Nick Power

Zak Vinter

Part 4

Our research question is a causal one. We are interested in whether a policy of video replay in football games increases the number of penalties per team in England vs. Scotland. This policy was enacted in 2018 for teams only in England. There are three models that we ran in order to find the relationship of this effect.

The first model is a robust model run with comparing all the teams from England and Scotland. It is a basic diff-diff model. There is data taken from before the policy takes place and then the trends between the treatment (England) and control (Scotland) are looked at after the policy takes place. Then a difference between the trends is taken. Regression 1 shows the results of this model. The variables time, and area account for the fixed effects of the country and year in which we are looking. The interaction is the differential effect of the policy across the teams in the two countries based on the treatment and control. The coefficients on the time and area fixed effects are not of particular importance. Their interpretation is not easily understood and does not provide any real insight. The interaction term is the main variable of interest. The interpretation is as follows, teams in England received one more penalty on average after the policy took effect in 2018 than teams in Scotland where the policy didn’t take effect. The intercept here represents the amount of penalties all teams received regardless of the policy or country of origin. The main takeaway from this is that using instant replay may lead to the discovery of more penalties that would’ve originally gone unseen. This doesn’t guarantee that instant replay will lead to this result, only suggesting. It is quite plausible that the results from this regression point to the difference in the amount of penalties between these two country’s leagues as being a result of instant replay being used in England. But this is one model, there is no clustering for standard-errors, no synthetic controls, no other counterfactuals used to help check the robustness. This model may provide an insight, but it is only the tip of the iceberg. The main sources of bias would be the lack of more observations. The policy was implemented in 2018 and there are not many years of data after this in order to really compare. This may not capture the full trend, and may push the results to be a bit of an overestimate since there may be more of a upward effect from one outlier and so few other datapoints.

The second model is the same structure of the first model. The only difference is that it is comparing the differences within the treatment group of England. In order to do this the teams in the England football league are divided among teams in the top 6 and not in the top 6. This may go against convention where the control group is also affected by the policy. However, this still allows a comparison between the teams that have a larger budget and fanbase compared to teams that do not and the effect of this policy within the treated league. The coefficients here are nearly identical, instead of a treatment group being all the teams in England, it is just the teams that are in the top 6 of England. The control would be the teams that are not in the top 6 of England. The year fixed effects are the same. But the place fixed effects are this time represented by each club, as each club may have characteristics that may vary from club to club. The intercept here represents the amount of penalties per game that each team received regardless of placement in the league. The interaction here says that top 6 teams received 2.25 more penalties that teams who are were not in the top 6 from 2018 onward after the policy took effect. This is not entirely enlightening as it doesn’t show whether these penalties went in the higher ranked teams’ favor. Nor is it a properly run model. The best way to run this model would be to have a synthetic control where the entire league has a simulated trend from 2018-present in order to capture what the games would look like without the instant replay. From here there would be a diff-diff run to distinguish the difference between the synthetic control league and the treatment league, to capture the aggregate effect. After this a binary variable can be created for teams that are in the top 6 and not in the top 6 in each dataset. And the means of each variable of penalties received between each category within the dataset can be calculated. From here the means of difference in penalties received from top 6 vs not top 6 from each dataset can be compared to see how the policy effected each category. Sort of a descriptive analysis within a diff and diff model. The most obvious source of bias in this model is the lack of a control group. This doesn’t allow any sort of counterfactual to be compared to so the results of this regression have no real basis in which to compare to. There is no way of saying whether this pushes the bias up or down. The best way to proceed from this point is to do the aforementioned experiment design, with some more adjustments to account for proper structure and robustness.

The third model is similar to the second model. However, in this model the top teams in England are compared to the top teams in Scotland in order to accurately run a proper diff and diff. This model just allows another viewpoint in which to view the original model, in which the top teams can be compared in each country to see how the policy effects teams with a larger budget and fanbase. The notation here is similar to every other regression. The only differences being the category of the treatment variable, and the location fixed effects. Here the treatment group is all the top 6 teams in England, where the control is the top 5 teams in Scotland. The location fixed effects are replaced with country fixed effects. The intercept coefficient is representative of the amount of penalties each top team gets when not distinguished for in any way. The interaction term here says that the top 6 teams in England received 2.6 more penalties on average than top teams in Scotland since the policy took effect in 2018, holding all else constant. This points out that the instant replay was able to identify an increase in penalties between top teams in England than in Scotland. Interestingly, compared to the first model where all the teams were considered. In the first model each team received an additional penalty as a result of the policy. When this is scaled to just looking at the top teams, this relationship seems to be increased. Which suggests that there may be a difference in how the instant replay effects top teams and non top teams, which regression two tried to investigate. These results pave the way for a more in depth analysis and is a jumping off point for more in depth ideas to explore with this data. This is all speculation however. These results are reasonable to be causal for the top teams in England and Scotland, and not any further. Again, there is a lack of observations that may cause an over or under estimate of the overall trend. Compared to the ideal experiment described below, there are only two countries here, with leagues that may have different characteristics than other leagues. So this limits the scope of the regression. Player roster varies across place and time and therefore may not be accounted for in the fixed effects, which could lead to a positive bias since better players increase team performance, but lead to lower penalties overall. The biases present in this model should be concerning if the reader wishes to come to a policy decision or to go further than a simple trend comparison in the countries where this policy is present. Beyond this there is not enough data or sophistication in the model in order to account for a more complex conclusion.

The ideal experiment for all models would be to have all clubs and all teams have the same budget and then have the same players be randomly assigned to teams that would have instant replay, and teams that would not. The teams would be randomly assigned to have instant replay or not as well. Then have them play the same number of games over a longer period. Then repeating the process once every several years, and then comparing the aggregate differences over time. In order to progress from this stage there would hopefully be more years to draw from the sample, and in addition other countries with similar characteristics in their football leagues. From here more in depth and robust models could be run in order to capture a bigger picture of how this effects football.