The Effect of
Sports Participation on GPAs:
A Conditional Quantile Regression Analysis

Andreas Santucci

The University of California, at Berkeley

Economics

May 7, 2012

Advisor: Professor Gregory M. Duncan

Dedicated to Saverio Santucci

Contents

List of Figures

List of Tables

Chapter 1

Acknowledgements

I would like to thank my family for helping to support me throughout my life, my peers who have helped nurture my interest in economics, and my team mates who helped me develop a passion for athletics. I would also like to give special thanks to Charlie Gibbons and Gregory M. Duncan for inspiring me to pursue econometric research, as well as for their invaluable assistance in preparing this analysis.

Abstract

There exists a persistent stereotype that student-athletes are not as academically inclined as non student-athletes. This viewpoint is supported through primarily anecdotal evidence and aggregate level data, which does not reveal the causal effect of sports participation on grades. Is it the case that student-athletes achieve different grades because they are time constrained, or is it because they possess a systematically different academic ability? This study attempts to uncover a causal relationship between sports participation and academic achievement by examining panel data of student-athletes attending the University of California at Berkeley. Results show that the effects of participating in Division I athletics on grade point averages vary conditional on gender, sport, and academic ability. In expectation, the term grade point average for football players with the lowest academic ability will increase by 0.770 grade points if they participate in sports. For women's crew members with a very high academic ability, their term grade point average is expected to decrease by 0.456 grade points if they participate in sports. Both of these estimators account for the prior two semester grade point averages, financial aid status, academic year, and other unobservable factors which are time-invariant. Among other cohorts of student-athletes, the density of support is insufficient and we cannot estimate the marginal effect of sports participation on term grade point averages.

Chapter 2

Introduction

A typical weekday for a student-athlete involves, at a minimum, four hours of intense training in addition to a full time academic schedule. In several sports, student-athletes are expected to exhaust their energy in a morning workout, and arrive to class before most students have even stepped out of bed. During season, student-athletes may be expected to travel across the country whilst keeping up with missed lectures and exams. Devoting additional time and energy to athletics leaves student-athletes with fewer resources available for school. Proponents of sports participation argue that the increased burden forces students to practice time management skills and goal setting, which may help their grades. Does participating in sports make grades go up, or down?

NCAA regulations limit the maximum number of "practice hours" to twenty per week during the season of competition (Josephs 2006). However, the estimate of hours is biased downward due to intentional loopholes. Competition days do not count for more than three hours of practice time, no matter how long the duration. Time spent travelling or rehabilitating injuries is disallowed from counting towards practice time. Furthermore, many coaches strongly encourage "optional" practices. The sheer amount of time spent participating in athletics is equivalent to holding a part time job, and the amount of daily energy expenditure required to train in a Division 1 program leaves student-athletes at a deficit of mental and physical resources. It should be no surprise that student-athletes tend to achieve less academically when compared with their peers.

Aggregate level data reveals significant heterogeneity in grade point averages between student-athletes and non student-athletes. According to an Academic Performance Survey conducted by the Athletic Study Center at the University of California, at Berkeley in 2011, both male and female student-athletes consistently underachieve relative to their peers. Specifically, male student-athletes disproportionately earn more GPAs below 3.0 relative to the rest of the student body; most GPAs between 3.0 and 3.5 are achieved by female student-athletes. However, the general undergraduate population has the highest representation within the highest GPA bracket. Although aggregate level data indicates student-athletes achieve lower grades than their peers, the data fail to explain why.

My personal anecdote does no better in discerning a causal relationship. As a child, I trained every day after school for several hours with hopes of becoming an Olympic gymnast. Due to an injury sustained in high school, my gymnastics career ended. I participated casually in high school diving, but did not begin to again participate in athletics seriously until I walked on to Cal's Swim and Dive team; this semester coincided with my transfer to Cal from another academic institution. Coincident to walking onto the team, my first semester grades dropped an insignificant amount relative to my grades previously. However, since then my grades have seen an upward trend. I achieved my first 4.0 grade point average at Cal during the same semester in which I became a member of the team's travel squad and competed all over the western region of the United States.

This is in no way useful of discerning the effects of athletics on academic achievement, because anecdotal evidence does not hold enough other factors constant. Did my grades go down at first because I walked onto the diving team? Or did they go down because I was transitioning to a new learning environment? Did becoming a member of the travel squad definitively improve on my academic performance? Was I always a "student-athlete" since childhood? There are too many possible confounding factors to draw any meaningful conclusions.

Take a moment to consider how grades vary within the population of student-athletes. What makes a sport unique also makes its participants unique in a systematic way. It is reasonable to assume that individuals who are attracted to a particular sport may have an unobservable characteristic that are commonly shared between them which also effect school performance. Different sports expect uniquely different things from participants, and so the effect of sports on schooling varies by sport.

Furthermore, we may expect that grades vary conditional on possessing varying levels of academic ability, for which we use SAT scores as a proxy. Intuitively, the costs of studying are lower for higher ability individuals because they do not have to work as hard to achieve the same results. If the marginal cost decreases, we will observe high ability individuals achieving higher grades not only because they are smarter, but also because they study more.

On the other hand, it may be reasonable to expect that an individual with a "low" academic ability, as measured by SAT scores, may find schooling harder, study less, and therefore earn worse grades. Taking a mean regression risks having these effects 'cancel' out, and so our regression analysis will condition on belonging to a particular quantile within the distribution of SAT scores. Figure 2.1 demonstrates that conditioning on ability changes our prediction for grade point averages.

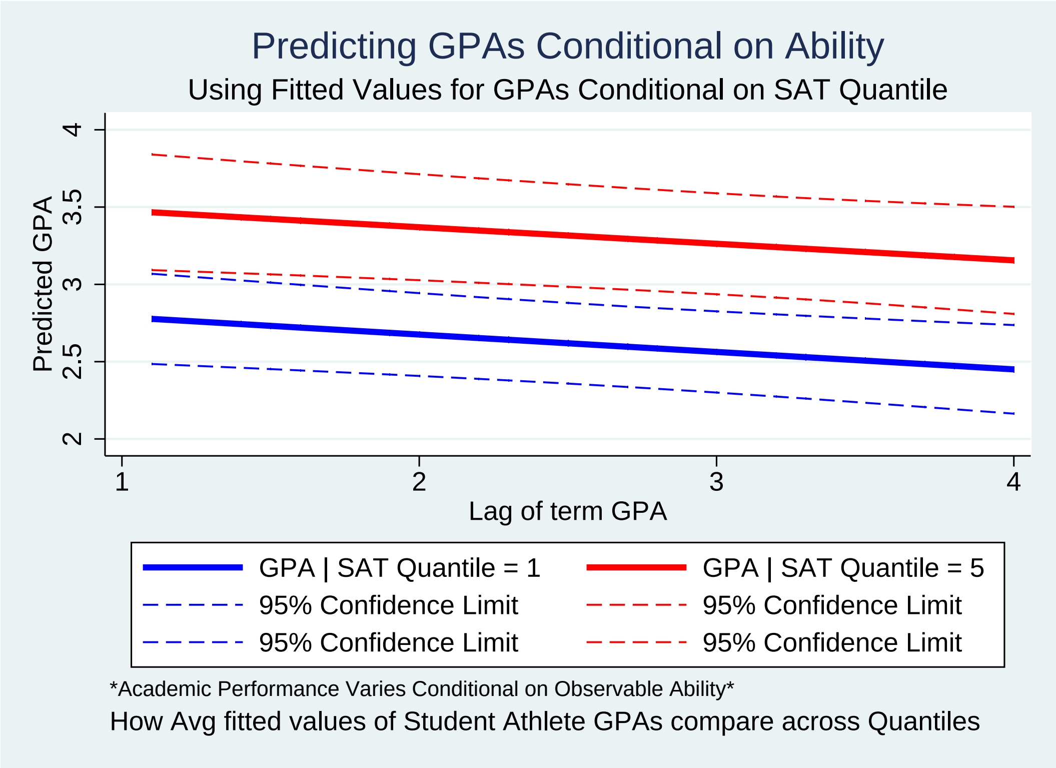

Figure 2.1: How Academic Ability Predicts Academic Achievement

By conditioning on more information, our predictions become more accurate. Conditioning on sport (and therefore gender) and ability will help us to better learn about the causal relationship between sports and grades.

Understanding the way in which athletics causes change in academic performance is important for several reasons. On an individual scale, utility maximization and realization of true preferences is only made possible given perfect information. An incoming athlete can better decide if they wish to participate in sports if they know how athletics will affect future grade performance. Imagine an individual who wants to play sports, but only under the condition that it doesn't hurt their academic performance by more than a certain margin.

However, the implications of this issue extend beyond the academic world and into the job market. Employers use observed grade point averages as a signal or proxy variable to predict innate ability; they believe observed grades contain important information necessary to gauge future job performance. With knowledge of marginal effects of sport participation, employers can more efficiently use grades as a signal of ability.

This study hopes to gain an insight of the causal relationship of sports participation and academic performance through the use of panel data and a fixed effects model which conditions on team affiliation and academic ability. The process allows us to derive an elegant intuition about how the effects of sports participation on grades vary both by sport and academic ability.

Chapter 3

Background and Assumptions

Scientists are able to create experiments which determine causal relationships. The design of an experiment is such that all other factors other than the variable of interest are held constant. A control group receives no treatment, and is compared with an observationally identical treatment group. The treatment and control group can be randomly determined such that, although the two groups are not observationally identical, differences between the two groups are not systematic. The observed differences can therefore be attributed solely to treatment because that is the only factor that varies systematically between the groups.

Social scientists do not have the privilege of creating such experiments for ethical reasons. Instead, we must rely on quasi-experiments to infer causality; treatment and control groups are still utilized, but selection into each group is not random. The worry with comparing the two groups relates to confounding variables; differences in outcomes stem not only from treatment, but also due to systematic differences between the two groups that affect the outcome of interest (Heckman and Robb 1986).

In this study, the observed outcome of interest is a student-athlete's term grade point average. Our control group consists of student athletes not currently participating in sports, and our treatment group consists of student-athletes who are currently participating in sports. We wish to compare the grade point averages between these groups to learn the effect of sports participation on academic performance.

Let us formalize some notation before proceeding. $GPA_{S=1}$ and $GPA_{S=0}$ represent the hypothetical grade point average that an individual would earn if they did or did not participate in sports, respectively. Each individual is associated with a hypothetical $(GPA_{S=0}, GPA_{S=1})$ pair in additional to their observed grade point average, which is denoted by GPA. Let S be an indicator variable that takes on a value of 0 if the individual is not currently participating in sports, and 1 if they are. X represents a vector of observable characteristics, and $X = x$ indicates that we are conditioning on a sub-population homogeneous in observable characteristics.

We want to learn about the true effect of sports participation, holding all other variables fixed. Unfortunately, we cannot observe a $(GPA_{S=0}, GPA_{S=1})$ pair for every semester of each individual. If we could, then simply taking the difference between the pairs of grade point averages would reveal the causal effect of sports participation on grades. The next best comparison we can observe is how an individual's grades change between semesters in which sports participation also changes. Presumably in this case, all other factors are held constant, and we are left with strictly the causal effect.

Imagine an individual who plays sports one semester, but is injured and cannot participate the following semester. It may be reasonable to assume that injuries are exogenously determined by factors that do not affect participation in sports or academic performance. In this case, it's as if this individual has randomly been assigned from the treatment to control group. The difference in term grade point averages before and after the injury would closely approximate the effect of sports participation. This is an example of a naturalized experiment.

Creatively envisioning naturalized experiments, in which an exogenously determined event forces individuals to change into or out of a treatment group, have become recently popular [1]. Ideally, we wish that within this quasi-experiment and among observationally similar individuals, sports participation is independent of the grade point averages that would have been earned as an athlete or non-athlete. Formally, we may describe this as *exogenous treatment assignment* (Card and Sullivan 1988) :

$$\Big(GPA_{S=0}, GPA_{S=1}\Big) \perp\!\!\!\perp S \Big| X = x \tag{3.0.1}$$

This is a very strong assumption. It literally states that after conditioning on observable characteristics, grades earned as a student-athlete and grades earned as a non student-athlete are independent of participation in sports. This is tough to conceptualize because the two outcomes seem to depend entirely on sports participation. However, if this assumption is true, we can compare grade point averages between student-athletes and their peers who are observationally similar and attribute the difference solely to sports; it plays a necessary role in allowing us to infer causality.

[1] My general understanding of finding quasi-experiments was enhanced by reading Levitt and Dubner (2005) and especially through discussions with Charlie Gibbons. Charlie's insights on naturally occurring quasi-experiments for sports participation at Cal have provided the foundation for this analysis. Without his help, this paper would not be possible.

3.1 Motivating Exogenous Treatment Assignment

Let us motivate a more intuitive notion of this concept. In general, we hope that economic agents make decisions based on things they know about, their personal characteristics. If we examine two individuals who are perfectly identical in every possible way, there is no more available information left with which we can further use to help predict sports participation. Therefore, after conditioning on all possible characteristics and comparing two identical individuals, the athletic status of a student-athlete is seemingly random.

Unfortunately, in reality we cannot possibly condition on all characteristics, because they are not all observable. Let us define academic performance or achievement by term grade point averages. Furthermore, consider the notion of academic ability being broken up into unobservable characteristics, such as motivation and discipline, and observable characteristics, such as SAT scores. It is reasonable to expect that academic ability is highly predictive of academic performance. Because academic ability cannot fully be observed, we will use strictly the observable component, specifically SAT scores, as a proxy variable.

3.1.1 Using SAT Scores as a Proxy Variable

Redundancy

Before we discuss how this relates to our notion of exogenous treatment assignment, let's note several mathematical assumptions we implicitly make when using proxy variables in a regression analysis (Graham 2012). The first assumption is known as *redundancy* and can be stated as follows:

$$\mathbb{E}\Big[GPA|Ability, SAT\Big] = \mathbb{E}\Big[GPA|Ability\Big] \tag{3.1.1}$$

where *Ability* represents the sum total of unobservable and observable ability. In expectation, conditioning on SAT scores in addition to academic ability does not help us to predict term grade point average. This is a weak assumption; despite not being able to measure academic ability, we may conceptualize it richly. Further conditioning on SAT scores adds no new information because observed academic ability is by definition a subset of total academic ability.

Conditional Mean Independence

Another assumption that is required for proxy variables to be considered a good proxy is *conditional mean independence.* This assumption is more formally stated as follows (Graham 2012):

$$\mathbb{E}\Big[Ability|SAT, S\Big] = \mathbb{E}\Big[Ability|SAT\Big] \tag{3.1.2}$$

Conditional on our proxy of SAT scores, observing participation in sports does not help us to explain academic ability. This is a much stronger assumption than 3.1.1. Conditional mean independence implies zero correlation, and so 3.1.2 states that participation in sports is not correlated at all with academic ability.

Let's step back and think about how we can better predict grades in general. If it were possible, conditioning on total academic ability would help us to form our expectation of term grade point averages. Putting these assumptions together with this idea paints a nice picture of how we can better determine the effect of sports participation on term grade point averages.

The two previous assumptions now help us to explain our *exogenous treatment assignment* or *selection on observables* assumption. We assume that ability is a good way to predict academic performance, and ability is best predicted through SAT scores. We further assumed that once we condition on SAT scores, sports participation is independent of academic achievement. This is important because we can now compare grades of observationally similar athletes and non-athletes, indexed by SAT scores and team affiliation, and attribute the differences to the causal effect of sports participation.

3.2 β^{ATT}, β^{ATE}, and β^{CATE}

In our data, we are only given *observed grade point average*, which can be represented as:

$$GPA = \Big(S\Big)\Big[GPA_{S=1}\Big] + \Big(1 - S\Big)\Big[GPA_{S=0}\Big] \tag{3.2.1}$$

In this example, GPA depends only on S, an indicator variable, and so the value of GPA will happen to coincide with one of the hypothetical grade point averages within the pair $(GPA_{S=0}, GPA_{S=1})$.

3.2.1 β^{ATE}: Average Treatment Effect

This equation helps us envision how we can learn of the *average treatment effect* of sports on grades (Imbens and Angrist 1994):

$$\beta^{ATE} = \mathbb{E}\Big[GPA_{S=1} - GPA_{S=0}\Big] \tag{3.2.2}$$

Note that the average treatment effect is an unconditional expectation, and it includes the effect of sports participation on non-athletes. β^{ATE} tells us the average effect of sports participation on grade point averages for all individuals, regardless of whether they are student-athletes. This information is not particularly interesting because it does not capture the true effect of playing sports for athletes and is a simple mean difference. The average treatment effect is of primary interest only if our assumption of *exogenous treatment assignment* holds true. In this case, there is no bias and $\beta^{ATT} = \beta^{ATE}$.

3.2.2 Comparing β^{ATT} & β^{ATE}

Let us now relate the average treatment effect to the average treatment effect on the treated using short and long regressions. Note that *selection on observables* implies that among observationally identical individuals, whatever jointly determines $GPA_{S=1}$ and $GPA_{S=0}$ is independent of what determines participation in sports. The joint distribution between our variables of interest can be described more accurately as follows:

$$f\left(GPA_{S=1}, GPA_{S=0}, S \middle| Ability\right) \\ = f\left(GPA_{S=1}, GPA_{S=0} \middle| Ability\right) \times f\left(S \middle| Ability\right)$$

Random assignment implies independence, which implies conditional mean independence, which implies zero correlation between sports participation and grades. The following equations describe an important result [2]:

$$\implies \mathbb{E}\Big[GPA_{S=0}\Big|S=1\Big] = \mathbb{E}\Big[GPA_{S=0}\Big|S=0\Big] = \mathbb{E}\Big[GPA_{S=0}\Big] \tag{3.2.3}$$

Plugging this result into 3.2.2, our definition for the average treatment effect, shows that there is no selection bias under random assignment. To show this, we start with an expectation of observed differences in grade point averages that can be computed subject to sampling variability:

$$\mathbb{E}\Big[GPA\Big|S=1\Big] - \mathbb{E}\Big[GPA\Big|S=0\Big] \tag{3.2.4}$$

Selection on observables allows us to manipulate observed differences to learn about the *average treatment effect on the treated.* Simply adding and subtracting unobservable expectations, $\mathbb{E}[GPA_{S=0}|S = 1]$, can help us group together terms in a way that develops a more parsimonious interpretation.

[2] Equations 3.2.3 through 3.2.6 implicitly condition on *Ability*.

$$\mathbb{E}\Big[GPA\Big|S=1\Big]-\mathbb{E}\Big[GPA\Big|S=0\Big]$$
$$=\beta^{ATT}+\Bigg(\mathbb{E}\Big[GPA_{S=0}\Big|S=1\Big]-\mathbb{E}\Big[GPA_{S=0}\Big|S=0]\Bigg) \tag{3.2.5}$$

Where we can describe the selection bias as

$$\mathbb{E}\Big[GPA_{S=0}\Big|S=1\Big]-\mathbb{E}\Big[GPA_{S=0}\Big|S=0\Big] \tag{3.2.6}$$

Selection bias is 0 if there is exogenous treatment assignment, or if people participate in sports randomly.

3.2.3 β^{ATT}: Average Treatment Effect on the Treated

Although we cannot observe a $(GPA_{S=0}, GPA_{S=1})$ pair for each individual in reality, our previous assumptions can help us to get an idea of how we may better approximate the effect of sports participation on grades. Consider the *average treatment effect on the treated* of sports on grades (Hirano et al. 2003):

$$\beta^{ATT}=\mathbb{E}\Big[GPA_{S=1}-GPA_{S=0}\Big|S=1\Big]. \tag{3.2.7}$$

Identifying β^{ATT} efficiently is the goal of many policy analysts, but we have discussed previously how this direct comparison is not possible. The above expectation effectively holds factors fixed that may be systematically different between the two groups of interest and could otherwise confound our results.

Our assumption of *selection on observables* allows us to omit sports participation from our conditional expectation without losing any information. We can therefore write out the difference in expectations as:

$$\beta^{CATE}=\mathbb{E}\Big[GPA_1\Big|Ability\Big]-\mathbb{E}\Big[GPA_0\Big|Ability\Big] \tag{3.2.8}$$

Equation 3.2.8 is known as the *conditional average treatment effect.* Given data, this expectation is identifiable, and it's correct if our assumptions are true. β^{CATE} tells us the average effect of participating in sports for athletes and non-athletes who possess a common academic ability.

3.2.4 Relating β^{CATE} to β^{ATT} using L.I.E.

Using the *law of iterated expectations* can help us to relate β^{CATE} to β^{ATE} and β^{ATT}.

$$\begin{aligned} \beta^{ATE} = \mathbb{E}\Big[GPA1 - GPA_0\Big] &= \mathbb{E}\Big[\mathbb{E}\big[GPA_1 - GPA_0\big]\Big|Ability\Big] \\ \implies \beta^{ATE} &= \mathbb{E}\Big[\beta^{CATE}\big(Ability\big)\Big] \end{aligned} \tag{3.2.9}$$

The expectations tells us that we can calculate the *average treatment effect*, β^{ATE}, by averaging the conditional average treatment effect, β^{CATE}, across all values of *Ability*.

Once we have the *average treatment effect*, we can derive the *average treatment effect of the treated*, our parameter of interest.

$$\begin{aligned} \beta^{ATT} &= \mathbb{E}\Big[GPA_1 - GPA_0\Big|S = 1\Big] \\ = \mathbb{E}\Big[\mathbb{E}\big[GPA_1 - GPA_0|Ability, S = 1\big]&\Big|S = 1\Big] \\ \implies \beta^{ATT} &= \Big[\beta^{ATE}\Big|S = 1\Big] \end{aligned} \tag{3.2.10}$$

The above result is not surprising. The average treatment effect on the treated is equivalent to the average treatment effect conditional on belonging to the treatment group.

Apples to Apples Comparison

Our assumptions leave us with an apples to apples comparison from which we can attempt to make a causal claim. The proceeding analysis assumes linearity in *Ability*, which is quite restrictive. However, this permits us to conveniently describe the expected grades of students athletes as $\mathbb{E}[GPA_1|Ability] = \alpha_1 + \beta_1 \times Ability$ and the expected grades of non-student-athletes as $\mathbb{E}[GPA_0|Ability] = \alpha_0 + \beta_0 \times Ability$. Plugging these expectations into our expectation of observed grade point averages yields the following:

$$\begin{aligned} \mathbb{E}\Big[GPA\Big|Ability, S\Big] &= \mathbb{E}\Big[(S)GPA_1 + (1 - S)GPA_0\Big|Ability, S\Big] \\ &= \Big(S\Big)\mathbb{E}\Big[GPA_1\Big|Ability, S\Big] + \Big(1 - S\Big)\mathbb{E}\Big[GPA_0\Big|Ability, S\Big] \\ &= \Big(S\Big)\mathbb{E}\Big[GPA_1\Big|Ability\Big] + \Big(1 - S\Big)\mathbb{E}\Big[GPA_0\Big|Ability\Big] \\ &= S\Bigg(\alpha_1 + \beta_1 Ability\Bigg) + \Big(1 - S\Big)\Bigg(\alpha_0 + \beta_0 Ability\Bigg) \implies \\ \mathbb{E}\Big[GPA\Big|Ability, S\Big] &= \alpha_0 + \beta_0 Ability + \Big(\alpha_1 - \alpha_0\Big)S + \Bigg[\Big(\beta_1 - \beta_0\Big)\Big(S \times Ability\Big)\Bigg] \end{aligned} \tag{3.2.11}$$

Given this understanding of the population expectation of grade point averages conditional on sports participation and academic ability, we can estimate a sample analogue through ordinary least squares. The procedure for estimating the sample analogue of the *average treatment effect on the treated* is as follows:

3.3 Feasible Estimation of $\widehat{\beta^{ATT}}$

- Calculate Sample Conditional Average Treatment Effect: $\widehat{\beta^{CATE}}$
  - Regress GPA_i on $\Big(\alpha_0\Big)$, $\Big([\alpha_1-\alpha_0]S_i\Big)$, $\Big(\beta_0 Ability_i\Big)$, & $\Big([\beta_1-\beta_0]S_i \times Ability_i\Big)$
  - $\widehat{\beta^{CATE}(Ability)} = [\widehat{\alpha_1-\alpha_0}] + [\widehat{\beta_1-\beta_0}] \times Ability$
- Compute Sample Average Treatment Effect: $\widehat{\beta^{ATE}}$
  - $\widehat{\beta^{ATE}} = \frac{1}{N} \times \sum_{i=1}^{N} \gamma[\widehat{Ability}_i]$
  - $\implies \widehat{\alpha_1-\alpha_0} + \widehat{\beta_1-\beta_0} \times \frac{1}{N}\sum_{i=1}^{N} Ability_i$
  - $\implies \widehat{\beta^{ATE}} = \widehat{\alpha_1-\alpha_0} + \widehat{\beta_1-\beta_0} \times \overline{Ability}$
- Compute Sample Average Treatment Effect on the Treated: $\widehat{\beta^{ATT}}$
  - $\widehat{\beta^{ATT}} = \frac{\sum_{i=1}^{N} S_i \times \gamma \widehat{Ability}_i}{\sum_{i=1}^{N} s_i}$
  - $\widehat{\alpha_1-\alpha_0} + \widehat{\beta_1-\beta_0} \times \frac{\sum_{i=1}^{N} S_i \times \widehat{Ability}_i}{\sum_{i=1}^{N} S_i}$
    - Where $\frac{\sum_{i=1}^{N} S_i \times \widehat{Ability}_i}{\sum_{i=1}^{N} S_i}$ is the average Ability within the S = 1 sub-sample

The above model is intended for a cross sectional data set. Note that dependent and explanatory variables are denoted by person ‘i’. The preceding analysis is important because it demonstrates that, given an ideal data set, we can in fact compute the average effect of sports participation among student-athletes specifically.

This study utilizes panel data, and therefore we cannot estimate $\widehat{\beta^{ATT}}$ using the above method. In our simple model, academic ability remains constant over time within each individual; there is zero variation of SAT score for each panel because it is a test that is only taken once. A variable that has zero variance will cause problems of multicollinearity in a regression, and SAT scores cannot be included as an explanatory variable using a fixed effects model.

We will still use the same principles derived in the preceding analysis. The goal is to compare observationally similar individuals, allowing sports participation to vary randomly. We will utilize SAT quantiles as a conditioning tool to derive efficient estimators for $\widehat{\beta^{ATT}}$ among select cohorts.

Chapter 4

Data

Data for this study come from the University of California at Berkeley, specifically the Athletic Study Center. The ASC is devoted to assisting student-athletes through their college careers and closing the achievement gap between student-athletes and their peers. Their most recent data set includes 3,809 student-athletes who attended Cal between 1999 and 2012.

The primary variables or parameters of interest within our data set include: semester GPA, sport teams, academic year, university start year, sat scores, and gender. This data set is advantageous because it overcomes the problem of comparing apples with oranges, as is done in a cross sectional data set. By tracking each individual over time, we can see how academic achievement varies with sport participation holding all other factors constant.

Individuals appear in the data set anywhere from 1 to 14 semesters; this data set is strongly unbalanced. It is worthwhile to consider why our data set is unbalanced, because if it relates to factors effecting academic performance, the problem may cause biased estimators.

4.1 Cleaning and Organization

The sample data has several questionable observations which must be handled before regression analysis can be performed. Observations which do not contain an identification tag are dropped from the data set; they are not useful in tracking specific individuals over time. Unfortunately, the data set is also subject to human input error, and therefore contains errors which are not systematic.

For example, there are at least two instances in which an individual is recorded as having a term grade point average of zero or one, but the cumulative grade point average for the semester is recorded as 4.0. Such observations have been removed from the data set, because this observation is mathematically impossible. The data set also contains several repeated observations which are problematic when attempting to organize the data in a 'long' panel format. Several duplicate observations have therefore been dropped.

Because the errors are not systematic, it is impossible to create a loop function which cleans the problematic observations. It is likely that within the remaining 20,000 observations, there are additional errors that have gone unnoticed.

However, there were several systematic methods used to clean the data set. For example, simply creating a gender indicator variable based on team affiliation is problematic, because gender appears to change when an individual walks on to a team, redshirts, or quits. These few instances can be tracked down and fixed systematically. Furthermore, females can participate on the men's crew team as coxswains. After flagging down problematic gendered observations for all sports in general, we can narrow down our search to instances in which the individual also participates in men's crew. There are only three reported cases in our data set of females participating as coxswains for men's crew teams and these have been fully accounted for within the data set.

In its original form, the data set also lists one observation for each year that a student participates or affiliates with Cal athletics; fall and spring term grade point average are listed for each row. Our model seeks to regress term grade point average on a vector of characteristics and a sports indicator variable rather than running separate regressions for each semester. Therefore, we stacked our data such that each observation represents a single semester for a single student. Because our data set was initially described in terms of years, it is difficult, although possible, to track down the exact semester in which an individual changes their status of affiliation with athletics.

4.2 Shortfalls of our Data Set

Despite the advantages of having panel data, our sample is far from perfect. Our analysis in the next section could be made much more efficient with access to term GPAs for all students across all observations as well as knowledge of major choice for each individual. The data set is not able to distinguish the difficulty between class loads. Although units are included in the data set, they are not a suitable proxy for course difficulty because expectations in time commitment vary across courses and units may be endogenously determined by participating in sports. Without using two stage least squares, using units as an explanatory variable for term GPA poses problems.

Our data set is also deficient of demographic identifiers that may be useful in predicting college performance. Variables such as age, parental income, parental education level, and ethnicity have all been omitted and have the potential to cause bias. The benefit of using panel data is that because parental education level and ethnicity are held constant over the duration of a student's college attendance, these variables are in fact controlled for in our regression analysis despite not being included as explanatory variables. Although our estimates are not biased, they could be more efficient if we could condition on these additional variables.

We are also missing further explanatory variables such as high school grade point average and whether the individual participated in sports in high school. In addition to SAT scores, high school grades could help to better proxy for unobserved ability. Also, whether or not an individual participated in sports in high school is likely to affect their high school GPA, and so knowing these two pieces of information would better help to account for academic ability.

Our data set could further be enhanced if it included observations for individuals before and after they participated in athletics. Currently, the data set only tracks individuals while they are semi-formally associated with Cal athletics. Our data set is fortunate to include information pertaining to the semesters in which students walk on to a sports team or when they are injured, which provides the basis for our quasi-experiments.

However, our observations would be more useful with follow up data. For example, a freshman who is cut from a sports team is only tracked during their freshman year. Observing their grades for the duration of their college career would help us get more efficient estimators. A similar problem exists for walk-ons and transfers, whose grades prior to Cal are not included within the data set.

The biggest shortfall is that even though SAT scores are contained within the data set, our analysis suffers because only approximately 1/2 of the individuals in the data set have a reported score. The standard for scoring SATs changed in 2005 and so we have no SAT scores for more recent observations. This is obviously problematic in terms of conditioning on observed ability. Conditional quantile regressions are therefore not possible for a large majority of our data set because there are an insufficient number of degrees of freedom within many of the cohorts of sport and SAT quantile configurations.

4.3 Systematic Tendencies - Knowing our Data

Before we dive into our regression analysis, let's get a feel for trends within our data set.

4.3.1 Comparing Ability

Among our entire sample, we have divided our distribution of SAT scores such that there is an even representation from each quantile.

If we run a simple regression conditional on belonging to the highest or lowest quantile of measurable ability, we see that our predictions for grade point averages are drastically different.

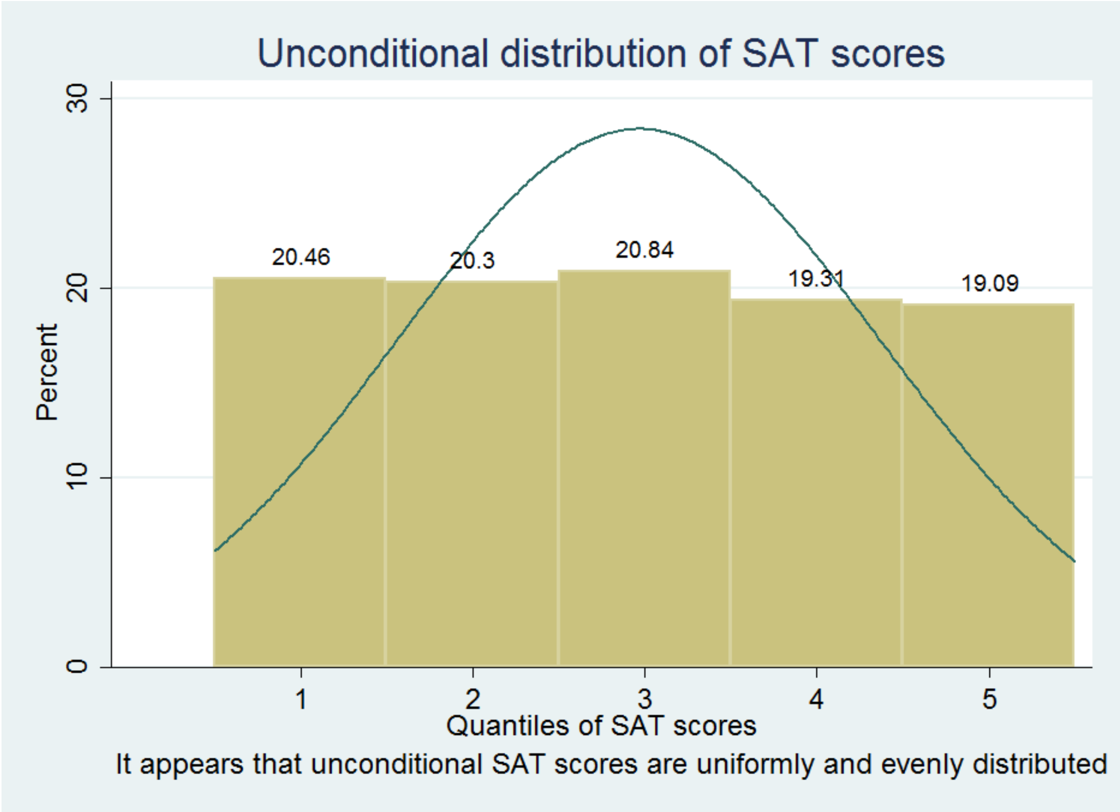

Figure 4.1: How Distribution of SAT Scores Varies Unconditionally

Table 4.1: Gender Summary

Male	Freq.	Percent	GPA	SAT Score
0	8,718	42.44	3.120997	1134.332
1	11,822	57356	2.88874	1118.007
Total	20,540	100.00	2.988226	1125.491

Figure 4.2 depicts fitted values for the cohort of individuals who have the highest quantile of observed ability, and compares them against fitted values for the cohort of individuals who have the lowest quantile of observed ability. Ability helps to predict academic performance; no matter what prior grades were, I will always predict an individual to earn higher grades if they scored higher on the SAT [1].

4.3.2 Comparing Gender

Within our sample, there are approximately 42% females and 58% males.

[1]The simple regression models the equation: $GPA = Sport + (LaggedGPA) + (TwiceLaggedGPA) + (FinancialAidIndicator) + (AcademicYear)$ using a fixed effects model with robust standard errors.

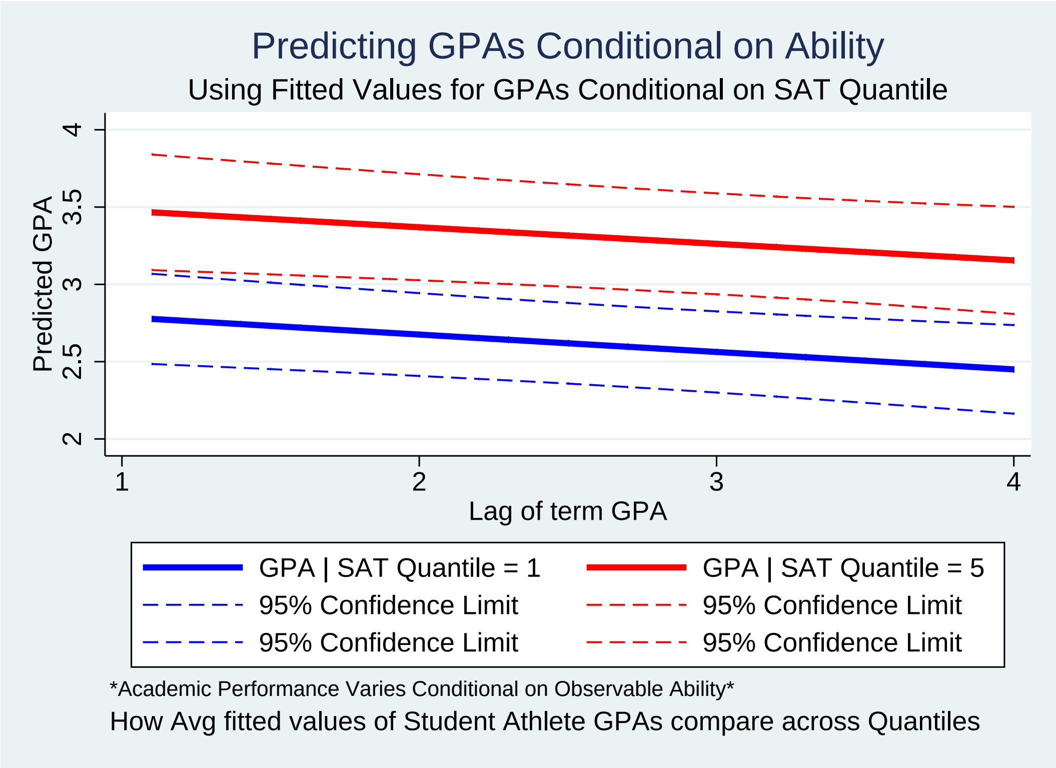

Figure 4.2: How Predicted GPA Varies Conditional on Ability

It would appear that within our sample, women on average have a higher observed academic achievement level relative to men. Our prediction for a female's grade point average, without any more information, is approximately 0.25 grade points higher relative to a male's predicted grade point average. Women also have a slightly higher observed ability, although the difference is not as significant; predicted SAT scores are approximately 16 points higher for females.

SAT Scores between Genders

Observing the distribution of SAT scores and comparing across gender shows that there are slight differences in the extreme quantiles.

If an SAT score was observed from the highest quantile, our best guess based on this information is that it was earned by a female. Males have the highest proportion of SAT scores within the lowest quantile, and women have the highest proportion of SAT scores with the highest quantile. The differences are negligible within the middle quantiles.

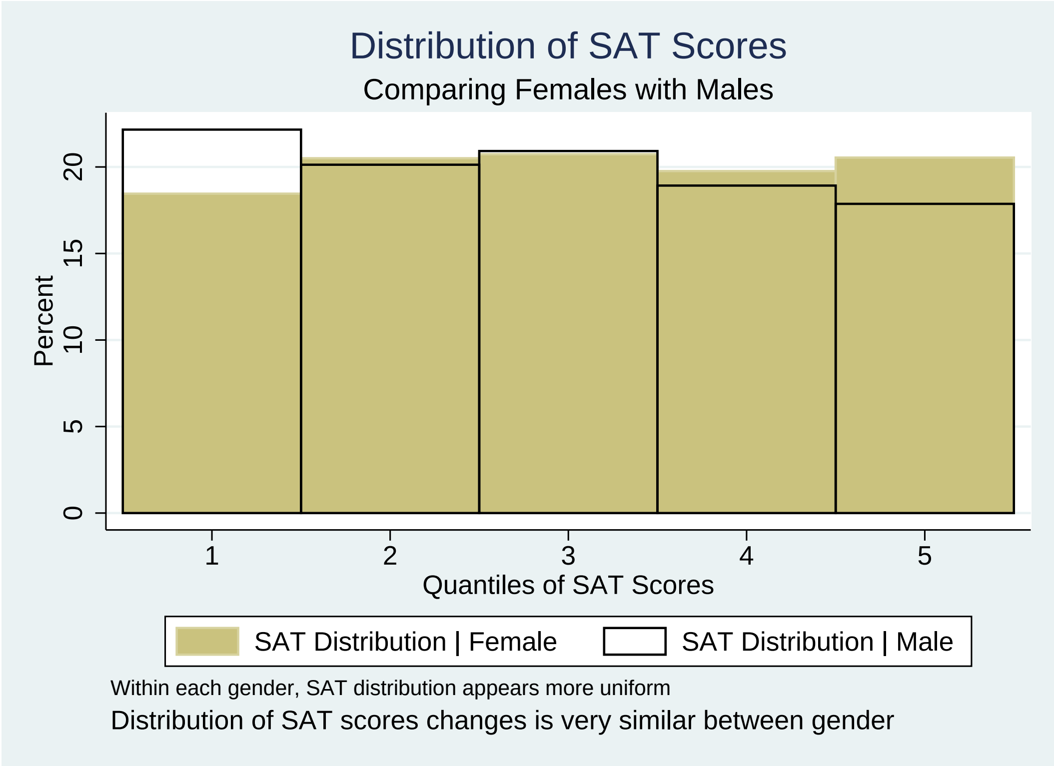

Figure 4.3: How Distribution of SAT Scores Varies with Gender

Predicting GPA on the basis of gender

On average in our data set, women earn higher grades than men. Although the ranges of grade point averages are similar, grades for males exhibit have a higher variance.

Table 4.2: Average GPA by Gender

Male	Observations	Mean	Std. Dev.	Min	Max
0	6,815	3.120997	0.5650616	0.2333333	4
1	9,095	2.88874	0.6431144	0.2	4

If we attempt to regress term GPA by gender and predict future grades as a function of lagged grade point average, we see that we predict women will achieve higher grades across the board [2].

Figure 4.4 captures the intuition of *regression towards the mean*: as last semester's grade point average increases, our prediction for next semester's grade point average decreases. Consider a student who has achieved a 4.0 and cannot mathematically achieve a higher grade point average the following semester. It is most reasonable to predict that a student who recently

[2] The simple regression models the equation: $GPA = Sport + (LaggedGPA) + (TwiceLaggedGPA) + (FinancialAidIndicator) + (AcademicYear)$ using a fixed effects model with robust standard errors.

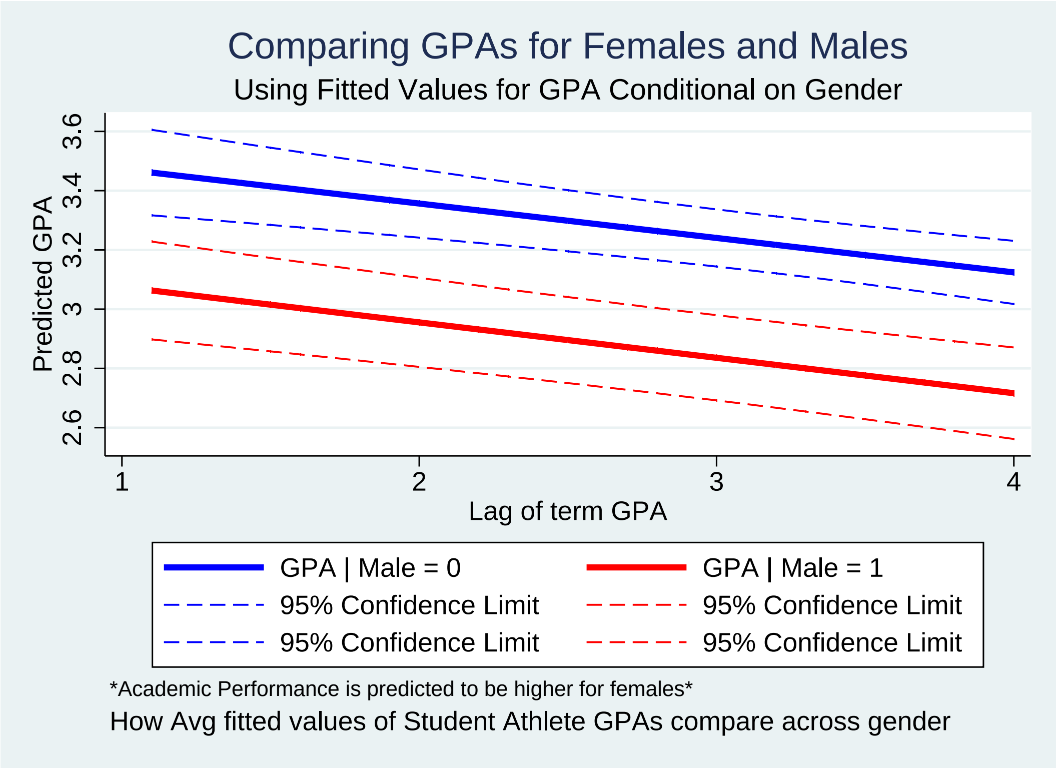

Figure 4.4: How Academic Achievement Varies with Gender

earned a high grade point average is likely to earn a lower grade point average in the upcoming semester. Comparable reasoning can be made to explain why students are predicted to earn better grades if their previous semester exhibited poor performance.

In conclusion of this section, it appears that within our sample, women are slightly “smarter” than men, both in terms of ability and achievement.

4.3.3 Comparing Sport Teams

Within our data set, football is the most frequently played sport, accounting for over 1/5th of our observations, whereas women’s badminton only has 2 observations. Although each sport is not evenly represented, the clustering of individuals within football, men’s crew, and women’s crew allows us to run a richer analysis within these groups.

Which team is the “smartest” in our data set? Table 4.3 describes the average grade point average across all years contained within the data set for select teams. Football, men’s baseball, and men’s basketball have the lowest average grade point averages out of all sports teams in our data set. Women’s crew, women’s track and field and cross country, and women’s water polo have the highest average grade point averages in our data set. I have additionally included the team average GPA within our data set for men’s crew, because it happens that they become an interesting bench mark later on.

Table 4.3: Team Summary

Sport	Frequency	Percent	Team GPA	Std. Dev.	Min.	Max.
Football	2,784	22.32	2.699676	0.6818955	0.25	4
Men's Crew	1,529	12.26	2.870704	0.6534745	0.2	4
Women's Crew	1,511	12.12	3.157063	0.5214417	0.971	4
Men's Rugby	1,251	10.03	2.916283	0.598578	0.378	4
Men's Track CC	1,233	9.89	2.911816	0.6415925	0.286	4
Women's Track CC	1,157	9.28	3.07357	0.6083145	0.889	4
Men's Water Polo	939	7.53	3.061487	0.5916037	0.443	4
Men's Baseball	903	7.24	2.863183	0.5425726	0.8	4
Women's Water Polo	697	5.59	3.083614	0.560725	0.6	4
Men's Basketball	467	3.74	2.820925	0.6959922	0.34	4
total	12,471	100.00				

Note that women's crew is the smartest team, on average. Their grades are not only highest in expectation, but they have the least variance and their worst recorded grade point average is higher than the worst instance for any other team. We would be fairly surprised to see an outlier in the women's crew team with a lower grade point average.

On the other hand, Football has the lowest average grade point average in our sample, and the variance of their average is among of the highest. We may predict football players to have a lower grade point average for each semester, but simultaneously would not be as surprised if we saw an outlier among football players with a higher grade point average.

The above averages are interesting because despite attempting a comparison between sports, our data still appear to be systematically divided by gender. Let us observe how ability changes on average within each sport and see if a similar result holds. Table 4.4 describes several SAT statistics, broken down by sport [3]. Note that mean SAT scores have been rounded to the nearest integer value.

Women's crew is once again the "brightest" team, and football sees itself ranking the lowest among average SAT scores; a pattern is developing between the two teams. However, this table is markedly different from the averages of grades because it is no longer systematically divided

[3] It's worth mentioning that men's water polo has a very high average SAT score, perhaps their sport attracts naturally bright individuals. Also observe that men's rugby has a minimum SAT score that is significantly higher than all other teams; perhaps playing the game at a high level requires a high "ability" for all players, not just on average. Of course, this is all subject to sampling variation. It would be interesting to see how these statistics compare with other samples from different schools.

Table 4.4: SAT Scores by Sport

Sport	SAT Mean	Std. Dev.	Min.	Max.
Women's Crew	1,225	141.9889	830	1,530
Men's Water Polo	1,204	136.6983	870	1,500
Men's Crew	1,155	141.9889	830	1,560
Men's Track CC	1,154	156.1185	760	1500
Men's Rugby	1,153	142.6184	1010	1,490
Women's Water Polo	1,147	131.3668	780	1,380
Men's Baseball	1,143	139.4782	860	1,440
Women's Track CC	1,105	176.6498	760	1,470
Men's Basketball	1,041	181.5500	710	1,480
Football	1,007	152.1244	420	1420

by gender.

It is not particularly surprising that the teams with the most extreme average academic performances also represent the most extreme cases in average academic ability, as measured by SAT scores.

These differences not only exist in expectation, but also persist throughout the distribution of SAT scores. Figure 4.5 shows that football is disproportionately comprised of students who have scored in the lowest SAT quantile, and women's crew is disproportionately comprised of students who have scored in the higher SAT quantiles.

There are 94 unique individuals who played football in college and also scored within the lowest quantile of the SAT distribution, contrasted with only 12 football players who scored in the highest quantile of the SAT distribution over the same time period. On the other hand, only 12 members of the women's crew team scored within the bottom SAT quantile, compared with 79 who scored in the highest SAT quantile. The mode SAT quantile is 1 for football players and 5 for women's crew members.

These results motivate why we should condition on quantiles rather than relying on mean averages to create predictions. SAT scores on average for the two groups only differ by approximately 200 points, but taking a look at the distribution realizes that the trend is more exaggerated than we may have anticipated.

It would appear that systematically, the lowest ability and the lowest performing students are represented by football, and that the highest ability and highest performing student-athletes are represented by the women's crew team. We previously noted that females tend to achieve higher grades relative to males, but that their observable ability is not necessarily higher on

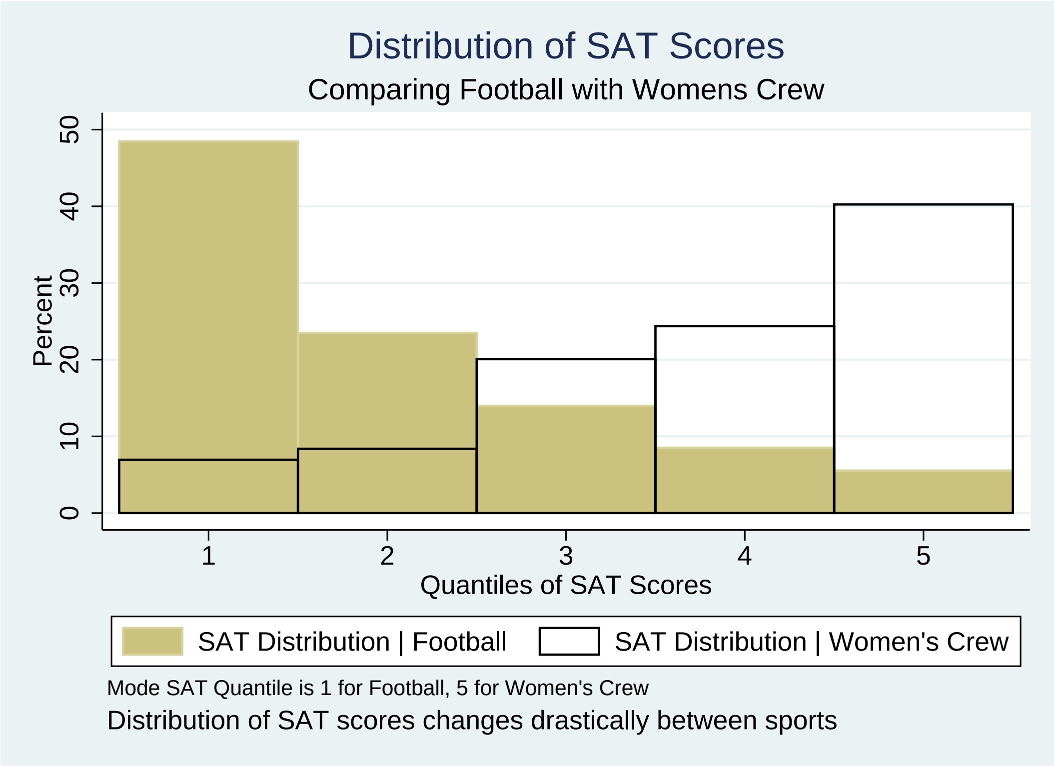

Figure 4.5: Distribution of SAT Scores for Football and Women's Crew

average. These results can be difficult to interpret: are the observed differences between teams confounded by gender?

Holding Gender Constant

A fairer comparison may be made by examining differences between male or female teams, holding gender constant. Comparing grades of football players with members of the men's crew team is interesting, because we are examining the male counterpart to a team which we previously examined.

Another reason why these two teams are chosen for comparison is because they are the only two men's sports for which there exists a sufficient support density to run a conditional quantile regression analysis. In at least one SAT quantile, football and men's crew have at least 38 points of support, which leaves a sufficient number of degrees of freedom in our later regressions. Seeing how our analysis changes for these two particular sports after conditioning on quantiles of observed ability later on is interesting, and so we will therefore restrict this preliminary analysis to these sports.

The fact that these are the only two sports which have a sufficient number of students taking SAT to run the regressions may be due to sampling variability[4]. Our sample of men's crew members yields a distribution of SAT scores indicated in figure 4.6.

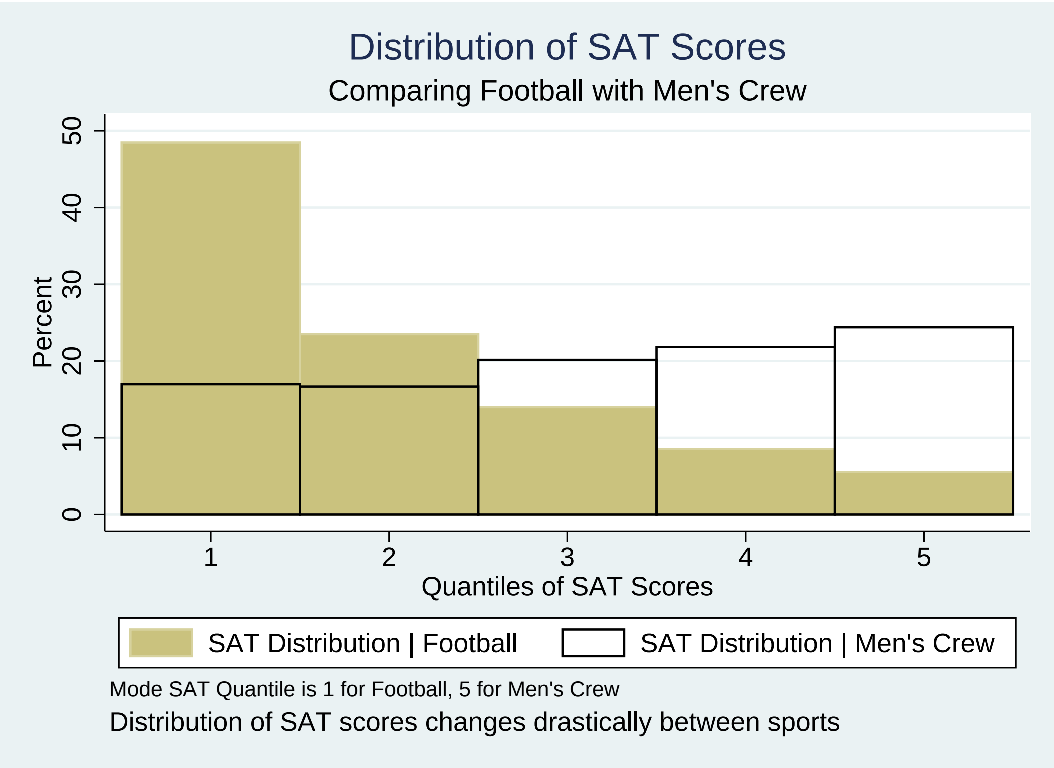

Figure 4.6: Distribution of SAT Scores for Football and Men's Crew

The differences in distributions still persist and follow a similar pattern as laid out by the women's crew team. Again, men's crew has the highest proportion of team members achieving the top quantile of SAT scores. Maybe the difference we saw earlier was partially related to differences in gender, but it appears part of it may be explained by team affiliation. There are systematic differences between the academic ability levels of football players and crew members.

In terms of academic achievement, there does not appear to be much discrepancy between fitted values for the two groups [5]. Estimates for predicted grade point average are well contained within each group's confidence limits.

This simple regression still omits observed ability from the equation. It is possible there may be a larger predicted difference in grade point average between the two groups conditional on belonging to a particular SAT quantile, but this is obscured when we use a simple mean regression. [6]

[4]In fact, additional SAT scores are available in a supplementary data set. However, appending this to our original data set was a task far beyond the scope of an undergraduate thesis.

[5]This simple regression models the equation: $GPA = Sport + (LaggedGPA) + (TwiceLaggedGPA) + (FinancialAidIndicator) + (AcademicYear)$ using a fixed effects model with robust standard errors.

[6]Unfortunately, although men's crew has 38 points of support within the highest SAT quantile, there is

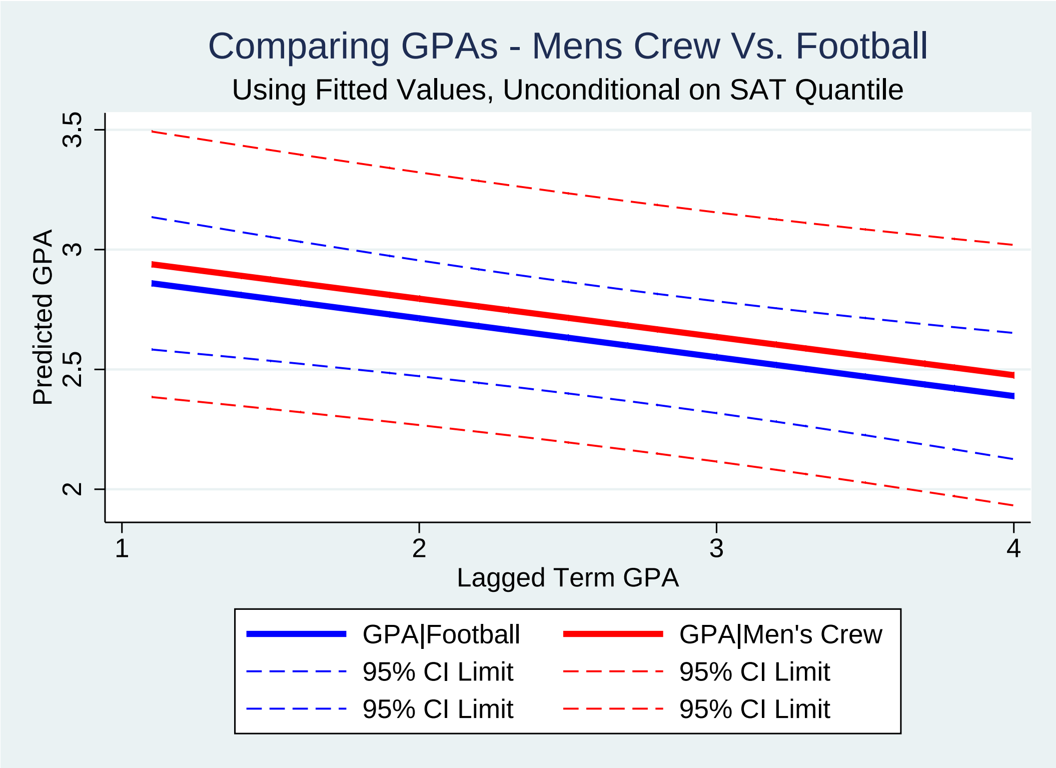

Figure 4.7: Fitted Values for Football and Men's Crew

Simply looking at averages of fitted values from basic regressions does not uncover a causal relationship because important variables have been omitted in each of the previous analyses. Furthermore, the analyses above do not consistently produce statistically different predictions of academic achievement or ability between groups; we seek a better model.

4.4 Motivating Conditional Quantiles

By conditioning on more information, we can get better predictions of term grade point averages. If our model is able to better explain grades, then it is also more likely to be able to parse out the effect of sports participation on grades. Figure 4.8 shows how our predictions for grade point averages drastically differ after conditioning on both team affiliation and ability [7]:

Our model used to predict grade point averages yields significantly different predictions when we condition on either being a football player in the lowest quantile of SAT scores, or on being a women's crew member in the second highest quantile of SAT scores. Figure 4.8 motivates

insufficient variation within our sport indicator variable of interest. Because the sport indicator variable is omitted due to problems of multicollinearity, I have omitted a simple regression analysis; men's crew is no longer a cohort of interest.

[7] The simple regression models the equation: $GPA = Sport + (LaggedGPA) + (TwiceLaggedGPA) + (FinancialAidIndicator) + (AcademicYear)$ using a fixed effects model with robust standard errors.

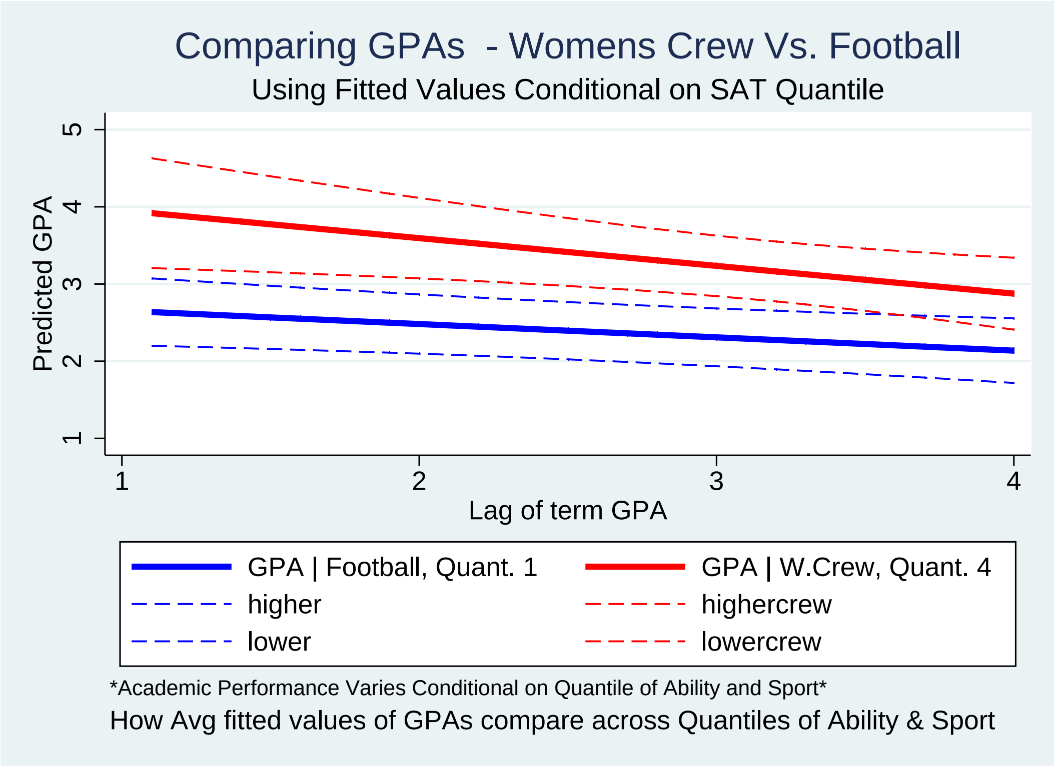

Figure 4.8: Fitted Values - Football and Women's Crew, Conditioning on Ability

the use of conditional quantiles: notice that our predictions are very statistically significant from each other [8], and that our fitted values have smaller standard errors where there is likely to be a clumping of observations in our data set. Reducing the variability in estimates is important in fitting a model that can explain grades and parse out the effect of sports participation.

Preliminary Data Analysis Conclusion

Football players in our data set tend achieve lower SAT scores and grades relative to women's crew members. But we have not answered our question of interest. How does sports participation effect grades, for either a football player in the lowest ability quantile or a women's crew member in the second highest ability quantile?

[8]Formally, we want a test statistic that relates both populations of interest, whose associated p-value indicates the probability of observing a difference as large as we did, conditional on the assumption that the fitted values should actually be identical between the two groups. Using a heuristic in this preliminary analysis is justified because we only want to get a better feel for our data.

Chapter 5

Results and Methodology

This study seeks to exploit quasi-experiments in order to identify a causal relationship between sports participation and academic achievement. We may worry that student-athletes are systematically different from their peers in ways that affect academic performance. Perhaps these characteristics are either unobservable or not measurable (i.e. dedication or leadership). We want to steer clear of direct comparisons between student-athletes and non student-athletes. Instead, we will pay close attention to cases in which a student exogenously transitions into or out of athletics, such that we can hold individual characteristics constant and examine only the effect of sports participation on grade point averages.

A naturally recurring experiment satisfies our assumption of *selection on observables*: sports participation is random and we can determine the causal effect of participation on grade point averages; we observe the grade point average of the same individual as a student-athlete, and not as a student-athlete, and learn precisely the *average treatment effect on the treated.*

5.1 Method

Imagining cases where a person is similar in every possible way except sports participation is a mental exercise. Our data set consists of only student athletes, and so we want our sports indicator variable to equal 0 whenever a student athlete is not currently participating in sports.

An example that most readily comes to mind is the case of an athlete who takes a "red shirt" year. NCAA regulations limit the number of years an individual may participate in college athletics. In the case of a red shirt year, the athlete is not allowed to compete for an entire year in exchange for an additional year of competition eligibility at a later date; this is often done for freshman or injured athletes who have hopes of competing with greater success in the future. Including students who take a red shirt year in our sports indicator variable, and flagging each instance as a 0 value, is problematic because in reality expectations from the athlete themselves, their coaches and their peers remains unchanged[1]. Athletes may even train harder during a red shirt year to make up for missed opportunities.

It is also tempting to think that grades may be compared between individuals when they are and are not in the midst of their competition season. This is also problematic for the same reasons: training patterns and expectations remain constant throughout the year, and coaches often times strongly encourage what are technically considered "club" team practices throughout the year; these practices often consist of exactly the same students, coaches, and training facilities and labeling the practice as "optional" is a formality exercise.

For an entire year, I naively assumed the Swim and Dive team had a year round season because competitions took place regularly. Distinguishing on and off seasons is difficult when observing only the amount of time commitment; the phrase "off season" is of little practical importance to most collegiate athletes.

Another idea is to compare individuals from teams that were cut midway through their college career. The University of California at Berkeley threatened to cut funding from several sports teams in 2011, and had this plan been followed through with we would have observed yet another quasi-experiment that could be utilized in our model [2].

5.1.1 Finding Quasi-Experiments

There are several naturally recurring instances of athletes who are placed into and out of athletics somewhat randomly, which are suitable for our analysis. Cases of interest include a student who walks on to an athletic team mid-way through their college career, or a student-athlete who is injured mid-way through their college career and is forced to no longer participate in sports. In both cases, we are gifted with a "before and after" comparison of semester GPAs which hold nearly everything else constant except for sports participation.

Another interesting case is for students who study abroad during their college career, because they are forced to be pulled away from their usual regimented training patterns. A skeptic may suggest that students who study abroad may be simultaneously training for the Olympics. Our data set describes instances in which student-athletes have traveled to the Olympics, and so this specific concern does not apply. It is possible, although perhaps less likely, that an individual who studied abroad within our data set adhered to the same training schedule as was required during home practices.

[1] I myself was flagged as having a red shirt year my first year in community college. I did not play sports at the time, but the athletics department issued a retroactive red shirt year such that as a transfer student I could get an extra year of eligibility.

[2] Fortunately for the teams in question, alumni contributions were successful in reinstating the programs before they were even officially cut.

5.2 Model Selection

We will utilize a fixed effects model because we assume that unobserved ability remains constant within each individual over time; this allows us to calculate the true effect of sports on grades. Our estimators are known as "within" estimators because they use variation within each individual panel to explain term grade point averages. By realizing somewhat naturally occurring randomized trials, we can better hope to satisfy our exogeneity assumption necessary for unbiased estimators.

The basic idea is that because our cases of interest have occurred "randomly", conditioning on observable characteristics do not help us to predict if an individual will participate in sports during a given semester. The quasi-experiments are seemingly random, and therefore sports participation is quasi-randomized. This assumption helps us to claim a causal relationship between sports and grades.

5.2.1 Mathematical Formulation

To estimate the causal effect of sports participation on academic achievement, we estimate the following equation:

$$\mathbb{E}\Big[GPA_{i,t}\Big|\mathbf{X_{i,t}}\Big] = \alpha + \beta S + \gamma \mathbf{X_{i,t}} \tag{5.2.1}$$

for all i=1,...,N and t=1,...,T

N is the number of student-athletes within our data set. T is the number of time periods observed in each panel. $GPA_{i,t}$ is the term GPA for person i at time t. S is a sport indicator variable, taking on a value of 0 if the individual is not participating in sports and a value of 1 if the individual is currently participating in sports. $\mathbf{X_{i,t}}$ is a vector of characteristics which includes lagged term grade point average, twice lagged term grade point average, an indicator variable indicating if the individual is on financial aid through the athletic department, and a variable that tracks time. The lags of grade point averages are beneficial because in addition to SAT scores, they help to proxy for ability.

Our coefficient of interest is β, and measures the average effect of participating in a division I athletic program at Cal on the semester grade point average for individual i at time t.

We can refine our predictions by conditioning equation 5.2.1 on being in a particular sport and gender, as well as a SAT quantile. Restricting our regression to a particular cohort is algebraically similar to the estimators we would obtain by adding a dummy variable for each sport and SAT quantile, which are then interacted with every other explanatory variable in our model.

By conditioning on observed ability, we can better estimate the effect of participating in sports for each cohort of individuals. We've discussed how belonging to a particular team can affect practice, travel, and time commitments, as well how team affiliation is correlated with academic ability. Running a simplified regression analysis as was done previously omits important variables, biases our results, and leaves us with statistically insignificant results. The hope is that conditioning on more information in conjunction with utilizing quasi-experiments will lead to more efficient estimates of the marginal effect of participation in sports.

If the coefficient of interest, β, is statistically significant, then this implies that participating in sports effects grades. In this case, a change in athletic status changes our prediction of future grades. The magnitude of the effect is determined by the magnitude of the coefficient.

If the sports indicator variable is statistically insignificant, then this implies that student-athletes would earn the same grades regardless of participating in sports. Based on the discrepancies in aggregate level data, this may negatively imply that student-athletes are perhaps systematically less academically intelligent than their non-athletic counterparts.

5.3 Results

5.3.1 Limitations: Cohort Support Density

Before regressing 5.2.1, we must ensure we have a sufficient density of support for each cohort such that we can get efficient estimators. Standard practice assumes nothing about the distribution of error terms, but instead relies on asymptotic results to obtain valid inference tests. If there is insufficient support density for a particular cohort, given the possibility our errors are not normally distributed, relying on asymptotic results may not be wise.

There are 26 sports contained within our data set. Conditioning on 5 quantiles of SAT scores within each sport yields 130 different cohorts of interest. Unfortunately, our data set only includes SAT scores for 1,750 students. If SAT scores were uniformly distributed within each quantile and within each sport, each cohort would contain approximately 14 points of support. Even despite the simplicity of our model which is careful to preserve degrees of freedom through a select few explanatory variables, multivariate regression analysis requires a great deal more points of support before a least squares regression can yield efficient estimators.

The issue of observed SAT scores being unevenly distributed becomes helpful in this regard. We previously observed that SAT scores exhibit clustering within certain cohorts, which helps to add more power to our tests for the cohorts whose regressions we are able to perform. The hope is that despite not having a lot of cohorts to compare, our estimates within these groups will be fairly precise.

The following cohorts have a sufficient support density to benefit from a conditional quantile regression: football players in the lowest two quartiles and women's crew members in the highest two quartiles. This is interesting because we happen to deal with the two extreme cases discovered in our preliminary data analysis: male football players who systematically perform worse and have a lower observed academic ability, as measured by SAT scores, and female crew members who systematically perform better scholastically and have a higher observed ability level.

Regressing 5.2.1 conditional on football players belonging to the first or second quantile of SAT scores, and women's crew members who belong to the 4th and 5th quantile, leaves us with at least 32 degrees of freedom. Because our data is unbalanced, degrees of freedom are uniquely calculated. Suppose T_i represents the number of time periods that a cross sectional individual, i, is observed. The total number of observations and therefore degrees of freedom is given by $T_1 + T_2 + ... + T_N$ (Wooldridge 2009). This model essentially uses first differences to remove fixed effects constant over time, and so we lose one degree of freedom for each observation in the process.

5.3.2 Significant Models

Regressing 5.2.1 for cohorts with sufficient degrees of freedom yields the results listed in table 5.1[3]:

Table 5.1 shows that the effect of sports participation is only significant in the 1st quantile of SAT scores for football players and the fourth quantile of SAT scores for the members of women's crew. Our predictions for grade point average change considerably depending on which cohort we condition on. Inspecting these estimates closer yields an interesting analysis.

The formal interpretation of our coefficient of interest is as follows. Conditional on playing football and having an observable ability in the lowest quantile, participating in sports is expected to increase grade point average by approximately 0.77 grade points holding the prior two semester grade point averages, financial aid status, academic year, and other unobservable factors which are time invariant constant. This is a tremendous positive difference.

Conditional on belonging to the women's crew team and scoring in the 4th quantile of SAT scores, participating in Division I sports at Cal is expected to decrease term grade point average by about 0.46 grade points, holding the prior two semester grade point averages, financial aid status, academic year, and other unobservable factors which are time invariant constant. This is

[3]Again, note that we exclude the cohort of men's crew members who are in the 4th quantile of SAT scores, because there is insufficient variation within the sports indicator variable to derive an estimator for the variable of interest. This yields an uninteresting analysis and has therefore been omitted from the regression results table

Table 5.1: Regression Results for Cohorts with Sufficient Support Density

	Fball, Q = 1	Fball, Q = 2	W Crew, Q = 4	W Crew, Q = 5
	b/se/p	b/se/p	b/se/p	b/se/p
Sport Indicator	0.770***	-0.155	-0.456***	-0.213
	(0.08)	(0.12)	(0.15)	(0.28)
	0.000	0.188	0.004	0.451
Lag of term GPA	-0.172**	-0.268*	-0.360**	-0.300***
	(0.07)	(0.15)	(0.14)	(0.10)
	0.013	0.075	0.017	0.004
2nd Lag term GPA	-0.192***	-0.110	-0.191*	-0.164*
	(0.06)	(0.11)	(0.11)	(0.09)
	0.002	0.321	0.079	0.067
Fin. Aid Indicator	-0.371**	0.023	.	.
	(0.16)	(0.34)	.	.
	0.024	0.945	.	.
Academic Year	-0.025	-0.027	0.083*	0.151***
	(0.03)	(0.06)	(0.04)	(0.04)
	0.452	0.630	0.060	0.001
Constant	53.755	58.691	-160.261*	-297.535***
	(67.67)	(113.03)	(85.38)	(86.50)
	0.429	0.606	0.069	0.001
Observations	417	196	116	191
No. of Individuals	85	41	37	58
Overall-R^2	0.0337	0.0524	0.1447	0.0709
R^2	0.0811	0.0590	0.1787	0.1714
F-Test	.	1.3763	6.5124	5.0952

*

p<0.10, ** p<0.05, *** p<0.01

quite a significant decrease, implying that grades could be expected to go from a plus to minus as a result of playing sports.

Figure 5.1 takes a graphical look at how sports participation causes a change in grade point averages within these two cohorts:

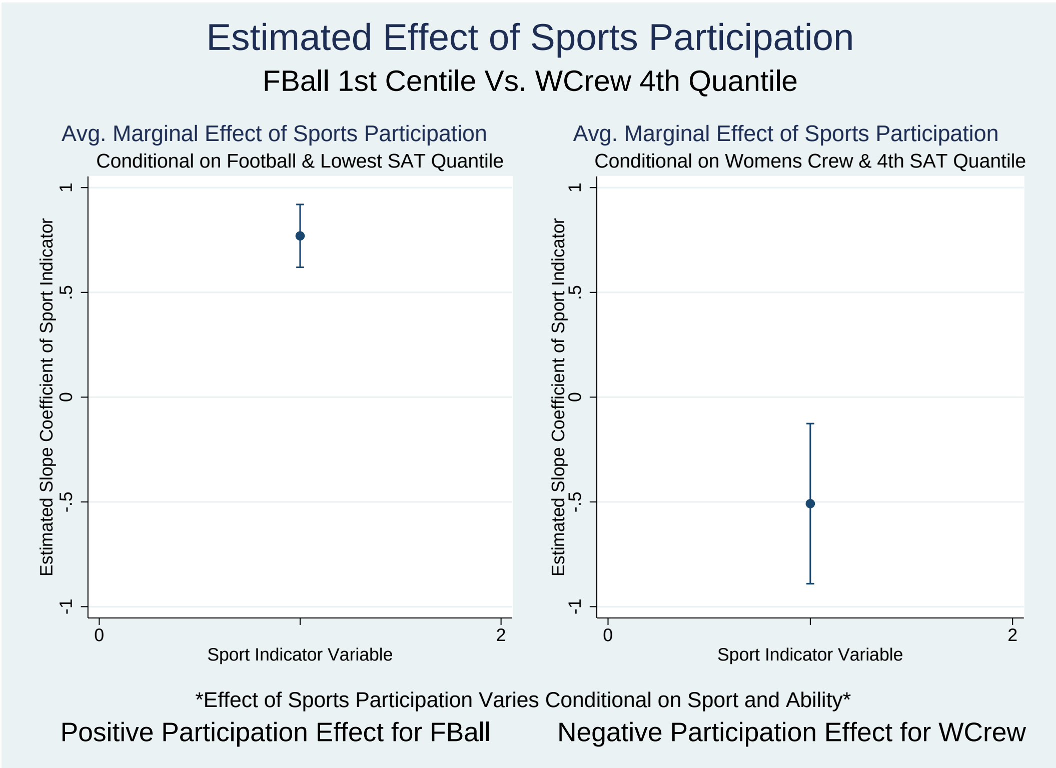

Figure 5.1: Marginal Effect of Sport Participation

Both of these estimators are not only statistically significant, but practically significant as well. The difference between a half grade point on average can make or break quite a few opportunities.

The other explanatory variable that is consistently significant between these two cohorts is the lagged grade point average. The second lag of semester grade point averages, the academic year variable, and our constant are sometimes significant depending on which model we are examining. However, we must be careful when interpreting these coefficients. Although it is tempting to develop parsimonious interpretation for each coefficient, we were not careful to create quasi-experiments for any of the other variables in our data set, and they are therefore subject to bias.

5.4 Robustness Checks

The model assumes fixed effects, and does account for heteroskedasticity within the error term through use of robust standard errors. Our model also accounts for the possibility of serial correlation within error terms. Testing for serial correlation and correcting it, given that our data is unbalanced, is actually quite complex because we must compare several more models. Preliminary testing in our model showed that our error terms likely suffer from serial correlation, and so our test statistics may not be accurate. Fortunately, our estimators are still consistent even if our error terms are not independently and identically distributed.

Our analysis also assumes academic ability is held constant within each person, but if other unobservable factors change our fixed effects model will fail to account for this. Perhaps it is realistic to think that an unobservable characteristic that confound grades could possibly change within an individual over the course of time. For example, a traumatic family experience, or health concern, may force an individual away from their studies, thereby hurting academic performance despite their innate ability. This kind of unobservable factor is not contained within our model, and it may be correlated with lagged grade point average, in which case it violates our exogeneity assumption which is necessary for best linear unbiased estimators.

In the case of an omitted variable as described above, our estimators would still in fact be consistent because we assumed random selection. Even if unobservable factors are changing within our sample, sports participation is randomly assigned due to the use of quasi-experiments, and the unobservable factors are not systematically tied to treatment or control group.

We can say with confidence that participating in sports does have a statistically significant effect on term grade point average for football players in the lowest SAT quartile, and for women's crew members within the 4th SAT quartile within our sample.

Discussion

Conditional on belonging to a particular cohort, there are either costs or benefits associated with sports participation, as measured by term grade point averages.

These results have an intuitive appeal that can be explained in laymen's terms. Ignoring sports team affiliation for a moment, consider only SAT scores. Suppose an individual has a high observed academic ability. They possess the potential to achieve a high grade point average. However, participation in sports is likely to cause a significant drain on time, energy, and accessibility to other resources. We might expect that an individual may not be able to achieve as high grades as otherwise predicted, because they are not employing their mental capacity to its fullest. Sports participation directly causes individuals to have fewer resources available to them, which in turn causes grades to decline. Sports participation can therefore cause academic performance to decline among bright-minded individuals.

Consider another individual who is ranked low in observed academic ability, as measured by SAT scores. Perhaps this individual has walked through adolescence with little structure in their lives or time management skills. In this unique case, participating in sport may promote goal setting, efficient use of time, and a sense of purpose. Through whatever means, said individual is pushed to perform better academically. Sports participation in this case causes the individual to be endowed with positive attributes, which help them in school. Sports participation in this case might cause academic performance to increase.

We may also be able to explain some of the variation in marginal effects through the required minimum grades necessary for sports participation, as well as tutoring benefits available to student athletes. Consider a football player with a low academic ability and a history of deficient academic success. Said individual may be forced to study harder than they did previously because in order to be eligible for sports participation at Cal, athletes are required to maintain a minimum grade point average. If the individual would normally achieve a grade point average below this threshold, participating in sports may force them to study harder. Furthermore, academic advising and tutorial services are provided at no cost to student-athletes, and are strongly encouraged or even required for individuals who are on academic-probation or at risk of academic failure. It's important to note that being forced to maintain a minimum grade point average or visit a tutor are direct consequences of sports participation, and so we can still conclude that sports participation is the cause of the increase in expected semester grade point average.

On the other hand, perhaps a women's crew player is taking very difficult courses, for which there are no tutors available. Their level of academic performance has a nearby upper bound, and it is not always possible for their grades to increase. In this case, we might theorize that as a direct result of sports participation, the crew member has less time available to study and so cannot perform as well on exams. If the member did not participate in crew, they would have more resources available to better succeed in schooling.

The implications of the preceding analysis are noteworthy at personal, collegiate, economic, and societal levels.

An incoming collegiate athlete may be empowered with this information to better optimize their actions in accordance to preferences. Perhaps an individual who is prospectively thinking about women's crew and has scored well on an SAT examination would rather maintain grades on average than participate in athletics. Losing a half grade point on average is practically significant, and can affect chances of future employment or acceptance to graduate school; these factors may push the individual to reconsider sports all together.

On the other hand, an incoming football player who has not scored well on the SAT examination may consider pursuing a collegiate athletic career to increase their expected grade point average. Of course there exists outliers, and our results and analysis are subject to sampling variation. However, inference testing based on this sample leads one to believe the benefits of sport participation are potentially large among football players with a low academic ability.

If a college ever chooses to conduct a cost benefit analysis of having sports programs in general, or a particular team, conditional quantile regression analysis serves useful. Perhaps a school may choose to financially support a team on the basis of how it is expected to affect the academic performance of participating students.

These findings are important to the economy because employers may potentially use term GPA and therefore cumulative grade point averages as a signal of labor market potential. Status quo, an employer may misallocate resources by mistakenly hiring an individual with a higher level of observed academic achievement, when in reality there exists a more qualified candidate. Understanding the effects of sports participation for all possible cohorts would help increase the efficiency at which employers are able to use grades as a signal.

Even for society as whole, the findings of this analysis are provocative. Given that there exists a different sports effect conditional on belonging to a particular cohort of individuals, it is not reasonable to create stereotypes that apply to all cohorts. The results of this study do not speak to the overall intelligence of athletes and non-athletes, but they do indicate that participating in sports effects different cohorts differently. Contrary to popular belief, having some football players may find their grades will improve by playing in college.

Chapter 6

Conclusion

The findings presented in this analysis pertain strictly to our sample of student-athletes attending the University of California at Berkeley between 1999 and 2012. It is fallacious reasoning to necessarily assume that statistically significant coefficients in one sample will still be significant in another sample. Therefore, we cannot infer anything about the causal nature of sports participation on grades within other samples.

That said, this sample size was reasonable and did allow us to perform conditional quantile regression analysis for several cohorts of interest. There is an observed difference in the effect of sports participation conditional on observable characteristics such as sport and SAT quantile. Our analysis used quasi-experiments to ensure *selection on observables* or *exogenous assignment of treatment*, and we therefore obtained unbiased estimators.

This analysis shows that within our sample, football players in the lowest SAT quartile face a positive incentive for participating in sports when considering future grade point averages. Their term grade point average is expected to increase by 0.770 grade points by participating in sports. Unfortunately for women's crew members, participating in sports is expected to decrease their future grades on average and therefore imposes a cost. Women's crew members in the fourth quantile of SAT scores can expect their term grade point average to decrease by 0.456 grade points if they participate in sports.

A couple of issues are worth being examined further. It appears that women's crew is a really smart group of individuals, and aggregate data supports a contrary argument with respect to football players. We also saw that men's rugby posses a high minimum ability level relative to other sports, and that men's water polo also poses a high average ability level relative to other sports. It would be interesting to examine these same sports in other samples to see if the various phenomenon observed in this sample continue to persist.

As the model and data set currently stand, there is still plenty of room for improvement. More analysis can be done on testing for serial correlation within the error terms, and it's quite possible there are other models that better fit our data. Even if there exists no serial correlation, perhaps our analysis would be more interesting if we allowed for a non-parametric functional form of our model. Our variable of interest is an indicator variable, but our data is still fit using lagged grade point averages, a practically continuous variable which likely does not have a constant marginal effect on future grades. Fitting a non-parametric model may help to better explain term grade point averages and therefore more precisely estimate the marginal effect of participating in sports.

If we were privileged with a larger data set consisting of the same variables, this analysis could be further enhanced. Our data set strictly pertains to student-athletes, and it may be interesting to see how the efficiency of our estimators would change if we could compare our data set to the general student body population at the University of California, Berkeley during the same time period. If we could increase the sample size either through chronological expansion, tracking students for a longer period of time, or geographical expansion, comparing across schools, this would help to make our analysis more precise. Tracking individuals after they are cut from the team, or collecting data on past grades for walk-ons can also give our inference more power and efficiency. Simply having no missing observations for SAT scores would also assist in examining the remaining cohorts. In this case, we may be able to learn more about cohorts we were not able to consider in this data set due to lack of support density.

If we were to obtain additional variables for our existing data set, our results could also be further improved. Consider the use of parental income, parental education level, high school achievement level, and other demographic characteristics as additional explanatory variables that could better explain our data. Even though demographic variables are time invariant, we could still include them in our model through interactions with time-variant variables. It would also be useful to approximate the difficulty in course load that a student faces such that we may weight academic achievement relative to work-load in some fashion. Something as simple as including whether a student belongs to the College of Letters and Science versus the College of Engineering can help to better explain grades.

Also consider the potential room for improvement with an instrumental variable analysis. Instruments could be use to better approximate the effect of sports participation without worrying about making assumptions such as *selection on observables*. Also consider the possibility for using an instrumental variable for the number of units a student is enrolled in. As it is, units may be endogenously determined because student-athletes tend to take lighter class loads to accommodate training schedules. Finding a way to include units in our model could also be of use.

Although our analysis is not perfect, it does help to shed some light on the causal effect of sports participation on grades, which is positive for football players in the lowest quantile of

SAT scores and negative for women's crew members in the fourth quantile of SAT scores.

In expectation, the term grade point average for football players with the lowest academic ability will increase by 0.770 grade points if they participate in sports. For women's crew members with a very high academic ability, their term grade point average is expected to decrease by 0.456 grade points if they participate in sports. Both of these estimators account for the prior two semester grade point averages, financial aid status, academic year, and other unobservable factors which are time-invariant. The average treatment effect on the treated in these groups is both statistically and practically significant. Among other cohorts of student-athletes, the density of support is insufficient and we cannot estimate the marginal effect of sports participation on term grade point averages. Further studies are required to learn how other cohorts are affected.

References

Card, David and Sullivan, Daniel. “Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment”, *Econometrica*, 56. 3 (1988): 497-530

Graham, Brian. “Proxy Variables and Panel Data” Lecture. Applied Econometrics Lecture. University of California at Berkeley, Berkeley. 05 Apr. 2012 Web.

Heckman, James, and Richard Robb. “Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes.” *Drawing Inferences from Self-Selected Samples*, (1986): 63-107 JSTOR

Heckman, James J., and Jeffrey A. Smith. “Assessing the Case for Social Experiments” *The Journal of Economic Perspectives*, 9. 2 (1995): 85-110. JSTOR

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score.” *Econometrica*, 71. 4 (2003): 1161-189

Imbens, Guido W., and Joshua D. Angrist. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica*, 62. 2 (1994): 467-75 JSTOR

Josephs, Ray. “Time to be Candid about 20-hour Rule.” NCAA News Archive. NCAA, 1 Jan. 2006 Web 7 May 2012

Levitt, Steven D, and Stephen J. Dubner. Freakonomics: A Rogue Economist Explores the Hidden Side of Everything. New York: Williams Morrow, (2005). Print.

Rheenen, Derek V. ASC: Academic Performance Summary 2011. Athletic Study Center. University of California at Berkeley, 2011.

Wooldridge, Jeffrey M. Introductory Econometrics: A Modern Approach E ed. Vol. 4. Mason: South-Western Cengage Learning, (2006): 488