The arrival of Guardiola in the Premier League: Success or Failure?

Lucas Veteikis

Introduction

Pep Guardiola is a former accomplished football player who in the last years has been cultivating his career as a coach. He started his career at Barcelona where he could work with players such as Messi or Iniesta to achieve to be named one of the best football teams in the history of football. Following this, he spent three seasons working for Bayern Munich.

The purpose of this study is to assess the results of his next step. In the season 2016/2017 he started working for Manchester City and many football fans accused him of mainly managing powerful teams in less competitive leagues, and they had serious doubts about how well he would adapt to the Premier League.

Data

To conduct this analysis, A total of 380 matches from five seasons (2014/2015 to 2018/2019) were studied.

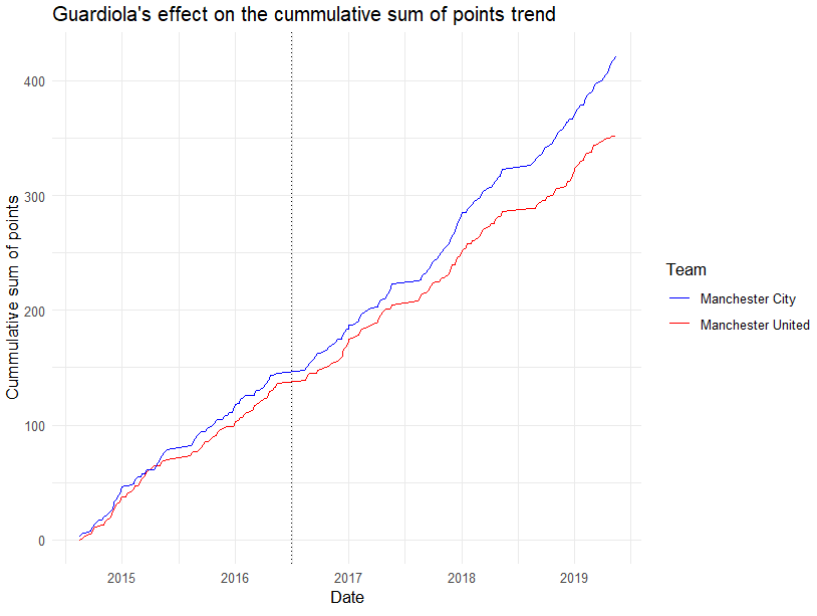
Matches data was extracted from https://fbref.com/en/ while data regarding transfers was retrieved from https://www.transfermarkt.com.ar/.

Groups under study

In accordance with the purpose of the study, the treatment group consists of Manchester City, the team which hired Guardiola for season 2016/2017.

Selecting a suitable control group was a complex decision, but the choice for this group is Manchester United. There are many reasons behind this choice. Firstly, Manchester United is the historical rival of Manchester City. Additionally, both teams have similar budgets and similar squad quality. Lastly, the main reason for their comparison is that they were not experiencing their best time on the premier league before Guardiola's arrival.

From a statistical approach, I have compared and established the similarity in trends prior to the treatment, which can be visually observed in the graph attached, where the vertical line marks Guardiola's arrival. This is a fundamental argument to perform an analysis like the one presented in this study.



Furthermore, t-tests were conducted to compare the means of our variable of interest (points) before and after treatment. The

results were favorable, suggesting that the average points earned per match for both teams was similar before treatment ($p > 0.05$) but different after its implementation ($p < 0.001$).

Methodology

To accomplish the study's of assessing Guardiola's impact on Manchester City, it has been performed a difference in differences (DID) methodology, which is a statistical technique that attempts to mimic an experimental research design using observational study data, by studying the differential effect of a treatment group versus a control group in a natural experiment. The seasons 14/15 and 15/16, that are previous to the arrival of Guardiola, were used as the period pre-treatment, whereas the seasons 16/17 to 18/19 as post-treatment.

Regression

The regression proposed followed a standard DID regression structure, including a variable that identifies the group under treatment (*treatment*), a variable whose purpose is to flag the pre- and post- treatment periods (*after*), and three different variables which are relevant to control the technique's outcomes.

$$Cumsum\_pts = \beta_0 + \beta_1 * Treatment + \beta_2 * After + \beta_3 * (Treatment * After) + \beta_4 * Transfer\_mkt\_balance + \beta_5 * Group\_stage + \beta_6 * Knock\_out\_stage$$

Dependent variable

Given the amount of statistics available in football that can be used to evaluate team performance, determining the appropriate dependent variable has been another challenge in the methodology. Ultimately, I opted to use the accumulated sum of points (*Cumsum_pts)* as the dependent variable. Initially, I attempted to employ points per match, but due to high variance within this variable, the study did not lead to meaningful results. By using the accumulated sum of points, capturing the lift on the trend in the team under treatment was possible and mitigating this variance problem mentioned.

Independent variables

With the objective of explaining the trend of the accumulated sum of points, it has been incorporated into the regression two kinds of variables, the conventional DiD variables and the control variables.

The standard DiD variables to perform the described methodology were *"Treatment*" and *"After*". These dummy variables represent, respectively, the group undergoing treatment and the time periods pre- and post- treatment.

The control variables, on the other hand, were considered as factors that could drive an increase or decrease in the points accumulation trend. One of these variables is *"Transfer_mkt_balance",* which stands for the difference between the amount earned from player sales and the amount spent on player acquisitions. The goal of this

variable is to identify if in a particular season there has been a significant disparity between how much both teams spent on the transfer market and that could drive a difference in points. The other set of control variables *"Group_stage"* and *"Knock_out_stage"* are dummy variables designed to flag if the team had an international cup match within the four days preceding or following the match in question, differentiating if the international match is of group or knockout stage. Typically, when a team has an international match near a premier league match, they may opt to rotate players and employ substitutes on the Premier League, which could potentially lead to a decrease in the points earned in that particular match.

Results

The regression had an r-squared of 0.8064, which is considered a good fit and a good sign of the quality of the model.

Upon variable analysis, the primary result of interest, the interaction between *treatment* and *after,* is statistically significant at a confidence level of 0.05. This suggests that Guardiola's arrival to Manchester City had a discernible impact on the points earned by the team, analyzing the coefficient associated to this particular variable, this impact results on 34.47 points won throughout the three seasons of treatment keeping the rest of the variables constant.

Additionally, the variable *treatment* did not result statistically significant which is a good sign that both teams were not different previous the treatment. The variable *after* did show statistical significance, which is reasonable as it is being used an accumulative sum as the dependent variable.

Furthermore, the *Group_stage* variable did yield statistical significance and according to its coefficient, the teams lost an estimate of 24.84 over the course of five seasons as a result of playing a group stage international match within the week of a Premier League match. However, the *Knock_out_stage* did not result statistical significant. This result can be understood as the group stage matches are more frequent than the knockout ones, and that could make the difference and force teams to use substitutes in the local league.

	Cumsum_pts
Intercept	134.62*** (7.15)
Treatment	14.86 (8.13)
After	187.11*** (7.42)
Transfer_mkt_ balance	0.62*** (0.04)
Group_stage	-24.84*** (6.65)
Knock_out_ stage	19.50 (10.55)
treatment:after	34.47** (10.49)
Adjusted R^2	0.8064

Note: Standard errors in parentheses. **p < 0.01 ***p < 0.001

Finally, the *Transfer_mkt_balance* showed statistical significance as well, but with a low coefficient. This suggests that while there was a statistical relationship, its practical impact on the points earned may be limited or require further investigation.

Conclusions

To conclude this work, the statistically significant coefficient of 34.47 for the "treatment:after" interaction in the regression suggests that Guardiola's arrival to the team has been a success.