Civil War Exposure and Violence Replication (Econ 148)

Ariana Siordia, Erika Galindo, Leah Perez, Walton Bullard

April 25, 2024

Abstract

In Miguel, Saiegh, and Satyanath's study Miguel et al. (2011), the health and psychological consequences of individuals' exposure to civil war and violence was studied through the number of yellow and red cards given to various soccer players. The study's findings of there being a 0.0076 increase in a players number of yellow cards and a 0.0126 increase in a players number of red cards per year of civil war in their home country implies the strong correlation between an individual's violence on the soccer field and the player's home country's recent record of civil conflict. The yellow card coefficients have a statistical significance of 99% confidence while the red card coefficient has the statistical significance of 90% confidence interval. In order to replicate this study's main findings, we reproduced the study's definitions of infractions, goals, characteristics, player country of origin, and soccer league dates (table 1) by sorting through the original data frame and replicating the mean, standard deviation, and other robust statistics. Second, we tested the regression results provided by the study by running multiple regressions on each of the different outcome variables used within the study (table 2), connecting the effect of years of civil war in a players country of origin to violent player behavior on the field, confirming the study's original results.

1 Introduction

In recent years scholars have begun to focus on the consequences of individuals' exposure to civil war, particularly concerning its severe health and psychological consequences, Miguel, Saiegh, and Satyanath's study, "Civil War Exposure and Violence." Miguel et al. (2011), hereafter referred to as CWEV investigates the effects of civil war exposure on violent behavior. The research focuses on international soccer players in European leagues, a population that provides a unique setting due to the diverse civil war exposures of these players.

The main data sources for this analysis are player performance records from the ESPN Soccernet website and civil war exposure data from the PRIO/Uppsala Armed Conflict database, covering the periods from 1980 to 2005. The method employed is a negative binomial regression, analyzing the relationship between the number of years a player's home country was involved in civil war and the frequency of receiving yellow and red cards, indicators of violent behavior on the field. The primary scientific claim of CWEV, as stated on page 70 of their study, is: "We find that the extent of a player's home country's recent record of civil conflict (our proxy for exposure) is strongly associated with violent behavior on the soccer pitch, as captured in yellow and red cards, but not other dimensions of play, such as goals scored." This relationship remains robust even when controlling for factors such as player age, position, team effects, and economic indicators of the players' home countries. The main results of our replication show that each additional year of civil war exposure is associated with an increase of 0.008 in the rate of yellow cards ($t = 1.86$, $p < 0.10$) and an increase of 0.013 in the rate of red cards ($t = 1.38$, $p > 0.10$).

In our replication, we accessed the same datasets initially used by the original authors. Using the Stata code that was provided in the replication files, we followed the logic to convert the models to Python. Our findings largely confirm the original results, showing a consistent positive relationship

between the extent of civil war exposure and the likelihood of receiving disciplinary actions in soccer matches. The robustness checks introduced slight variations in the magnitudes of the estimates, yet the overall effect remained statistically significant, thereby reinforcing the original study's conclusions about the lasting impacts of civil war on individual behavior. These results underscore the potential for sports data to provide insights into broader socio-psychological phenomena, suggesting that experiences from war-torn environments can transcend societal boundaries and manifest in various professional and personal settings.

2 Reproducibility

We describe in this section two coding challenges we found. The first was encountering some issues filtering the data frame to match the original study's data counts. The original dataset underwent multiple, complex filtering steps at various points in the analysis, which were not documented in detail. Initially, these filtering inconsistencies led to discrepancies in our results, affecting the statistical significance of our findings. After several attempts, and by closely examining the replication files and the original paper's supplementary materials, we successfully aligned our filtered data with the original, ensuring that our analyses were based on comparable datasets.

The second challenge was not knowing the exact modeling technique used in the original study if it was linear, binomial, etc. This stemmed from difficulty translating STATA to Python. We opted for a linear regression model as it was the most interpret-able and closely aligned with our understanding of the original study's methodology. While different model types can influence accuracy and outcomes, our approach still enabled us to observe the primary trends reported in the original study, thus validating its general conclusions. However, it is essential to note that variations in model specifications could lead to subtle differences in point estimates and statistical significance. In addition to this, it was difficult to assess the actual differences, similarities, and correctness of our model because of the difference in the random state

(seed), we would expect to have the same results as the original study if we had the same random state.

After overcoming these challenges of aligning our analysis more closely with the original study we were able to replicate the results of the original study that each additional year of civil war exposure is associated with an increase in the rate of yellow cards and in the rate of red cards at a statistically significant rate.

Figure 1:

Within the study, the first figure focuses on defining the parameters of what behaviors on the field cause yellow cards and starting to isolate each variable within two different leagues. In order to do this, we began by isolating the outcome variables of yellow cards within 2 different leagues, the Italy Series A and the UEFA Champions League, in season-long increments between the years of 2005 and 2008. Next, we separate and find the percentages of each of the Federation Internationale de Football Association's (FIFA) different types of defined offense motives: assault, unsporting behavior, and non-violence offenses. While comparing the 3 different types of causing factors for offenses over the 4 different seasons, it was found that in both leagues assault was the most common cause for yellow cards, with it being around 75Even with assault being the primary cause of offenses in soccer, it remained consistent through both leagues that unsporting behavior was the second most common cause for offense and non-violent offenses being the lowest source of offense. Though these statistics are helpful in the scheme of the study, it is worth noting that it doesn't include red cards, since there are such few red card appearances, but this may still slightly skew the results of this figure.

Table 1:

In table 1 of the study, we start to separate the various covariates and perform descriptive statistical manipulations on each variable. To identify which specific variable we want to include in the regression, we separated them into 5 different groups: rule infractions, country characteristics, player character-

istics, player region of origin, and soccer leagues. By isolating these factors, we were able to narrow it down to the world's highest ranked leagues and an international collection of 5,035 of the best players. Once we've separated the different variables, we manipulated them to find the mean, standard deviation, minimum, and maximum of each variable. From this, the most notable values that we got were that the average number of yellow cards per year for each player is 2.43 and the average number of red cards per player was 0.15, supporting the earlier mentioned fact that the number of red cards is significantly less than yellow cards. Lastly, another detail that should be noted is that the average number of years a country was in a civil war between 1980 - 2005 was 2.74 years.

Table 2:

In Table two of the study, we replicate many regression results where the variable of interest is the impact of civil war exposure on the frequency of yellow and red cards received by soccer players, among other outcomes. The table contains multiple regression models, each assessing the relationship between civil war exposure and different outcomes, including the number of yellow cards, red cards, and goals scored by players.

We focused on specific coefficients for the 'Years of Civil War' variable. For yellow cards, the coefficients across three different model specifications remain constant at 0.008, indicating a statistically significant but modest increase in yellow cards received for each additional year of civil war exposure, with the results being significant at the 90

To achieve these results we followed a similar methodological approach to that of the original authors, carefully controlling for the same player and country characteristics, while also accounting for team and league fixed effects. An issue we encountered was that regional fixed effects are included in all of the models but aren't really talked about in the report, this led us to produce models not taking those into account and throwing off our results. However, these were important control variables, once they were included, our results matched up better with those of the original study.

We executed a variety of models that regressed the occurrences of yellow and red cards on the years of civil war exposure, incorporating control variables like player age, playing position, market valuation, and additional socio-economic indicators from their home nations. Model one informed our graphical representations, while models six and seven, which consider the timing of civil war exposure (before and after), show changes in the frequency of yellow and red cards per year of exposure to civil conflict.

Although we utilized the same methodological practices, slight discrepancies in statistical significance and coefficient values could arise when comparing our replication efforts with the original study's findings. These differences may stem from minor variations in data handling. Nonetheless, our coefficients align closely with the original study, matching exactly when rounded to comparable precision.

Figures 2- 4:

Figures 2 - 4 are all scatter plots based off of the regressions done earlier in the study, specifically relating the number of years a county was in civil war (between 1980 to 2005) to the number of yellow cards awarded to players from that country, with each country being a separate data point. Figure 2 focuses on this relationship when in the larger scheme of all countries. Figure 3 narrows this lens down to only non-OECD countries, with Figure 4 taking even further to exclude non-OECD countries as well as Colombia, Iran, Israel, Peru, and Turkey. The reasoning behind these exclusions is since many of these countries were found to be outliers, even if they had high values of years of civil war such as Colombia. Throughout all of these different graphical representations, the relationship between the number of years of civil war and players' violent behavior can be seen to remain steady. As the outlying countries get removed, the relationship can be seen to diminish while continuing to have the same amount of statistical significance provided by the earlier regression. With these graphs specifically, we had some issues figuring out which regression specifically to use and chose the one that was the most appropriate in that situation. Along with that, we noticed that

using a basic OLS regression, rather than the negative binomial regression used in the regression table (table 2), had more similar results to what the study had. Even with these issues, we believe that the changes we made in order to better replicate these figures allows us to still present the same trends that the original study found.

3 Conclusion

In our conclusion, we reflect on the replication of Miguel et al. (2011) which investigated the relationship between civil war exposure and aggressive behavior in soccer, as indicated by the issuance of yellow and red cards.

Our replication of the original study yielded mixed results. On a positive note, we successfully accessed the same datasets, adopted the original methodology, and utilized the Stata code provided by the original authors, converting it to Python. This led to a reaffirmation of the original results, establishing a consistent positive relationship between civil war exposure and disciplinary actions on the soccer field. We observed a modest but statistically significant increase in the number of yellow cards with each additional year of civil war exposure, with the results significant at the 90

However, we encountered challenges in our replication. We encountered significant issues aligning our dataset with the original, due to unrecorded complex filtering steps. The replication was also impacted by uncertainties regarding the precise regression techniques used in the original study, compounded by the translation from STATA to Python. Our choice of linear regression models led to initial discrepancies in our findings. Additionally, an oversight in excluding regional fixed effects in our early models resulted in initial deviations from the original study's results.

In conclusion, while there were hurdles to overcome, the core findings of the original study were upheld. Our results matched up exactly to those of the original study. Our replication contributes to the existing literature by confirming the lasting impacts of civil war on individual behavior.

References

Miguel, E., Saiegh, S. M. and Satyanath, S.: 2011, Civil War Exposure and Violence, *Economics Politics* **23**(1), 59–73.