create endogeneity in the error term and bias the coefficient estimates of the model.[22] There may be other variables similar to Greek life that are not directly controlled for within my model, thus potentially biasing my coefficient estimates; however, I attempt to minimize such bias by including other controls to limit endogeneity in the error term.

Aside from potential omitted variable bias, there may also be measurement error within my model. The crime statistics within my model only account for *reported* crimes; those who report crimes to non-mandatory reporters are not accounted for within my model.[23] Detailed more in the following section, my combination of one dataset that reports statistics for the calendar year

[21] Greek life parties, where alcohol is typically served and consumed in copiously amounts, may feature more sexual assault as alcohol tends to be prevalent in violent crimes, such as sexual assault per Grossman (1999).
[22] Greek life may lead to higher retention rates as students involved in it may develop close friends, enticing them to stay at their institution. On the contrary, Greek life may also decrease retention rates as those involved may have less time to dedicate to their studies, making them potentially more likely to fail and thus drop out.
[23] Although mandatory reporting rules vary per campus, counselors and religious leaders such as chaplains tend to not be mandatory reporters, so any crimes reported to the two may not be accounted for in Clery Act data.

and another dataset that reports for the academic year may introduce measurement error. However, since I consistently merge each academic and calendar year from 2014 to 2017/2018, it is unlikely that the difference in time periods for the two datasets causes an overestimate or underestimate the effect of college crime on retention rates.

Overall, my final model substantially attempts to assure that the coefficient estimates are the best linear unbiased estimators, per the Gauss Markov Theorem.

Section 4: Data

I utilize data from the DOE College Scorecard and Campus Safety and Security Data Analysis Cutting Tool. The College Scorecard data includes panel data on many variables for each institution in the US regarding academia. Such variables are collected after each academic year and provided through federal reporting from institutions, specifically institutions that receive federal financial aid dollars. The Campus Safety and Security Data includes panel data on college crime for each institution in the US broken down by campus, collected annually and provided through federal reporting from institutions per the Clery Act that requires colleges that receive federal funding to report college crime statistics to the US DOE each year (McCallion 2014).

For this paper, I utilize College Scorecard data from the 2014-2015 academic year to the 2017-2018 academic year, and Campus Safety and Security Data from 2014 to 2017.[24] Since the collection period for the two datasets differs, I merge the 2014-2015, 2015-2016, 2016-2017, and 2017-2018 College Scorecard academic year data with the 2014, 2015, 2016, and 2017 Campus Safety and Security Data annual data, respectively.

From the College Scorecard dataset, I choose variables that allow me to use data for only the main campuses of four-year institutions, eliminating other campuses of each institution such

[24] I select this time period as the Campus Safety and Security Dataset did not include statistics on rape, fondling, or statutory rape (the variables defined as sexual assault) in datasets before 2014.

as abroad campuses, and non-four-year institutions such as community colleges and technical schools. Of the 2,146 variables, I choose 6 variables to include in my future regressions based upon the variables used in previous literature's regressions.[25]

From the Campus Safety and Security Dataset, I also eliminate non-four-year institutions and non-main campuses, including the sector of the institution labeling each school as a public or private two-year or four-year institution. I choose variables to measure crime and discipline on each campus,[26] broken down by campus, non-campus, and public property.[27] I create a variable to represent total crime on campus and non-campus, as well as total crime on public property. I also develop a variable to represent total alcohol violations on campus and non-campus.

I then merge the College Scorecard and Campus Safety and Security Data Analysis set by using each institutions' unit ID and the corresponding year for each data point. The merged dataset contains 6,455 observations, each observation a college-year pair with some institutions not reporting crime statistics in the Campus Safety and Security Dataset and/or institutional statistics in the College Scorecard dataset for each year in the three-year period.[28] Since I want to use observations that report *all* necessary data for each year in the three-year time period, I limit my dataset to only 1,143 observations— 286 institutions that report all data for 2014, 2015, 2016, and 2017. Sample selection bias may arise due to the schools not reporting all necessary data each

[25] These variables include retention rates, share of African American students, share of female students, share of fulltime students, average cost of attendance, and average SAT score.

[26] Crime consists of murders, negligent manslaughters, rapes, fondling instances, statutory rapes, robberies, aggravated assaults, burglaries, motor vehicle thefts, and arsons.

[27] On campus, as defined by the Clery Act, is an institution's core, main campus—its residence halls, academic buildings, cafeteria, buildings frequented by students, etc. Non-campus, as defined by the Clery Act, is non-campus properties that are not contiguous to the core campus but are used by students for the educational purposes of the institution, including properties owned or controlled by a student organization officially recognized by the campus. Public property, as defined by the Clery Act, is the public property that immediately borders the core campus, including public sidewalks, streets, etc.

[28] Some data was excluded or "null" in both datasets, while other institutional data was privacy-suppressed in the College Scorecard Dataset.

year and my subsequent removal of those schools from my forthcoming regressions; while there is little I can do to limit this bias, a potential future research study may be examining the relationship between non-reporting of college crime and/or institutional data and previous years' statistics.[29]

I specifically focus on 11 variables in my regressions supported by previous literature: retention rate, campus sexual assault, share of African American students, share of female students, share of fulltime students, average cost of attendance, average SAT score, public property crime, campus (both on campus and non-campus) liquor law discipline, total campus crime (again both campus and non-campus), and year. Table 1 in the Appendix provides summary statistics on these variables.

Although my model implements measures to best address such potential issues, I test and correct for model misspecification, heteroskedasticity, multicollinearity, spurious correlation, unit root behavior, and serial correlation in my forthcoming regressions to provide unbiased and efficient results.

Section 5: Evidence

Table 2 provides the entirety of the results from my model.

Csexualassault represents that when campus sexual assault increases by one crime per fulltime student annually, the predicted value of retention rate decreases by approximately 0.04566 percentage points, holding constant the influence of other included variables. Since the p-value on *csexualassault* (0.017) is statistically significant at the 5% level, I can reject the null hypothesis that the *csexualassault* coefficient is 0. In other words, the *csexualassault* coefficient estimate is statistically significant in difference from 0. These results align with my hypothesis: higher

[29] Non-reporting may potentially be indicative of an institution that reported data that poorly represented the institution in years past.

campus sexual assault per fulltime student inclines students to drop out or transfer in response, thus decreasing retention rates. However, the magnitude of the coefficient is not relatively large, implying it would take a substantial increase in the number of campus sexual assaults per student to decrease retention rates at a meaningful level. Such relatively small coefficient may be due to measurement error in the campus sexual assault variable—there are likely more sexual assaults occurring on college campuses than what is reported in the Campus Safety and Security dataset. Thus, the negative relationship between campus sexual assault per fulltime student and retention rates is likely underestimated due to measurement error.

Although the campus sexual assault result, as well as the result of other controls, aligns with my hypothesis and subsequent theory and literature used to develop such hypothesis, other variables' coefficients do not align with such theory and literature. Specifically, the positive and statistically significant campus alcohol violation and overall campus crime coefficients digress from the theory.

Calcoholvio represents that when campus alcohol violations increase by one violation per fulltime student annually, the predicted value of retention rate *increases* by approximately 0.00329 percentage points, holding constant the influence of other included variables. Since the p-value on *calcoholvio* (0.001) is statistically significant at the 1% level, I can reject the null hypothesis that the *calcoholvio* coefficient is 0. In other words, the *calcoholvio* coefficient estimate is statistically significant in difference from 0. These results imply that increasing campus alcohol violations per fulltime student *increase* retention rates, which does not align with the theory and literature because alcohol violations likely means alcohol is prevalent on campus and such alcohol consumption increases the prevalence of sexual assault due to altering one's behavior and cohesiveness, per Grossman (1999). Aside from its implications on sexual assault, alcohol

violations also likely decrease retention rates because egregious alcohol consumption means students are likely not dedicating ample time to their studies, meaning they have a higher chance of failing out and thus diminishing retention rates. However, campus alcohol violation may also function as a proxy for the social scene at a college—a campus with more alcohol violations likely has more parties and thus a better social scene, which may increase or decrease retention rates.[30] Overall, the positive campus alcohol violation coefficient does not align with theory and literature.

Ccampuscrime represents that when campus crime increases by one crime per fulltime student annually, the predicted value of retention rate *increases* by approximately 0.0430 percentage points, holding constant the influence of other included variables. Since the p-value on *ccampuscrime* (0.002) is statistically significant at the 1% level, I can reject the null hypothesis that the *ccampuscrime* coefficient is 0. In other words, the *ccampuscrime* coefficient estimate is statistically significant in difference from 0. These results imply that increasing campus crime per fulltime student *increases* retention rates, which does not align with the theory and literature because campus crime functions as a proxy for overall safety and security which students, which students are least satisfied with per Elliott and Healy (2001), thus implying that students are likely to drop out or transfer in response to poor safety (i.e. high campus crime).

Overall, although nonsensical in terms of sign, the magnitude of the two coefficient estimates was relatively small. To further assess the variables, I conduct a joint hypothesis to determine if campus alcohol violation and campus crime jointly impact retention, to which I find that the two *are* jointly significant (see Table 3 in the Appendix). The positive and statistically

[30] Allen et al (2008) finds that social connectedness in college increases the likelihood of a student staying at their school as opposed to dropping out, but that social connectedness also decreases first-year academic performance which may decrease the likelihood of staying versus dropping out. Thus, "the positive direct effect of social connectedness on staying versus dropping is negated by its negative (suppressed) effect on first-year academic performance," meaning that social connectedness may increase or decrease retention rates.

significant campus alcohol violation and campus crime coefficients, as well as their joint significance, may be driven by endogeneity in the error term because of omitted variables. For example, campus alcohol violation and campus crime are likely correlated with the share of Greek life students on campus, as increasing the share of Greek life students likely increases campus alcohol violations and campus crime since Greek life tends to revolve around alcohol consumption, and such alcohol may contribute to crime because of diminished mental capacity. The nonsensical variable coefficients may also be driven by measurement error, as campus alcohol violations and campus crime may go unreported because of some Greek life events being held off campus and thus being less likely to be "busted" by campus security forces, and campus crime may also go unreported because of individuals reporting such crimes to a confidential source. Since issues typical with time series work are addressed within my model,[31] it is likely that either omitted variable bias, measurement error, or a combination of both are biasing the campus alcohol violation and campus crime coefficients, and potentially the entire model.

Aside from the positive and statistically significant campus alcohol violation and campus crime coefficient estimates, all other variables' coefficient estimates aligned with theory in terms of sign and mainly significance, with only three variables not being statistically significant in difference from 0: the share of female students, the share of fulltime students, and the amount of public property crime. However, the three variables' signs aligned with theory, with the former two having a positive sign, indicating that increasing the share of female and fulltime students increases predicted retention rates, and the latter having a negative sign, indicating that increasing public property crime per fulltime student decreases predicted retention, although not at a

[31] I include robust techniques to test for and address (when necessary) spurious correlation, serial correlation, unit root behavior, multicollinearity, model misspecification, and heteroskedasticity. However, I was not able to fully test for unit root behavior and cointegration because the Dickey-Fuller test could not be conducted with my model due to "repeated time values in sample," per Stata.

statistically significant level. To summarize the additional variables' coefficient estimates, increasing the share of African American students decreases predicted retention rates, and increasing the average cost and average SAT score increases predicted retention rates.

Overall, while there may be endogeneity in the error term and/or measurement error potentially biasing the coefficient estimates and resulting in the violation of the Gauss Markov Theorem, my model supports my hypothesis that increasing campus sexual assault results in students transferring and dropping out in response, decreasing retention rates by approximately 0.4566 percentage points for every additional campus sexual assault per fulltime student.

Section 6: Conclusion

After addressing and correcting for spurious correlation, serial correlation, unit root behavior, model misspecification, multicollinearity, and heteroskedasticity when applicable (see Tables 4 through 9 in the Appendix for tests regarding model misspecification, heteroskedasticity, and multicollinearity), I arrive at a model with numerous controls to best address the relationship between campus sexual assault and retention rates. My results support my hypothesis that increasing campus sexual assaults per fulltime student results in students transferring or dropping out in response, thus decreasing retention rates. Specifically, I find that increasing campus sexual assault per fulltime student significantly decreases predicted retention rate by approximately 0.04566 percentage points. This would imply that 10 additional campus sexual assaults per fulltime student would decreases retention rates by nearly half a percentage point, and that reduction could result in a decline in future donations and applications to the institution. However, the magnitude of the coefficient is not relatively large, implying it would take a substantial increase in the number of campus sexual assaults per student to decrease retention rates at a meaningful

level. Overall, the data supports my hypothesis that students value their safety at school more than any potential change in their future job market success due to transferring or dropping out.

As the results suggest that college students respond to campus sexual assaults by transferring or dropping out of their unsafe school, institutions are more likely to address campus sexual assault. Awareness of how increasing campus sexual per fulltime student assault decreases retention will likely incite colleges to take more aggressive measures to combat such sexual assaults so that institutions will not lose money because of less donations and/or applications (and thus less students and funds from their tuition) due to institutions' lower retention rates. While institutions care about the safety of their students from campus sexual assault, institutions also care about their money flow and thus are more likely to respond to such sexual assault if it is likely to impact their retention rate and subsequently their funding.

Future literature regarding campus sexual assault and retention rate may step back and analyze the relationship between campus crime overall and retention rates, as my results indicated that increasing campus crime *increases* retention rates, which seems to contradict my results regarding campus sexual assault. Future literature may also examine the relationship between non-reporting of college crime and/or institutional data and previous years' statistics, as numerous institutions did not report crime statistics for each year of my panel. Some schools also reported no campus crime or campus sexual assault, a statistic I find hard to believe.

Works Cited

Allen, Jeff, Steven B. Robbins, Alex Casillas, and In-Sue Oh. 2008. "Third-Year College Retention and Transfer: Effects of Academic Performance, Motivation, and Social Connectedness." *Research in Higher Education* 49 (7): 647–64.

Elliott, Kevin M., and Margaret A. Healy. 2001. "Key Factors Influencing Student Satisfaction Related to Recruitment and Retention." *Journal of Marketing for Higher Education* 10 (4): 1 11. https://doi.org/10.1300/J050v10n04_01.

Fisher, Bonnie S. 1995. "Crime and Fear on Campus." *The Annals of the American Academy of Political and Social Science* 539: 85–101.

Grossman, M., & Markowitz, S. (1999). "Alcohol Regulation and Violence on College Campuses" (No. w7129). National Bureau of Economic Research.

Janosik, Steven M., and Donald D Gehring. 2003. "The Impact of the Clery Campus Crime Disclosure Act on Student Behavior." *Journal of College Student Development* 44 (1): 81–91. https://doi.org/10.1353/csd.2003.0005.

Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics Act of 1990, 20 U.S.C. §1092(f).

Marcus, Richard D. 1989. "Freshmen Retention Rates at U.s. Private Colleges: Results from Aggregate Data." *Journal of Economic & Social Measurement* 15 (1): 37–55. https://doi.org/10.3233/JEM-1989-15103.

Marsh, Gregory. 2014. "Institutional Characteristics and Student Retention in Public 4-Year Colleges and Universities." *Journal of College Student Retention: Research, Theory & Practice* 16 (1): 127–51. https://doi.org/10.2190/CS.16.1.g.

McCallion, Gail. 2014. "History of the Clery Act: Fact Sheet." *Fact Sheet*, 3.

McPheters, Lee R. 1978. "Econometric Analysis of Factors Influencing Crime on the Campus." *Journal of Criminal Justice* 6 (1): 47–52. https://doi.org/10.1016/0047-2352(78)90038-7.

Ménard, Kim S., Gordon C. Nagayama Hall, Amber H. Phung, Marian F. Erian Ghebrial, and Lynette Martin. 2003. "Gender Differences in Sexual Harassment and Coercion in College Students: Developmental, Individual, and Situational Determinants." *Journal of Interpersonal Vio8lence* 18 (10): 1222–39. https://doi.org/10.1177/0886260503256654.

Morrall, P., P. Marshall, S. Pattison, and G. Macdonald. 2010. "Crime and Health: A Preliminary Study into the Effects of Crime on the Mental Health of UK University Students." *Journal of Psychiatric and Mental Health Nursing* 17 (9): 821–28. https://doi.org/10.1111/j.1365-2850.2010.01594.x.

Morriss, Susan B. 1993. "The Influences of Campus Characteristics on College Crime Rates,"

Annual Forum of the Association of for Institutional Research.

Pampaloni, Andrea M. 2010. "The Influence of Organizational Image on College Selection: What Students Seek in Institutions of Higher Education." *Journal of Marketing for Higher Education* 20 (1): 19–48. https://doi.org/10.1080/08841241003788037.

Ullman, S. E., Karabatsos, G., & Koss, M. P. 1999. "Alcohol and Sexual Assault in a National Sample of College Women." *Journal of interpersonal violence*, *14*(6), 603-625.

United States Department of Education. Office of Postsecondary Education. *Campus Safety and Security 2014/2015-2017/2018*. https://ope.ed.gov/campussafety/#/datafile/list

United States Department of Education. *College Scorecard Data 2014-2018*. https://collegescorecard.ed.gov/data/

Volkwein, J. Fredericks, Bruce P. Szelest, and Alan J. Lizotte. 1995. "The Relationship of Campus Crime to Campus and Student Characteristics." *Research in Higher Education* 36 (6): 647–70.

Warwick, Jacquelyn, and Phylis Mansfield. 2003. "Perceived Risk in College Selection: Differences in Evaluative Criteria Used by Students and Parents." *Journal of Marketing for Higher Education* 13 (1/2): 101–25. https://doi.org/10.1300/J050v13n01_07.

Wooldridge, Jeffrey M. 2011. "RE: St: Fixed Effect Correcting Auto Correlation and Heteroskedasticity." Stata. March 1, 2011. https://www.stata.com/statalist/archive/2011-03/msg00035.html.

Appendix

Table 1: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
retention	1,144	.8203791	.0978118	.4007	.9952
sexualassa~t	1,144	10.67907	13.98545	.5388	210.6772
shareblack	1,144	.0955905	.1261738	.0047	.9518
sharefemale	1,144	.5435628	.081982	.2073	1
fulltime	1,144	.9032165	.0897603	.512	1
avgcost	1,144	34455.13	16375.21	13542	72717
avgsat	1,144	1146.889	149.6161	712	1558
ppcrime	1,144	2.669592	6.655355	0	77.75999
alcoholvio~n	1,144	206.8422	291.5297	0	2516.346
campuscrime	1,144	29.50438	31.9104	.6393	324.5844
year	1,144	2015.5	1.118523	2014	2017

Table 2: Primary regression

```
Fixed-effects (within) regression               Number of obs      =       1,143
Group variable: unitid                          Number of groups   =         286

R-sq:                                           Obs per group:
     within  = 0.5913                                         min =           3
     between = 0.7724                                         avg =         4.0
     overall = 0.6290                                         max =           4

                                                F(10,285)          =       83.10
corr(u_i, Xb)  = 0.1002                         Prob > F           =      0.0000

                                  (Std. Err. adjusted for 286 clusters in unitid)
------------------------------------------------------------------------------
             |               Robust
  clretention |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
csexualassault |  -.0004566    .000191    -2.39   0.017    -.0008326   -.0000807
   cshareblack |  -.1594076   .0630261    -2.53   0.012    -.2834633    -.035352
  csharefemale |   .0075372   .0443517     0.17   0.865    -.0797614    .0948357
     cfulltime |   .1117234   .0531694     2.10   0.036     .0070689    .2163779
     clavgcost |   .0088933   .0108769     0.82   0.414    -.0125159    .0303024
      clavgsat |   .5880709   .0398558    14.75   0.000     .5096219    .6665199
      cppcrime |  -.0004703   .0003138    -1.50   0.135    -.0010879    .0001474
   calcoholvio |   .0000329   9.44e-06     3.49   0.001     .0000143    .0000515
  ccampuscrime |   .0004298   .0001359     3.16   0.002     .0001624    .0006972
          year |  -.0213436   .0019102   -11.17   0.000    -.0251036   -.0175837
         _cons |   43.01815   3.850086    11.17   0.000     35.43993    50.59636
-------------+----------------------------------------------------------------
       sigma_u |  .01970549
       sigma_e |  .05665476
           rho |  .10792082   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Table 3: Joint significance test between campus alcohol violation and campus crime coefficients

```
. test calcoholvio ccampuscrime

 ( 1)  calcoholvio = 0
 ( 2)  ccampuscrime = 0

       F(  2,   285) =   18.02
            Prob > F =    0.0000
```

Table 4: Davidson MacKinnon Test to rule out levels model relative to log model (with retention, average SAT, and average cost as log variables in the *fitted* coefficient)

Source	SS	df	MS		
Model	8.37472769	10	.837472769	Number of obs	= 1,144
Residual	2.5605133	1,133	.002259941	F(10, 1133)	= 370.57
				Prob > F	= 0.0000
Total	10.935241	1,143	.00956714	R-squared	= 0.7658
				Adj R-squared	= 0.7638
				Root MSE	= .04754

retention	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sexualassault	.0003795	.0002403	1.58	0.115	-.0000919	.0008509
shareblack	.13087	.0377118	3.47	0.001	.0568772	.2048629
sharefemale	-.0649539	.0261473	-2.48	0.013	-.1162564	-.0136514
fulltime	-.0841342	.0411346	-2.05	0.041	-.1648428	-.0034256
avgcost	3.24e-08	1.35e-07	0.24	0.810	-2.32e-07	2.97e-07
avgsat	-.0003682	.0001385	-2.66	0.008	-.00064	-.0000964
ppcrime	.000026	.0002341	0.11	0.912	-.0004333	.0004852
alcoholviolation	-.0000223	9.50e-06	-2.34	0.019	-.0000409	-3.63e-06
campuscrime	-.0002949	.000155	-1.90	0.057	-.0005992	9.28e-06
fitted	1.449212	.2397629	6.04	0.000	.9787823	1.919641
_cons	1.64732	.2510261	6.56	0.000	1.154791	2.139848

Table 5: Davidson MacKinnon Test to rule out log model relative to log-differenced model (with all variables differenced in the *fitted2* coefficient)

Source	SS	df	MS		
Model	13.0766145	10	1.30766145	Number of obs	= 1,143
Residual	4.69127843	1,132	.004144239	F(10, 1132)	= 315.54
				Prob > F	= 0.0000
Total	17.7678929	1,142	.015558575	R-squared	= 0.7360
				Adj R-squared	= 0.7336
				Root MSE	= .06438

logretention	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sexualassault	-.0008148	.0002281	-3.57	0.000	-.0012624	-.0003672
shareblack	-.1209446	.0173665	-6.96	0.000	-.1550189	-.0868704
sharefemale	.072485	.0246001	2.95	0.003	.024218	.120752
fulltime	.149935	.0279305	5.37	0.000	.0951336	.2047363
logavgcost	.0017418	.0062606	0.28	0.781	-.010542	.0140255
logavgsat	.699089	.0245162	28.52	0.000	.6509867	.7471913
ppcrime	.0000177	.0003172	0.06	0.956	-.0006046	.00064
alcoholviolation	.0000303	7.69e-06	3.94	0.000	.0000152	.0000454
campuscrime	.0006163	.0001086	5.67	0.000	.0004032	.0008295
fitted2	-.1296938	.0314575	-4.12	0.000	-.1914154	-.0679721
_cons	-5.321727	.1478411	-36.00	0.000	-5.611801	-5.031654

Table 6: Breusch-Pagan Test for heteroskedasticity for primary regression

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of clretention

         chi2(1)      =     2.35
         Prob > chi2  =   0.1250
```

Table 7: White Test for heteroskedasticity for primary regression

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(54)     =    293.79
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	293.79	54	0.0000
Skewness	46.65	9	0.0000
Kurtosis	16.37	1	0.0001
Total	356.81	64	0.0000

Table 8: Variance Inflation Factors for primary regression

Variable	VIF	1/VIF
ccampuscrime	3.17	0.315419
clavgsat	3.02	0.331344
csexualass~t	2.65	0.376829
clavgcost	1.89	0.527899
cfulltime	1.51	0.663439
year	1.49	0.670295
cshareblack	1.39	0.721810
calcoholvio	1.32	0.756165
cppcrime	1.13	0.887326
csharefemale	1.09	0.918055
Mean VIF	1.87	

Table 9: Pairwise Correlation Coefficient Matrix for primary regression

e(V)	csexua~t	cshare~k	cshare~e	cfullt~e	clavgc~t	clavgsat	cppcrime
csexualass~t	1.0000						
cshareblack	0.0153	1.0000					
csharefemale	-0.0489	0.0325	1.0000				
cfulltime	-0.0287	-0.0397	0.0935	1.0000			
clavgcost	-0.0194	-0.1939	-0.1643	-0.3499	1.0000		
clavgsat	0.0198	0.4931	0.1739	-0.1796	-0.4999	1.0000	
cppcrime	0.1005	-0.0834	0.0595	0.0886	-0.0723	-0.0723	1.0000
calcoholvio	0.0621	0.0201	0.0155	-0.1327	0.1595	-0.1215	-0.0235
ccampuscrime	-0.7429	-0.0571	0.0478	-0.0021	0.1162	-0.1736	-0.2194
year	-0.1446	-0.2789	-0.1080	0.1032	0.1871	-0.5337	0.0526
_cons	0.1446	0.2789	0.1080	-0.1032	-0.1871	0.5337	-0.0526

e(V)	calcoh~o	ccampu~e	year	_cons
calcoholvio	1.0000			
ccampuscrime	-0.2721	1.0000		
year	0.1512	0.1271	1.0000	
_cons	-0.1512	-0.1271	-1.0000	1.0000