Oversigning; How the Phenomenon Affects Competitive Advantage.

Econometrics 485

KOURTNEY WALSH

12/10/2013
Abstract

The purpose of this study is to see if by oversigning division I college football players, teams are more likely to win more games in a season, or have a competitive advantage. Competitive advantage is known as having an advantage over your competitors allowing you to be more successful than your competitors due to a certain reason; for example oversigning players. This topic was chosen due to the recent rule changes and crack down on oversigning players. The data used ranges from 2006-2010. A regression analysis was ran to compare all the independent variables to see how they effected wins in the 2010 season. The data used was cross sectional, and it was found that oversigning players does not increase a teams competitive advantage.
**Introduction**

Recruiting is very important to Division I football teams; coaches are only allotted a certain amount of scholarships on a team per year. In recent years coaches have been oversigning significantly more. For example they are only allowed to have 25 new scholarships, but coming into the season they have signed 31 new incoming athletes to join the team. They then have to drop 6 of these players from the team before the season starts. Coaches are doing this because they can pick who they want to keep or let go of, possibly giving them a competitive advantage. The goal of this study is to compare teams that oversign to teams that do not, to see if they gain a competitive advantage. Multiple variables will also be included to determine how oversigning players affects Division I football team performance. The act of oversigning is thought to be unethical, but teams every year continue to be repeat offenders due to the fact that there is this perceived benefit that by doing so you could potentially win more games.

**Literature Review**

An important part of oversigning that a lot of people do not know about is a letter known as, National Letter of Intent (NLI). The NLI was first used so that once a student decides to go to a certain school, no other colleges can try and contact them, which this concept does still hold, but it used to also essentially hurt the players, or potentially can (Bateman2012). The reason this letter can hurt players is due to the phenomenon of oversigning. As stated earlier, coaches are only allowed to sign at the most 25 new players, but coaches are often signing 28 or 30 students. This comes back to hurt the students due to the fact that when signing the NLI, they have a binding agreement to play at this school. With coaches oversigning they have to drop down to 25 before the start of the season; thus coaches have to drop players that have signing the NLI and
are moved in and ready to practice. With this happening scholarships are taken from the players, and they are told that they have to figure out a way to pay for school, or they have to move their stuff out of the school. Some athletes are able to find different universities to transfer too, but this still turns into a big ordeal for the student.

This practice is “frowned” upon, but there are many ways around it. Teams in the division SEC are known for oversigning every year. In 2009 Ole Miss signed 37 student athlete scholarships coming into the new season, and in 2010 Auburn sign 32 new incoming players (Krohn 2011). Teams like Alabama and LSU have been given the nick name as “The Super Bowl of Oversigning” due to the amount that they are known for oversigning year after year. From 2006-2011 Alabama average of players signed per year was 27.4 per year, and LSU had an average of 25.2 per year, which places them amoung the top ten schools nationally to over sign (Bachman 2011).

Due to the number of schools increaslingly oversigning, in 2011 the NCAA set a limit of oversigning only +3, or 28 incoming athletes (NCAA 2011). This rule is often referred to as the “Houston Nutt Rule” because it was put into place after the head coach of the University of Mississippi signed 37 scholarships for the 2009 season (Engel 2011). There is no official penalty for oversigning players, so team have decreased this number, but still continue to do so.

**Theory**

The two models to be developed for this study use Percent wins for 2010 and Sagarin rank as dependent variables. Percent wins for 2010 was chosen as the dependent variable to see how dependent variables would affect wins. It is predicated that the independent variables will show if they have an affect on wins for the season. The dependent variable used for model two was the
Sagarin rank. This Sagarin rank not only takes into consideration the wins and loses of a team, but also takes into consideration the strength of the opponents and home venue advantage. The same independent variables are used to develop the second model. The independent variables are described below.

**Percentage of upperclassmen (Upperclass)**

This variable was chosen because teams that have more junior and senior players are likely to have a more successful seasons compared to a team that uses more freshmen and sophomore players. A team that has played together longer, and have had more experience playing college level football, one would predict would increase the number of wins per-season.

**Years of Coaching Experience (Coach)**

One would think years of experience would help coaches to improve skills and better understand the game and players. Thus a team with a coach that has more experience would expect to win more games.

**Average number of players signed over a 5 year span (FiveYrAvg)**

This looks at the amount of players signed each from the year 2006-2010 and averages them for the 5 year span. More players signed gives coaches more opportunity to pick better players. The number of players signed from 2006-2010 divided by 5 years results in an average number of players signed per year.

**Number of players oversigned (numoversig)**

The last variable chosen is the number of players over signed in a single year. This number needs to be taken into consideration for a single year to try and observe how it effects the
base years winning percentage, and in the future as well. It would be predicated that the more players that are oversigned would increase the amount of wins per year.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Expected Sign</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upperclass</td>
<td>+</td>
<td>.388</td>
<td>.061</td>
<td>.577</td>
<td>.248</td>
</tr>
<tr>
<td>Coach</td>
<td>+</td>
<td>26.35</td>
<td>7.892</td>
<td>51</td>
<td>10</td>
</tr>
<tr>
<td>FiveYrAvg</td>
<td>+</td>
<td>23.237</td>
<td>2.459</td>
<td>32.8</td>
<td>18.2</td>
</tr>
<tr>
<td>NumOverSig</td>
<td>+</td>
<td>23.234</td>
<td>3.969</td>
<td>34</td>
<td>10</td>
</tr>
</tbody>
</table>

**Data**

The data that was used for this model is found through ESPN, USA Today, and Oversign.com, which have been cited in the attached reference page. The total number of observations is 114 teams, over the years of 2006-2011.

**Theoretical Models:**

\[ \text{PercentWins2010} = \beta_0 + \beta_1 \text{Upperclass} + \beta_2 \text{Coach} + \beta_3 \text{FiveYrAvg} + \beta_4 \text{NumOversig2010} \]

\[ \text{SagRank} = \beta_0 + \beta_1 \text{Upperclass} + \beta_2 \text{Coach} + \beta_3 \text{FiveYrAvg} + \beta_4 \text{NumOversig2010} \]
Model 1

The results of the estimated model are reported in figure 1.

Figure 1:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.626281</td>
<td>0.261620</td>
<td>2.393860</td>
<td>0.0184</td>
</tr>
<tr>
<td>UPPERCLASS</td>
<td>-0.344222</td>
<td>0.354455</td>
<td>-0.971130</td>
<td>0.3336</td>
</tr>
<tr>
<td>COACH</td>
<td>-0.002025</td>
<td>0.002884</td>
<td>-0.702289</td>
<td>0.4840</td>
</tr>
<tr>
<td>FIVEYRAVG</td>
<td>0.003661</td>
<td>0.011196</td>
<td>0.327018</td>
<td>0.7443</td>
</tr>
<tr>
<td>NUMOVERSIG</td>
<td>0.000355</td>
<td>0.007197</td>
<td>0.049265</td>
<td>0.9608</td>
</tr>
</tbody>
</table>

R-squared: 0.015522
Mean dependent var: 0.532562
S.D. dependent var: 0.228292
S.E. of regression: 0.230632
Akaike info criterion: -0.053115
Schwarz criterion: 0.066894
Hannan-Quinn criter.: -0.004410
Durbin-Watson stat: 2.334556
Prob(F-statistic): 0.786942

Analysis & Estimation

The above model uses percent wins for 2010 as the dependent variable, and upperclassmen, years of coaching experience, five year average and number oversigned in 2010 as the independent variables. This model has an R-squared of .01, or in other words, 1% of the variation in the data in the dependent variable is explained by the independent variables. The signs are also the opposite of what was expected; upperclass, and coach all had a negative relationship with the dependent variable. These variables were expected to have positive signs. This could potentially due with the fact that when you look at the t-chart created below, the variables are not significant, causing potential problems.
The next test that was run on the first model was a t-test to look and see how significant each variable is. Below is the t-values for one tailed tests because none of my variables had unexpected signs, therefore only a one tail test is necessary.

\[
\text{T-critical n=114= 114-5 DF=109}
\]

\[
\text{One Tail}
\]

\[
.10: 1.296
\]

\[
.05: 1.671
\]

\[
.01: 2.309
\]

The above numbers are the values for a t-test. When looking at these values, you can see that all the variables used in figure 1 are not significant because none of the values are greater than 1.296. The variable that is closest to being significant would be upperclass, which would have explanatory power at the 33% level. In other words, if the number of upperclassmen increased by 1 student, this would decrease the dependent variable by .34%, holding everything else constant.

**Figure 2:**

<table>
<thead>
<tr>
<th></th>
<th>COACH</th>
<th>FIVEYRAVG</th>
<th>NUMOVERSIG</th>
<th>UPPERCLASS</th>
<th>S2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>COACH</td>
<td>1.000000</td>
<td>-0.048110</td>
<td>-0.265889</td>
<td>-0.020586</td>
<td>-0.071652</td>
</tr>
<tr>
<td>FIVEYRAVG</td>
<td>-0.048110</td>
<td>1.000000</td>
<td>0.607193</td>
<td>0.001681</td>
<td>0.046569</td>
</tr>
<tr>
<td>NUMOVERSIG</td>
<td>-0.265889</td>
<td>0.607193</td>
<td>1.000000</td>
<td>0.065716</td>
<td>0.042764</td>
</tr>
<tr>
<td>UPPERCLASS</td>
<td>-0.020586</td>
<td>0.001681</td>
<td>0.065716</td>
<td>1.000000</td>
<td>-0.090686</td>
</tr>
<tr>
<td>PERCENTWIN</td>
<td>S2010</td>
<td>-0.071652</td>
<td>0.046569</td>
<td>0.042764</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

Figure 2 is a correlation matrix that was ran to check for multi-collinearity in Model 1. A problem was found with the two variables highlighted; FiveYrAvg, and NumOverSig. Both of
these variables are measuring number of classmen that been over signed. In order to fix this problem, the variables could be combined into one, or one of them can be dropped. In this case, it was determined that NumOverSig should be the variable dropped from model 1. Though there is multi-collinearity, this model does not have autocorrelation because the data is cross-sectional.

**Figure 3:**

Dependent Variable: PERCENTWINS2010  
Method: Least Squares

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.627175</td>
<td>0.259805</td>
<td>2.414023</td>
<td>0.0174</td>
</tr>
<tr>
<td>UPPERCLASS</td>
<td>-0.342844</td>
<td>0.351744</td>
<td>-0.974698</td>
<td>0.3318</td>
</tr>
<tr>
<td>COACH</td>
<td>-0.002067</td>
<td>0.002740</td>
<td>-0.754404</td>
<td>0.4522</td>
</tr>
<tr>
<td>FIVEYRAVG</td>
<td>0.004002</td>
<td>0.008758</td>
<td>0.457010</td>
<td>0.6486</td>
</tr>
</tbody>
</table>

R-squared: 0.015501  
Mean dependent var: 0.532562  
Adjusted R-squared: -0.011349  
S.D. dependent var: 0.228292  
Akaike info criterion: -0.070636  
Schwarz criterion: 0.025371  
Hannan-Quinn criter.: -0.031673  
Durbin-Watson stat: 2.334694  
Prob(F-statistic): 0.631129

After finding that multi-collinearity existed in the model, a new model was run, leaving out the variable numoversig. In figure 4 when removing the variable numoversig, though R-squared decreased, adjusted R-squared increased, which means by taking out this variable costs out weigh the benefits, and removing this variable was a good decision.

Lastly for model 1, an f-test needed to be performed. An f-test looks at the entire model to see how it fits the data, and how good the model truly is. It can be used to compare models, which is what will be done after running tests for model 2.
F-critical (5-1) = 4 (114-5) = 109

.10: 2.04

.05: 2.53

.01: 3.65

The F-stat for figure 3 is .577. When comparing this to the f-critical values, you can see that it does not have explanatory power, and is not a good model.

Model 2

The results of the estimated model are reported in figure 4.

Figure 4:

Dependent Variable: SAGRANK
Method: Least Squares

Sample: 1 115
Included observations: 114

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>67.78549</td>
<td>13.57413</td>
<td>4.993725</td>
<td>0.0000</td>
</tr>
<tr>
<td>UPPERCLASS</td>
<td>-3.381093</td>
<td>18.39087</td>
<td>-0.183846</td>
<td>0.8545</td>
</tr>
<tr>
<td>COACH</td>
<td>-0.151387</td>
<td>0.149613</td>
<td>-0.1011858</td>
<td>0.3138</td>
</tr>
<tr>
<td>FIVEYRAVG</td>
<td>0.471269</td>
<td>0.580895</td>
<td>0.811281</td>
<td>0.4190</td>
</tr>
<tr>
<td>NUMOVERSIG</td>
<td>-0.097398</td>
<td>0.373416</td>
<td>-0.260829</td>
<td>0.7947</td>
</tr>
</tbody>
</table>

R-squared   0.016470  Mean dependent var 71.16825
Adjusted R-squared   -0.019623  S.D. dependent var 11.85064
S.E. of regression  11.96635  Akaike info criterion 7.844943
Sum squared resid   15608.09  Schwarz criterion 7.964951
Log likelihood      -442.1617  Hannan-Quinn criter. 7.893647
F-statistic         0.456326  Durbin-Watson stat 1.862077
Prob(F-statistic)   0.767594

Figure 4 is the second model which uses SagRank as the dependent variable. This model has an R-squared of .01, like model 1, 1% of the variation in the data in the dependent variable is explained by the independent variables, but the coefficients of each independent variable has changed, along with the level of significance for some variables. Two of the variables have signs
that were opposite of what was expected. As in the first model, none of the variables are significant when running a t-test.

\[
\text{T-critical n=114} = 114-5 \ DF=109
\]

**One Tail**

.10: 1.296  
.05: 1.671  
.01: 2.309

The above numbers are the t-critical values used to see if the variables are significant.

When looking at model 2, none of the variables are significant because they are all smaller than 1.226. The variable that is closest to being significant is coach, which has explanatory power at the 31% level. In other words if the years of coaching experience decreases by 1 year, it will decrease SagRank, the dependent variable by .151 points. This variable would not normally be considered significant.

**Figure 5**

<table>
<thead>
<tr>
<th></th>
<th>SAGRANK</th>
<th>UPPERCLASS</th>
<th>COACH</th>
<th>FIVEYRAVG</th>
<th>NUMOVERSIG</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAGRANK</td>
<td>1.000000</td>
<td>-0.017435</td>
<td>-0.096466</td>
<td>0.083106</td>
<td>0.052500</td>
</tr>
<tr>
<td>UPPERCLASS</td>
<td>-0.017435</td>
<td>1.000000</td>
<td>-0.020586</td>
<td>0.001681</td>
<td>0.065716</td>
</tr>
<tr>
<td>COACH</td>
<td>-0.096466</td>
<td>-0.020586</td>
<td>1.000000</td>
<td>-0.048110</td>
<td>-0.265889</td>
</tr>
<tr>
<td>FIVEYRAVG</td>
<td>0.083106</td>
<td>0.001681</td>
<td>-0.048110</td>
<td>1.000000</td>
<td>0.607193</td>
</tr>
<tr>
<td>NUMOVERSIG</td>
<td>0.052500</td>
<td>0.065716</td>
<td>-0.265889</td>
<td>0.607193</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

Figure 5 is the correlation matrix that was ran for the second model. In this model, like model 1, there is multi-collinearity in the same variables; numoversig and fiveyravg.

Numoversig will again be the variable dropped due to multi-collinearity.
Figure 6:

Due to the fact that multi-collinearity was found between two variables, Figure 6 is the new model that was ran, while dropping the variable numoversig. In this model, R-squared and adjusted R-squared both decreased, which would indicate that benefits out weigh the costs in removing this variable.

Finally in this model 2 and f-test had to be ran to see if the model has explanatory power. The same F-critical values are used as stated earlier.

\[
\text{F-critical (5-1)= 4 (114-5)= 109}
\]

\[
.10: 2.04
\]

\[
.05: 2.53
\]

\[
.01: 3.65
\]

When comparing the models f-stat (.591) to the f-critical values, you can see that this small than any of the f-critical values; therefore not containing any explanatory power.
The final model chosen was the second model, using Sagrank as the dependent variable.

Though both of the models are very poor, and essentially have no relationship or explanatory power, the second model has a higher adjusted R-squared, R-squared and F-statistic.

\[ \text{SagRank} = 67.540 \cdot 3.759 \text{ Upperclass} - 3.759 \text{ Coach} + 0.377 \text{ FiveYrAvg} \]

The last test ran on the final model was to test for Heteroskedasticity (figure 7). This refers to a spread of uneven data, and when plotted on a graph can be seen as a “cone like shape”. For the model I ran the Harvey heteroskedasticity test which showed that I had problems with one variable because the probability was less than .10, which was upperclass. In order to correct this problem, the white test was ran to removed the Heteroskedasticity (Figure 8).

**Figure 7:**

| Test Equation: LRESID2 Method: Least Squares Date: 12/10/13 Time: 05:05 Sample: 1115 Included observations: 114

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-1.109557</td>
<td>2.175678</td>
<td>-0.509982</td>
<td>0.6111</td>
</tr>
<tr>
<td>UPPERCLASS</td>
<td>6.175894</td>
<td>2.945601</td>
<td>2.096650</td>
<td>0.0383</td>
</tr>
<tr>
<td>COACH</td>
<td>0.010614</td>
<td>0.022949</td>
<td>0.462496</td>
<td>0.6446</td>
</tr>
<tr>
<td>FIVEYRAVG</td>
<td>0.100042</td>
<td>0.073340</td>
<td>1.364084</td>
<td>0.1753</td>
</tr>
</tbody>
</table>

| R-squared    | 0.054870    | Mean dependent var | 3.892026 |
| Adjusted R-squared | 0.029094 | S.D. dependent var | 1.951197 |
| S.E. of regression | 1.922503 | Akaike info criterion | 4.179695 |
| Sum squared resid | 406.6043 | Schwarz criterion | 4.275702 |
| Log likelihood | -234.2426 | Hannan-Quinn criter. | 4.218658 |
| F-statistic   | 2.128708    | Durbin-Watson stat | 1.952979 |
| Prob(F-statistic) | 0.100694 |           |           |
Figure 8:

Dependent Variable: SAGRANK
Method: Least Squares

Sample: 1 115
Included observations: 114
White heteroskedasticity-consistent standard errors & covariance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>67.54012</td>
<td>13.55686</td>
<td>4.981987</td>
<td>0.0000</td>
</tr>
<tr>
<td>UPPERCLASS</td>
<td>-3.759604</td>
<td>21.21838</td>
<td>-0.177186</td>
<td>0.8597</td>
</tr>
<tr>
<td>COACH</td>
<td>-0.139774</td>
<td>0.151225</td>
<td>-0.924279</td>
<td>0.3574</td>
</tr>
<tr>
<td>FIVEYRAVG</td>
<td>0.377561</td>
<td>0.455593</td>
<td>0.828724</td>
<td>0.4091</td>
</tr>
</tbody>
</table>

R-squared 0.015856  Mean dependent var 71.16825
Adjusted R-squared -0.010984  S.D. dependent var 11.85064
S.E. of regression 11.91555  Akaike info criterion 7.828023
Sum squared resid 15617.83  Schwarz criterion 7.924030
Log likelihood -442.1973  Hannan-Quinn criter. 7.866986
F-statistic 0.590763  Durbin-Watson stat 1.868628
Prob(F-statistic) 0.622337

Conclusion

Before starting this model, it was expected that teams that are more prone to oversigning players, would have a significant competitive advantage over teams who choose not to do so. After running two different models, you can see that this is not true. Both of the models ran did not have any variables that would actually be considered significant. Also, the regression ran for both had poor r-squared values, and failed the f-test. This study shows that oversiging does not influence a team’s competitive advantage.

With this being known, if coaches knew that in the long run, this does not have as big of an effect on a teams wins per season, would they change the way they recruit; more than likely not. This is done because coaches want to be able to choose between players, incase one would
not be academically available, or would be potentially hurt as well. Though the NCAA in recent years has been cracking down, this phenomenon is far from ceasing to exist.

**Limitations and Extensions**

Even though rules have been made against oversigning, these need to be made me strict. After doing this research and discovering all the negatives behind colleges doing this, teams should only be allotted the 25 new scholarships and that is all. Though some students will potentially fall through, it is not fair when all of the students can play, and players are left being dropped from the team.

When it comes to my data, I think a lot could’ve been done differently. It was hard to find accurate data about the number of players that are oversigned because some schools do not like to admit to doing so, or the numbers were not made public. I found different articles with different numbers, but this was the biggest challenge. I think that my data just didn’t support my theory very well, and if I was to do this again I would try to use different variables and find better data.
Work cited


Michelle, Hosick. N.p.. Web. 9 Dec 2013. 


