Lizzie Healy
Professor Preston
Introduction to Econometrics
8 December 2022

Salaries and Sports:
A Look into the Financial Effect of College Athletics

With just about half a million college students playing sports each year, it is no wonder there has been recent research on the reasons and benefits of this athletic participation. Previous research has indicated the physical, mental, and cognitive benefits, yet examining fiscal benefits is a relatively new area of study. Thus, in this research paper, I will focus on the monetary returns and aim to answer the following research question: Does participation in college sports impact salaries? Specifically, I will also investigate this in terms of varsity versus club sports and whether this impact is realized in the same way for both men and women. The dataset I will be using is a person-year dataset collected on labor and personal statistics of college students from four colleges that began during undergraduate and ended five years post-graduation taking place between 2012 and 2019. The data consists of survey answers that asked the respondents about personal information, their education, their job and salary, and information about their college sport if they participated. This information was collected at the same time, however, the datasets of college sports information needed to be merged with the original dataset.

In order to determine an initial regression, I first used residuals and fitted variable graphs to determine which specification would best suit the variable. From this I found that a log-log regression was best suited due to the cloud-like appearance of the scatter plot as well as having the highest R^2 value, indicating this regression had an independent variable that explained the most amount of variance in the dependent variable. Thus, my dependent variable is the natural

log of yearly salary plus a bonus of the respondent, measured in thousands of dollars. The independent variables are *lncolsport1_hrs,* the log of the hours spent per week playing a college sport, *varsity* and *otherlevel*, which are dummy variables that indicate whether the respondent played a varsity sport in college or junior varsity/club. The equation will be run as follows: $lnyearsalbon = \beta_0 + \beta_1 lncolsport1\ hrs + \beta_2 varsity + \beta_3 otherlevel + e_i$. The e_i indicates the error term. The equation will be clustered by id to correct for heteroscedasticity, autocorrelation, and the possibility of multiple observations for one individual as a result of a person playing two or more sports in college. The coefficient of interest here is that of *varsity*, which is statistically significant at the 5% level of significance, and can be interpreted as holding hours of sport played constant, varsity college players will earn 51.26% less, on average, yearly salary as compared to non-college athletes. With this equation, we also tested the null hypothesis that the coefficient on *varsity* is equal to the coefficient on *otherlevel,* which could not be rejected at the 5% significance level, thus we can not reject the null hypothesis that these two coefficients have the same effect on salary.

To investigate further whether the effect of participating in a college sport is different for men versus women and how it differs by level, I added variables to the initial regression as such $lnyearsalbon = \beta_0 + \beta_1 lncolsport1\ hrs + \beta_2 varsity + \beta_3 otherlevels + \beta_4 female + \beta_5 femalevarsity + \beta_6 femaleotherlevel + e_i$. In this equation, I added the independent variables, *female,* which is a dummy variable that indicates whether a person is a female, and interactions terms between *female* and *varsity* and *otherlevel.* The coefficients of interest are *varsity* and *femalevarsity*. The coefficient on *varsity* decreased, meaning that holding all else

equal, being a varsity athlete decreases an individual's salary by 68.66%, on average, as compared to non-athletes. The other coefficient of interest can be interpreted as holding all else equal, being a female varsity athlete as compared to a male varsity athlete in college increases yearly salary by 45.86%, on average. This effect was not statistically significant, however, the effect of being a varsity athlete was statistically significant. The null hypothesis that the coefficient on *varsity* and *otherlevel* has the same effect on salary as well as the null hypothesis that *femalevarsity* and *femaleotherlevel* have the same effect on salary were both tested. Both of these null hypotheses failed to be rejected.

To further control for possible error, fixed effects were added to the regression. The regression is thus as follows: $lnyearsalbon = \beta_0 + \beta_1 lncolsport1\ hrs + \beta_2 varsity + \beta_3 otherlevels + \beta_4 female + \beta_5 femalevarsity + \beta_6 femaleotherlevels + \alpha_i + u_t + e_{it}$. The final three terms represent the error, where α_i is the error associated with entity fixed effects which is the sector variable that is constant across time, but different for each observation. The term u_t is the error associated with time-fixed effects which are used for the year of survey variable as well as the number of months worked variable which are constant for each individual but varies across time. The coefficient of interest leads to the interpretation that holding all else constant including the year, the sector, and the number of months worked, playing a varsity sport as opposed to someone who doesn't, decreases a person's salary by 64.85%, on average. Similarly, holding all else constant including year, sector, and months, a female that plays a varsity college sport as opposed to a man who also does, will earn 62.15% more salary, on average. Both of these effects are significant at the 5% significance level. We were not able to

reject the null hypothesis that *femalevarsity* and *femaleotherlevel* have the same effect on salary, however, we were able to reject the null hypothesis that each of the coefficients on each of the variables used for fixed effects are jointly equal to zero.

For the last regression fixed effects were added to control for what type of sport the respondent played in college. The regression remained the same as the previous one, however, u_t is now also associated with the sport variable which is constant across time but varies across individuals. *Varsity,* which is the coefficient of interest, yielded that holding all else equal including now the sport a person plays, a college varsity athlete as compared to a non-athlete, decreases salary by 73.11%, on average. The coefficient on *femalevarsity* increased, meaning that holding all else equal and now sport equal being a female varsity athlete as compared to a male, increases salary by 60.65% on average. The result of *varsity* is less precise as this result is now not significant at the 5% level, but only at the 10% level, while the result of *femalevarsity* is slightly more precise as it is now significant at the 10% significance level.

The results, overall, were a bit surprising. The initial regression indicated a relatively large negative effect of playing a varsity sport in college on salary. However, more controls and fixed effects were included, and the negative effect grew larger and even increased in statistical significance. This indicates that college athletics do not seem to have a benefit on salary later in life, and may actually be detrimental. In terms of the other coefficient of interest, the coefficients on females that play varsity athletics in college initially produced a coefficient of 0.459, however, this was not significant. When we added controls and fixed effects, however, there was a significant positive effect with a peak at 62.2% increase in salaries for female varsity athletes.

Both of these results indicate the importance of observing how controls and fixed effects may change a regression to see more accurate results overall. For further analysis, there is still the possibility of error and omitted variable bias, which could be investigated with more variables and data on college athletes.

Table 1 Does Participation in College Sports Impact Salary?

VARIABLES	(1) lnyearsalbon	(2) lnyearsalbon	(3) lnyearsalbon	(4) lnyearsalbon
lncolsport1_hrs	0.0689	0.0214	0.0907	0.0752
	(0.0789)	(0.0748)	(0.118)	(0.158)
varsity	-0.513*	-0.687***	-0.648***	-0.731*
	(0.261)	(0.104)	(0.227)	(0.385)
otherlevel	-0.415	-0.584***	-0.645***	-0.956**
	(0.266)	(0.0914)	(0.230)	(0.421)
female		-0.775**	-0.822***	-0.757***
		(0.327)	(0.212)	(0.276)
femalevarsity		0.459	0.622**	0.607*
		(0.349)	(0.251)	(0.350)
femaleotherlevel		0.352	0.566**	0.512
		(0.348)	(0.252)	(0.382)
Constant	11.52***	12.03***	11.45***	12.03***
	(0.332)	(0.188)	(0.418)	(0.677)
Observations	195	195	181	181
R-squared	0.019	0.121	0.600	0.707

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Works Cited:

Bowen, D. H., & Hitt, C. (2016). History and evidence show school sports help students win. *The Phi Delta Kappan*, *97*(8), 8–12. http://www.jstor.org/stable/24893327

Preston, Anne. (2019). *Laboratory Labor Data* [Data file and code book]. Anne Preston [distributor].