Kaitlyn Hernandez

End of Summer Report

Introduction

The idea for my summer project was to see if investments in American youth soccer development programs have any influence on U.S. soccer players' success. I took a list of United States professional soccer players from the past five years and found information on their youth teams and the counties they operated from. The plan was to use API-Football to download player data across multiple leagues and countries, and then from there manually collect their youth team data and youth team financial data. From there, I would take player salary data and league performance information to use to inform "player success." I would then use Stata to compile all of the data and run a regression analysis to see if there was correlation between better funded youth teams and more successful professional soccer players. I wanted to do this research because of how different soccer is treated in the United States versus the rest of the world, especially how the U.S. is widely considered a "pay-to-play" country when it comes to youth soccer. As such, I wanted to see if more money would then lead to more professional success.

Results/Summary

By the end of the FURSCA period, my research plan changed a lot due to unforeseen complications with my data, and I did not reach my goal of having a completed regression analysis to interpret by the end of the eight weeks. I had originally intended to include a comparison with women's soccer, however there was no available data like we originally thought and so I ended up solely focusing on men's soccer. Additionally, once I downloaded the list of United States soccer players, where my advisor and I had originally expected around 400 players to begin youth team data collection for, we found that we were looking at a pool of nearly 800

players instead, which meant manually finding the youth teams of 800 players. As such, it took two weeks instead of one to collect that data and so we decided to use the county GDP of the youth teams and the latest year each player was in a youth team as a proxy for youth team funding since youth soccer teams in the United States operate as non-profit organizations, and thus receive funding based on county wealth.

Luckily, I managed to cut down the list of players as some played soccer for an overseas youth team or only played for their high school team, which was unrelated to my research of specifically United States youth soccer club teams. From there, I collected the real GDP of each team's county and state, and merged the data on Stata to be able to connect player's names to their youth teams data and the corresponding years. I also changed the factors for analyzing success, adding whether or not player's played in an overseas league, and if not, their current team's league standing in Major League Soccer, as well as game statistics from recent seasons (game appearances and total minutes played), to better account for other factors could be a part of "player success" when running my regression analysis.

The next snag in my data collection was merging salary data, as there was a mismatch in names, as on Stata variables must be the exact same (in this case, exact same spelling of names) in order to merge, and the salary data we downloaded from Major League Soccer used player nicknames rather than legal names, meaning I had to manually combine those datasets on top of downloading overseas player data by myself, which I am still currently working on. As there was so much more data than expected, I had to extend certain portions of data collection and change up my timeline to focus on data collection throughout the duration of the eight weeks. Going forward, I will finish my last bit of salary data collection and then run my regression analysis, interpreting the results and beginning writing my thesis, and then presenting at Elkin Isaac.

Conclusion

      This research project is really important to me in furthering my economics statistics knowledge and will also help me to be the basis of my thesis, which I am doing under the Economics department. This project really helped me with time management, as I ended up with a lot more work than I expected in eight weeks, and so it made me a lot more efficient, even if I did not completely finish by the end of it. I will also be presenting my final findings in the Spring, at the Elkin Isaac Research Symposium, which I am very excited for as I will have my thesis completed by then and thus be able to include aspects of that in my presentation. I plan to work more with sports economics in the future, and this research could help get my foot in the door with graduate schools as well as future job/internship opportunities. Hopefully, in the future, in graduate school, I can expand upon this research and maybe focus more heavily on specific variables, or extend it into a cross analysis by replicating this work with other countries and seeing if there are any trends. I really appreciate FURSCA and the Christ Student Research Endowment for allowing me to pursue this research and for funding my interests in sports economics, and I thank both very much for this opportunity.