Crossing in soccer has a strong negative impact on scoring:

Evidence from the English Premier League and the German Bundesliga

Jan Vecer, Frankfurt School of Finance and Management

NESSIS 2013, Harvard University

September 21, 2013

Abstract

Crossing in soccer plays a significant role in scoring, about 15% of all goals scored in the recent seasons of the English Premier League are the result of open play crosses. However, crossing from an open play is hugely inefficient, only 1 open cross out of 91.92 leads to a goal on average. When we estimate the impact of open crossing on scoring of the individual teams using multilevel Poisson regression, we conclude that the net effect of crossing is typically negative or neutral at best. An average team can score up to additional 0.656(?) goals per game if it reduced open crossing. The quality of the team is the major explanatory factor on the number of such missed scoring opportunities, stronger teams miss more goal opportunities in general when crossing than weaker teams.

Abstract

Stronger teams have more options how to score and open play crossing seems as one of the suboptimal ways of a goal creation. Teams such as Arsenal, Chelsea, Liverpool, Manchester City or Tottenham have a potential of scoring an extra goal per match if they reduced open crossing. A reversed picture is seen in the defense analysis, more goal opportunities are missed in general when crossing against weak teams than crossing against strong teams. Interestingly, the actual conversion of open crosses to goals plays only a minor role for explaining the impact of open crossing on goals.

Original Motivation

The original question leading to this research was if there is any way to get a better prediction of the outcome of the soccer game from statistics obtained during the game. There is a large and liquid in-play betting market on soccer that trades various events:

- Win,
- Draw,
- Loss,

a team in a given game plus additional contracts such as the

- Total Number of Goals (including more than N + 0.5 goals),
- Exact Score,
- Team to Score Next + No Goal.

There is an extensive paid database of the betting quotes from Betfair.

How to Compute these Odds?

A reasonable approximation of the dynamics of the soccer score is a Poisson process for the goal distribution. The goals scored in the remainder of the game should follow

$$\mathbb{P}(X_T - X_t = k) = e^{-\lambda_t} \frac{\lambda_t^k}{k!}$$

for the home team and

$$\mathbb{P}(Y_T - Y_t = l) = e^{-\mu_t} \frac{\mu_t^l}{l!}$$

Here, the λ_t and μ_t play the role of scoring intensities for the two teams, the expected number of goals to be scored in the remainder of the match. Furthermore, if we assume independence of the goals scored, it is relatively straightforward to obtain all the betting quotes from the Poisson model, where parameters λ_t and μ_t serve as inputs.

Known Limitations of Poisson Model

- The scores are not independent. The correlation of the score in the English Premier League since 2006 is -0.057. Moreover, the realized fraction of draws is higher than implied from an independent Poisson model.
- There is some memory in goals, but this effect is reasonably small. One can fully estimate this effect from betting contracts on the Next Goal (which team scores next). The Poisson model implies that the quote on the Next Goal should stay the same before and after each goal.

Implied Intensities

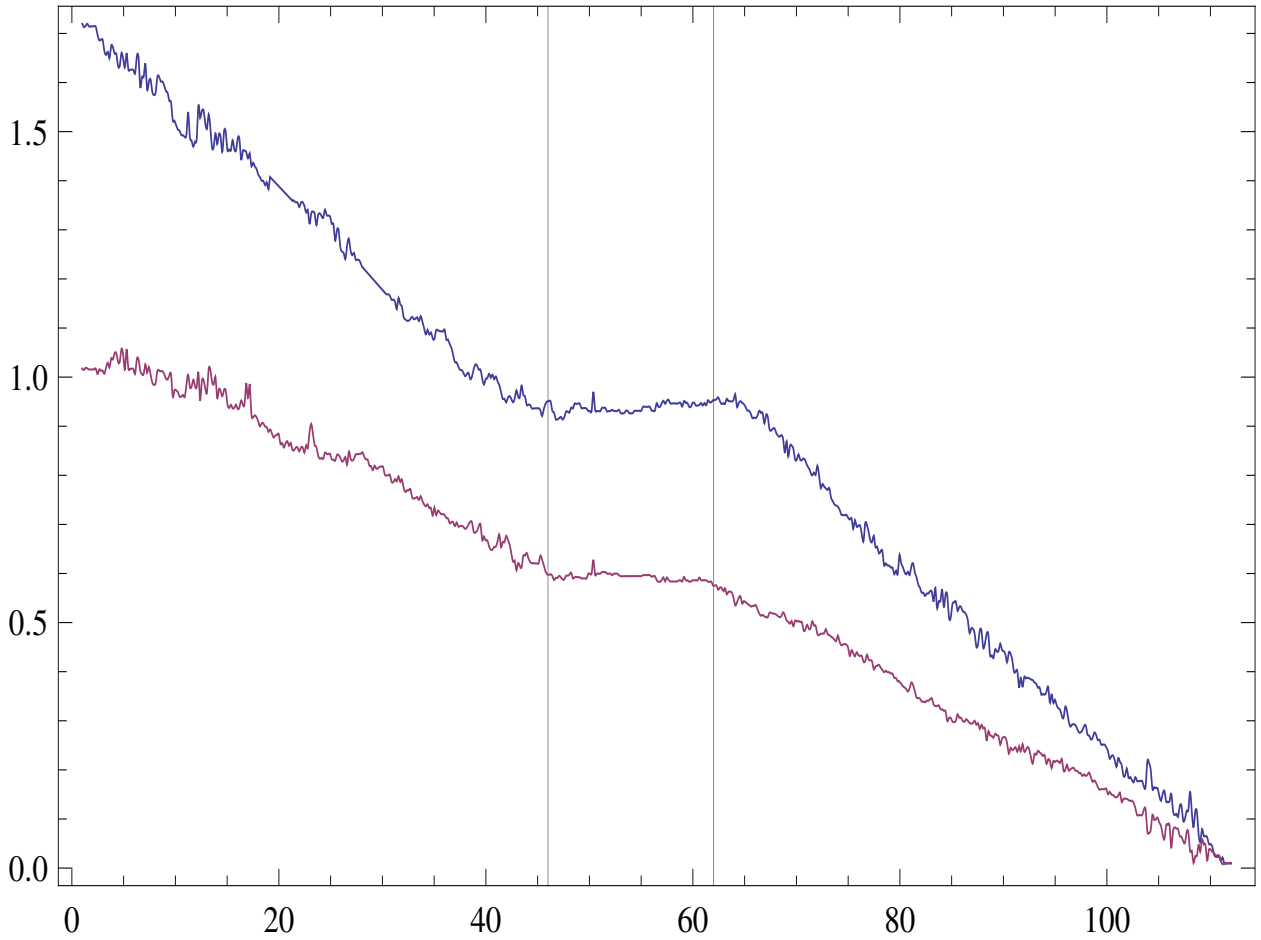


Figure: Arsenal-Chelsea 0:0, April 21, 2012

Inference of the Intensities from the Athletic Performance

Do the scoring intensities depend on some athletic performance data?
Source of data:

- OPTA (since 2008, 1900+ games)
- www.espnfc.com (match reports since 2008)
- www.premierleague.com (since 2006, 2700+ games)
- www.bundesliga.de (since 2009, 1250+ games, tracking since 2011)

What are the significant variables?

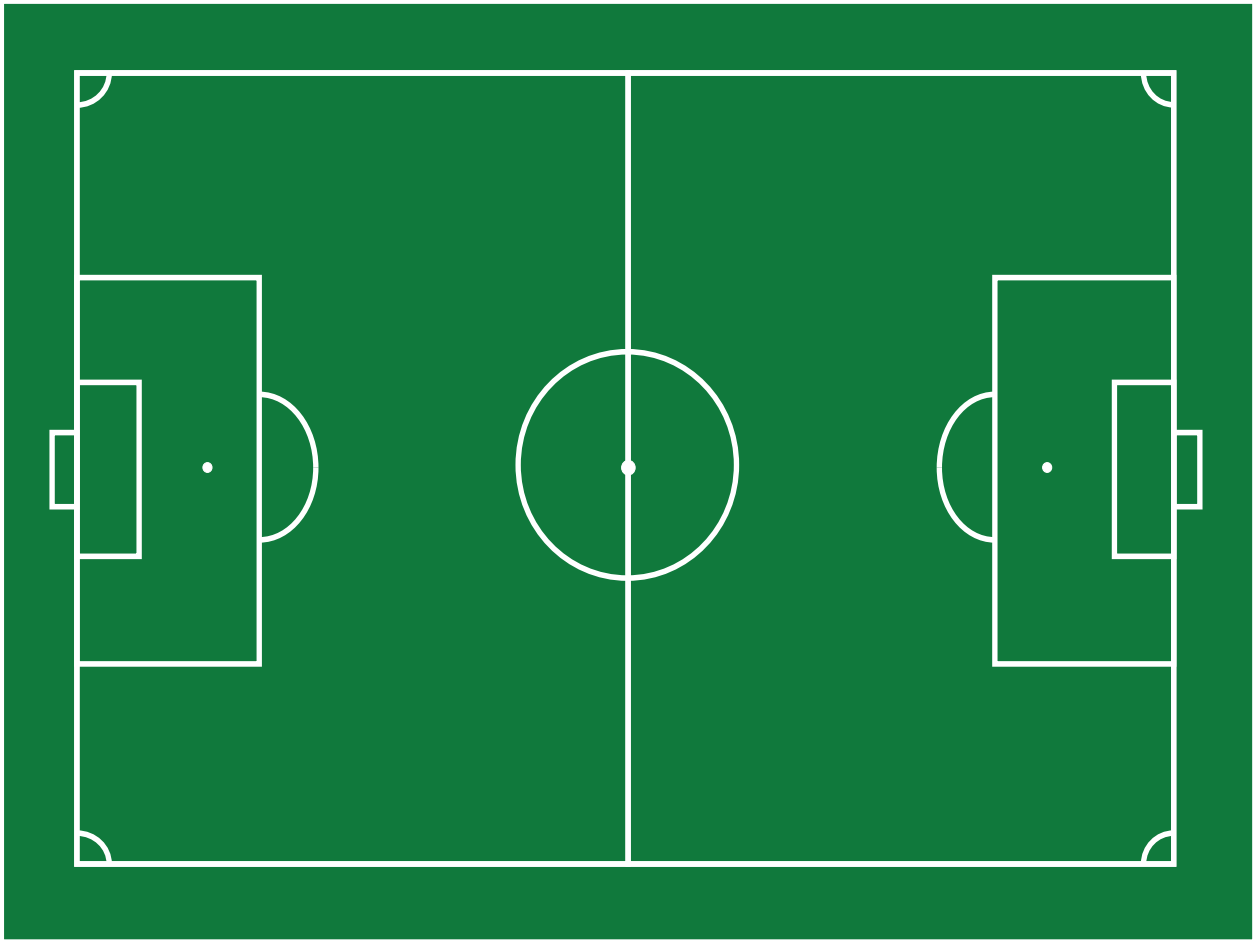
- Top Speed (Bundesliga)
- Discipline + Stoppages
- Open Crosses(!)

Some Facts about Crossing

A cross is an airborn delivery of a ball from the side of the field across to the front of the goal.

- An average EPL team makes 18.2 open crosses per game and scores 1.33 goals per game, an average Bundesliga team makes 11 open crosses per game and scores 1.45 goals per game.
- In the EPL, 18.2 open crosses produce 3.7 good crosses and 14.5 bad crosses, meaning that the vast majority of open crosses results in a loss of the possession in a favorable position.
- A goal is scored per 92 open crosses.
- The quality of crossing is highly variable among the teams, Manchester United needs 43.8 crosses to score a goal, Southampton needs 143.2 crosses to score a goal.
- Stong observational bias on TV highlights that show mostly good crosses and crosses leading to goals.
- There is an ongoing discussion about effectiveness of open crossing among football bloggers, but the analysis has been limited only to descriptive statistics.

Football Pitch



Conversion Statistics - Attack

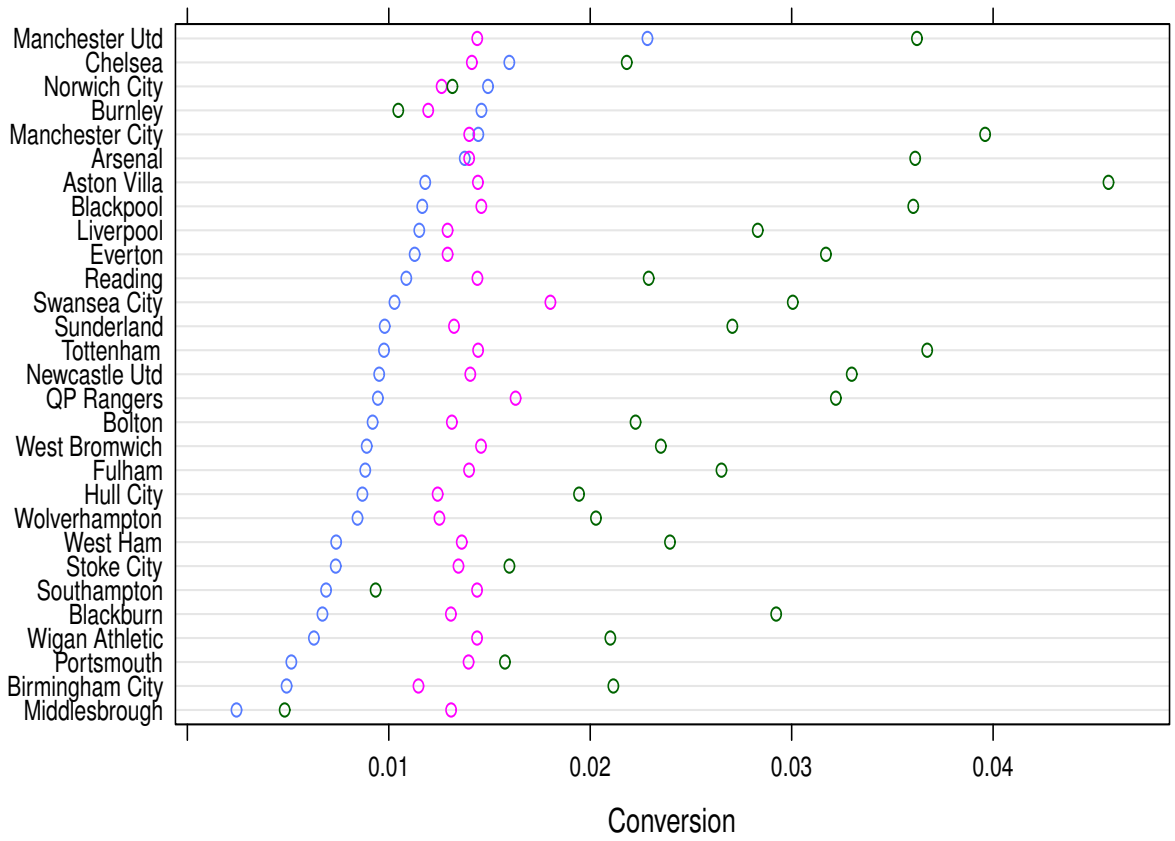


Figure: The fraction of open crosses (blue), final third entries (red) and outside the box shots (green) that results in a goal for individual attacking teams.

Conversion Statistics - Defense

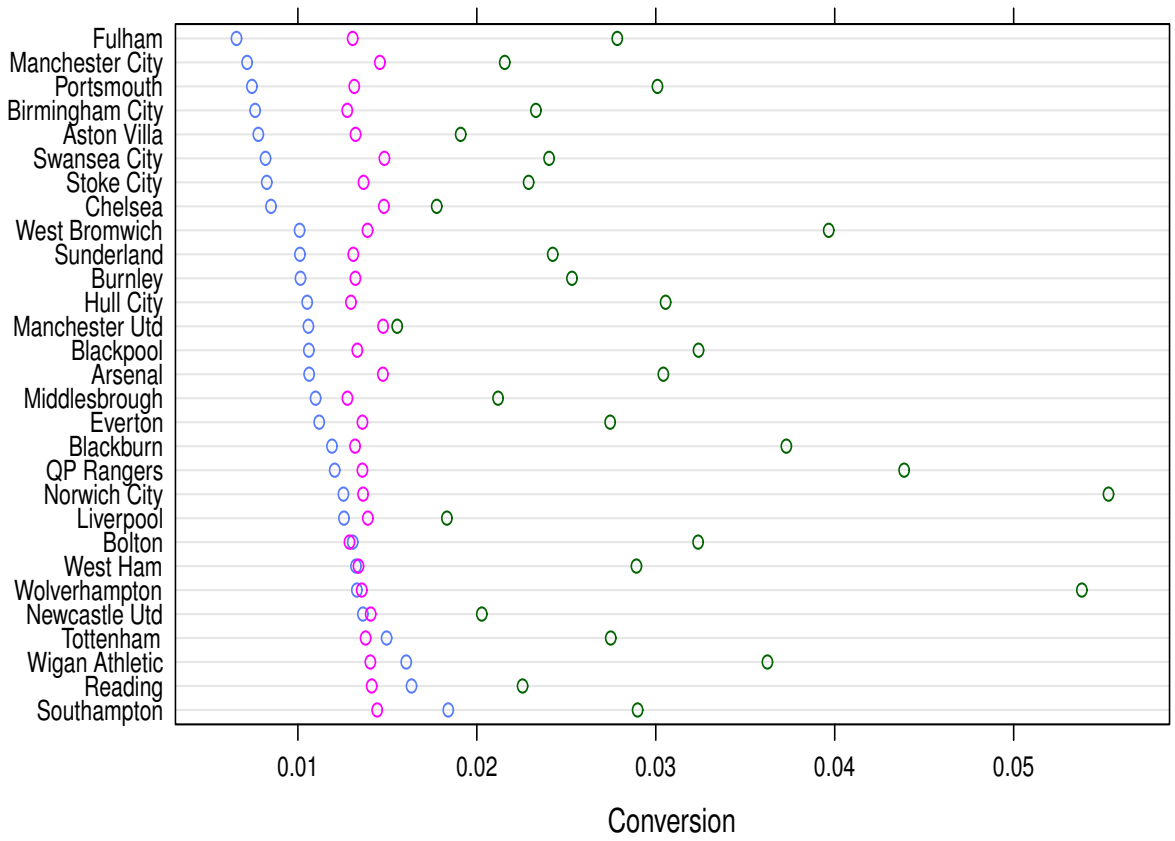


Figure: The fraction of open crosses (blue), final third entries (red) and outside the box shots (green) that results in a goal for individual defending teams.

Concerns

Estimation of the impact of crossing on goals should addresses the following issues:

- Crosses may lead to goals indirectly in a follow up play.
- By crossing the team is giving up an alternative way of playing.

Analysis of Goals regressed on Open Crosses addresses that.

The Statistical Model

Multilevel cross sectional Poisson regression: the teams are grouped according to the attack (using variable $j[i]$) and the defense (using variable $k[i]$):

$$\begin{aligned}\text{Goals}_i \sim \text{Poisson}\big(\exp\big((\beta^I + u^I_{j[i]} + v^I_{k[i]}) \\ + \beta^H \cdot \text{Home}_i + (\beta^C + u^C_{j[i]} + v^C_{k[i]}) \cdot \text{Cross}_i\big)\big) \quad (1)\end{aligned}$$

$$u_j \sim N(0, \Sigma_u)$$

$$v_k \sim N(0, \Sigma_v)$$

EPL Model

	Model 1
(Intercept)	0.473376
	(0.100849)
OpenCross	-0.022861
	(0.003196)
Home	0.417204
	(0.029636)
AIC	4229
BIC	4286
Log Likelihood	-2106
Deviance	4211
Num. obs.	3800
Num. groups: Team	29
Num. groups: Against	29
Variance: Team.(Intercept)	0.121923
Variance: Team.OpenCross	0.000110
Variance: Against.(Intercept)	0.112291
Variance: Against.OpenCross	0.000028

EPL Model - Attack

	(Intercept)	OpenCross	Home
Arsenal	1.042368	-0.030510	0.417203
Aston Villa	0.425882	-0.019876	0.417203
Chelsea	1.097310	-0.031954	0.417203
Everton	0.671642	-0.024262	0.417203
Fulham	0.273865	-0.017672	0.417203
Liverpool	0.860616	-0.029654	0.417203
Manchester City	0.846105	-0.029128	0.417203
Manchester Utd	0.798300	-0.011707	0.417203
Newcastle Utd	0.565930	-0.024579	0.417203
Norwich City	0.377409	-0.017618	0.417203
Southampton	0.593570	-0.025884	0.417203
Stoke City	0.011150	-0.011515	0.417203
Sunderland	0.472198	-0.026832	0.417203
Swansea City	0.387274	-0.022690	0.417203
Tottenham	1.024221	-0.040664	0.417203
West Bromwich	0.539959	-0.027303	0.417203
West Ham	0.241390	-0.015480	0.417203

EPL Model - Defense

	(Intercept)	OpenCross	Home
Arsenal	0.048665	-0.016203	0.417203
Aston Villa	0.446672	-0.022443	0.417203
Chelsea	-0.093416	-0.013975	0.417203
Everton	0.146778	-0.017741	0.417203
Fulham	0.331304	-0.020634	0.417203
Liverpool	-0.062757	-0.014456	0.417203
Manchester City	0.003006	-0.015487	0.417203
Manchester Utd	-0.152201	-0.013053	0.417203
Newcastle Utd	0.569296	-0.024365	0.417203
Norwich City	0.646480	-0.025575	0.417203
Southampton	0.541325	-0.023927	0.417203
Stoke City	0.358285	-0.021057	0.417203
Sunderland	0.439709	-0.022333	0.417203
Swansea City	0.340518	-0.020778	0.417203
Tottenham	0.149979	-0.017791	0.417203
West Bromwich	0.670652	-0.025954	0.417203
West Ham	0.612075	-0.025036	0.417203

Bundesliga Model

	Model 1
(Intercept)	0.374347
	(0.088441)
OpenCross	-0.020001
	(0.004450)
Home	0.281266
	(0.034298)
AIC	3009
BIC	3061
Log Likelihood	-1495
Deviance	2991
Num. obs.	2536
Num. groups: Team	24
Num. groups: Against	24
Variance: Team.(Intercept)	0.064105
Variance: Team.OpenCross	0.000080
Variance: Against.(Intercept)	0.070237
Variance: Against.OpenCross	0.000070

Bundesliga Model - Attack

	(Intercept)	OpenCross	Home
1899 Hoffenheim	0.343475	-0.015867	0.281266
1.FC Nurnberg	0.278478	-0.024494	0.281266
1.FSV Mainz 05	0.384550	-0.026493	0.281266
Bayer 04 Leverkusen	0.662520	-0.021821	0.281266
Borussia Dortmund	0.860156	-0.027790	0.281266
Borussia Mgladbach	0.390397	-0.021744	0.281266
Eintracht Frankfurt	0.326103	-0.019196	0.281266
FC Bayern Munchen	0.903426	-0.015981	0.281266
FC Schalke 04	0.506178	-0.012962	0.281266
Hamburger SV	0.419429	-0.021852	0.281266
Hannover 96	0.483677	-0.024369	0.281266
SC Freiburg	0.332071	-0.020679	0.281266
SV Werder Bremen	0.462857	-0.011292	0.281266
VfB Stuttgart	0.624621	-0.027100	0.281266
VfL Wolfsburg	0.511445	-0.017644	0.281266

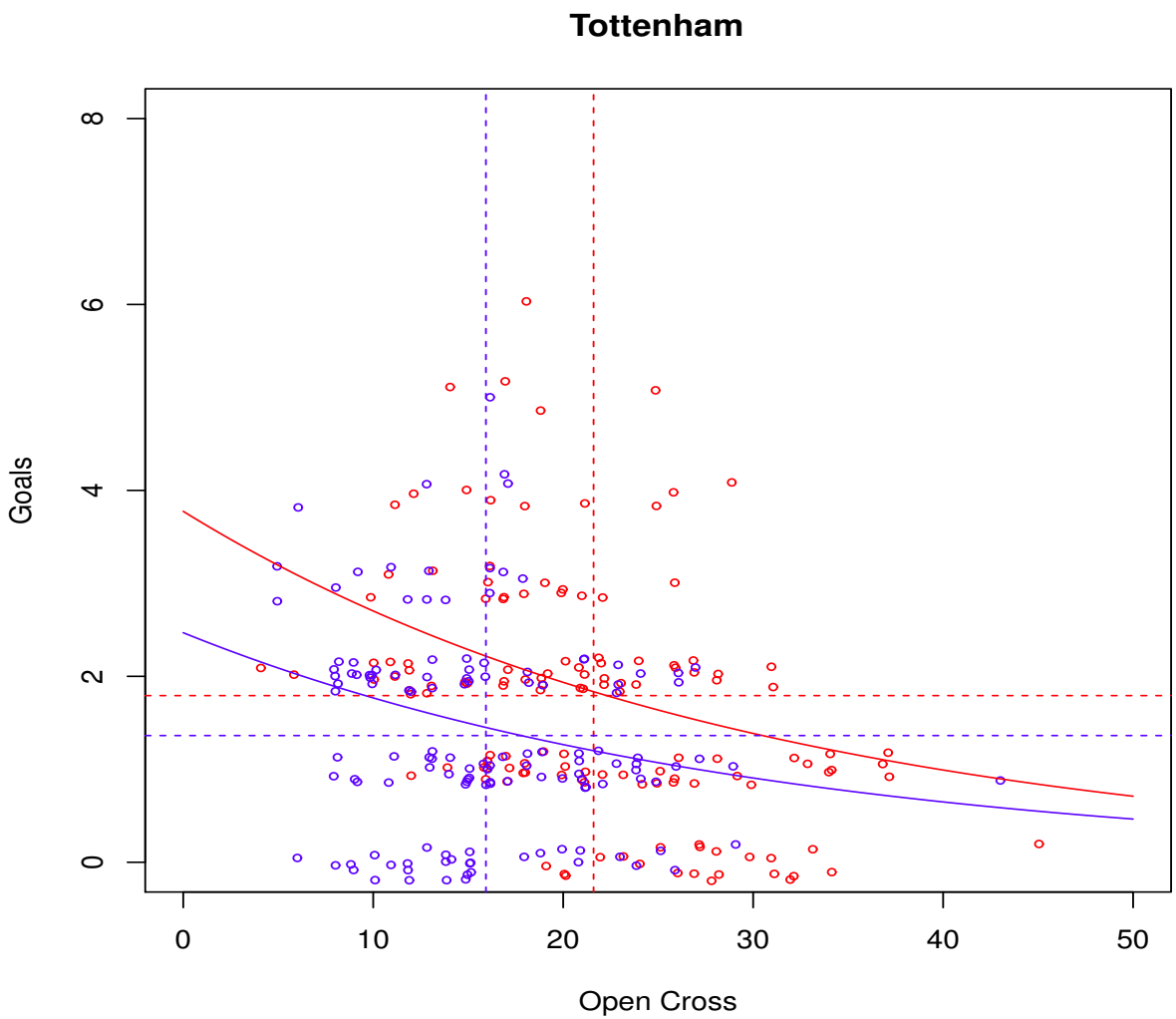
Bundesliga Model - Defense

	(Intercept)	OpenCross	Home
1899 Hoffenheim	0.397648	-0.020738	0.281266
1.FC Nurnberg	0.326835	-0.018498	0.281266
1.FSV Mainz 05	0.260753	-0.016407	0.281266
Bayer 04 Leverkusen	0.162643	-0.013303	0.281266
Borussia Dortmund	-0.158434	-0.003145	0.281266
Borussia Mgladbach	0.426736	-0.021658	0.281266
Eintracht Frankfurt	0.356889	-0.019448	0.281266
FC Bayern Munchen	-0.327126	0.002192	0.281266
FC Schalke 04	0.134238	-0.012404	0.281266
Hamburger SV	0.384058	-0.020308	0.281266
Hannover 96	0.482066	-0.023409	0.281266
SC Freiburg	0.398418	-0.020762	0.281266
SV Werder Bremen	0.517723	-0.024537	0.281266
VfB Stuttgart	0.403881	-0.020935	0.281266
VfL Wolfsburg	0.479178	-0.023317	0.281266

Things to Notice

- The impact of crossing on goals is negative for most of the teams, it is neutral at best.
- Stronger attacking teams tend to have a more negative impact on scoring than weaker teams with a single exception of Manchester United. This is due to the fact that aerial delivery of the ball has less precision and thus more luck than skill is involved. Stronger teams benefit more from situations that depend on skill in contrast to situations that depend on luck.
- The negative impact on scoring is more visible for weaker defending teams. It may be neutral against strong teams (FC Bayern Munchen).
- Long balls and corners played inside the box (set play cross) have similar negative impact pattern on scoring (but with lower statistical significance), suggesting that alternative play that keeps the possession of the ball can be more optimal.

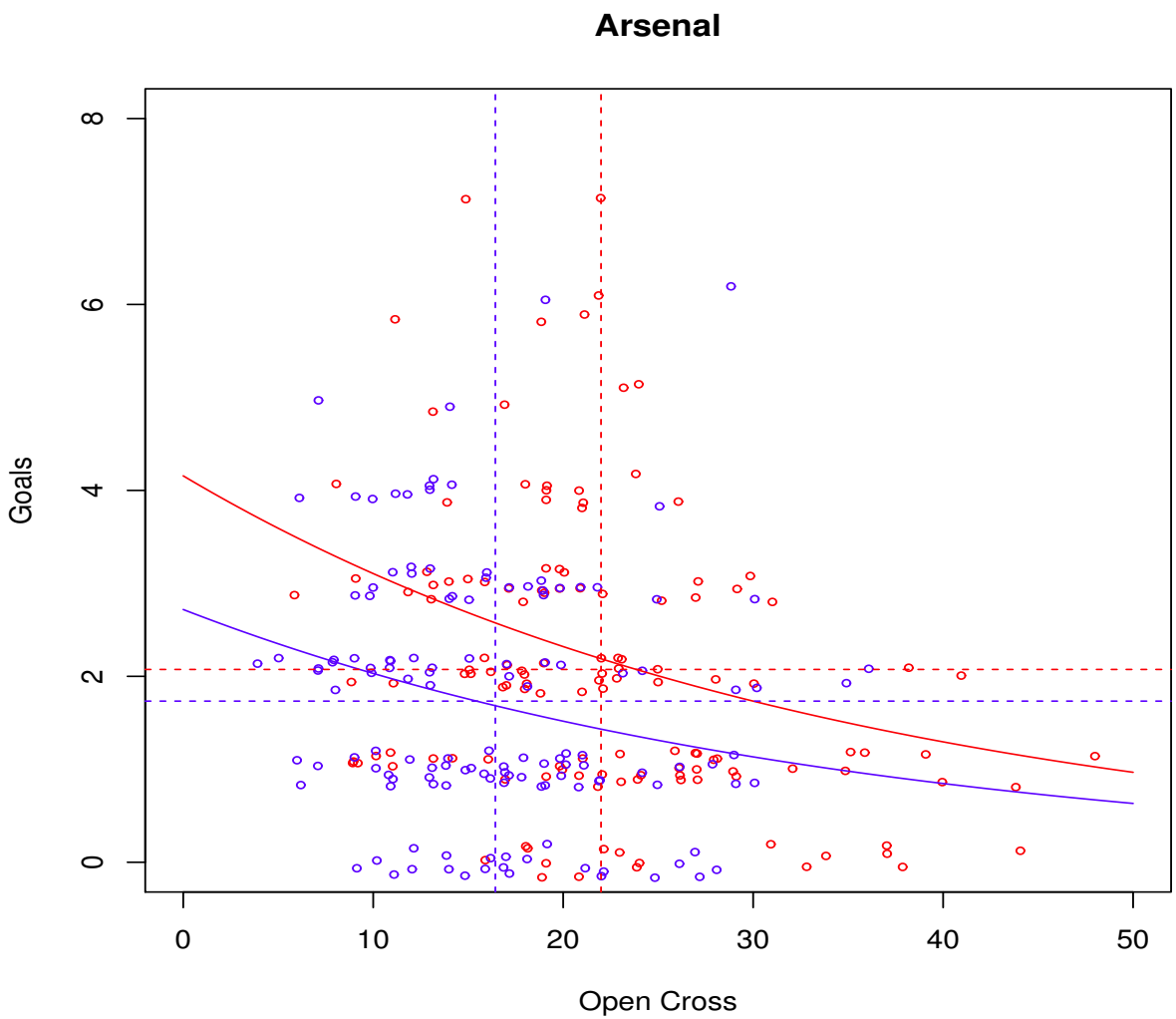
Tottenham Attack



Arsenal Attack

Crossing in soccer has a strong negative impact on scoring:

Jan Vecer, Frankfurt School of Finance and Management



Manchester United Attack

Crossing in soccer has a strong negative impact on scoring:

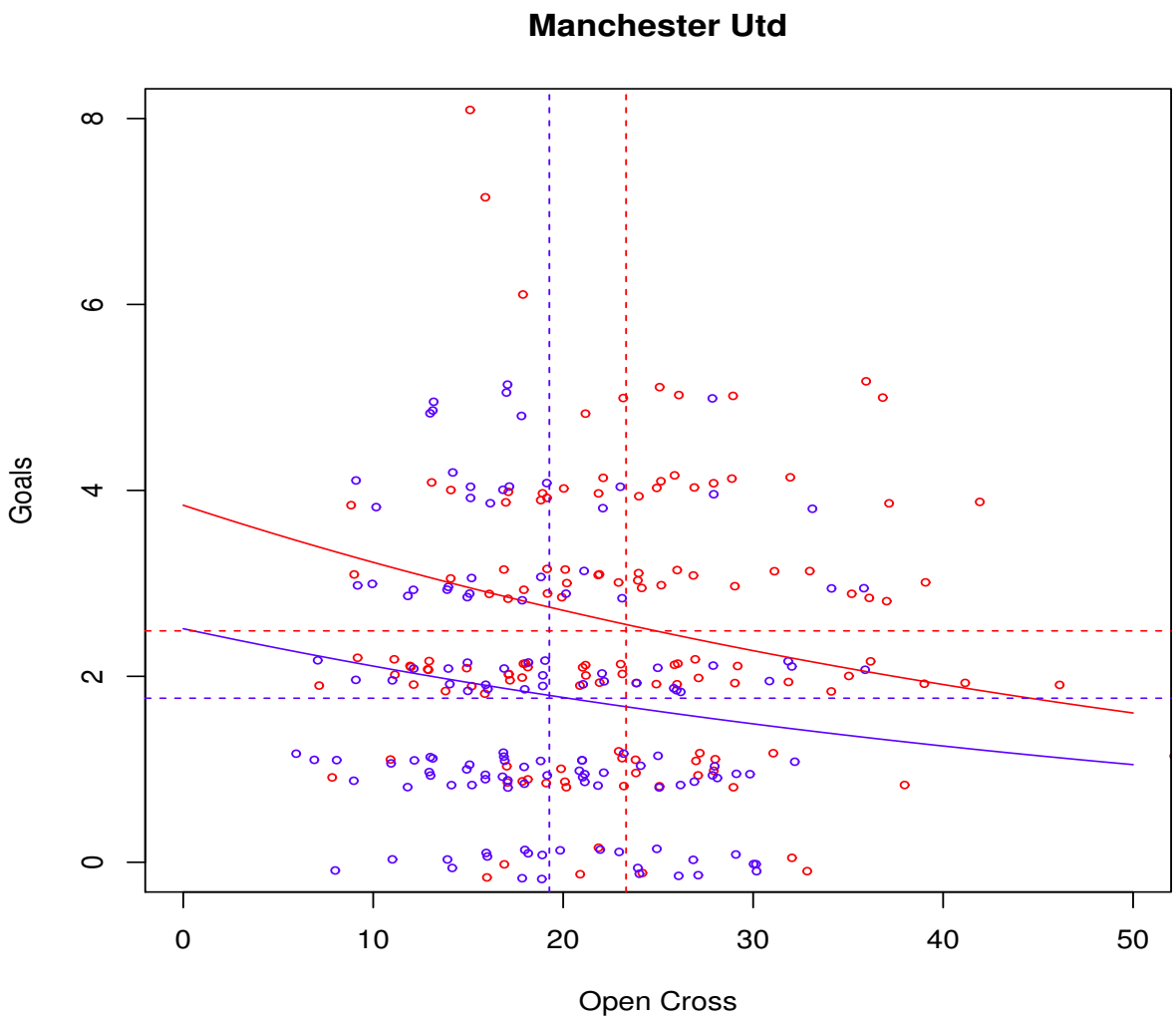
Jan Vecer, Frankfurt School of Finance and Management

Motivation

The Statistical Model

Graphs

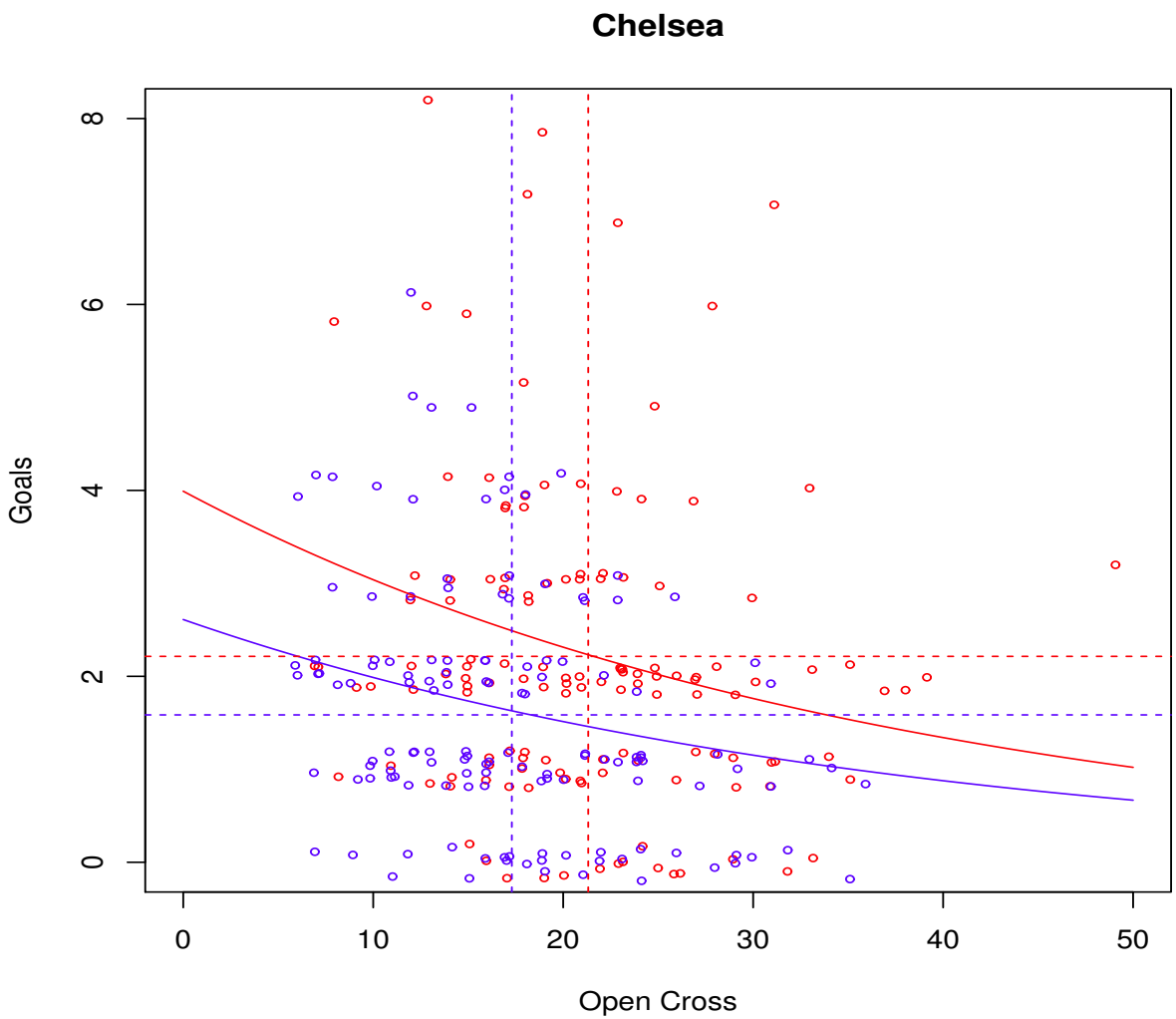
Conclusions



Chelsea Attack

Crossing in soccer has a strong negative impact on scoring:

Jan Vecer, Frankfurt School of Finance and Management



Liverpool Attack

Crossing in soccer has a strong negative impact on scoring:

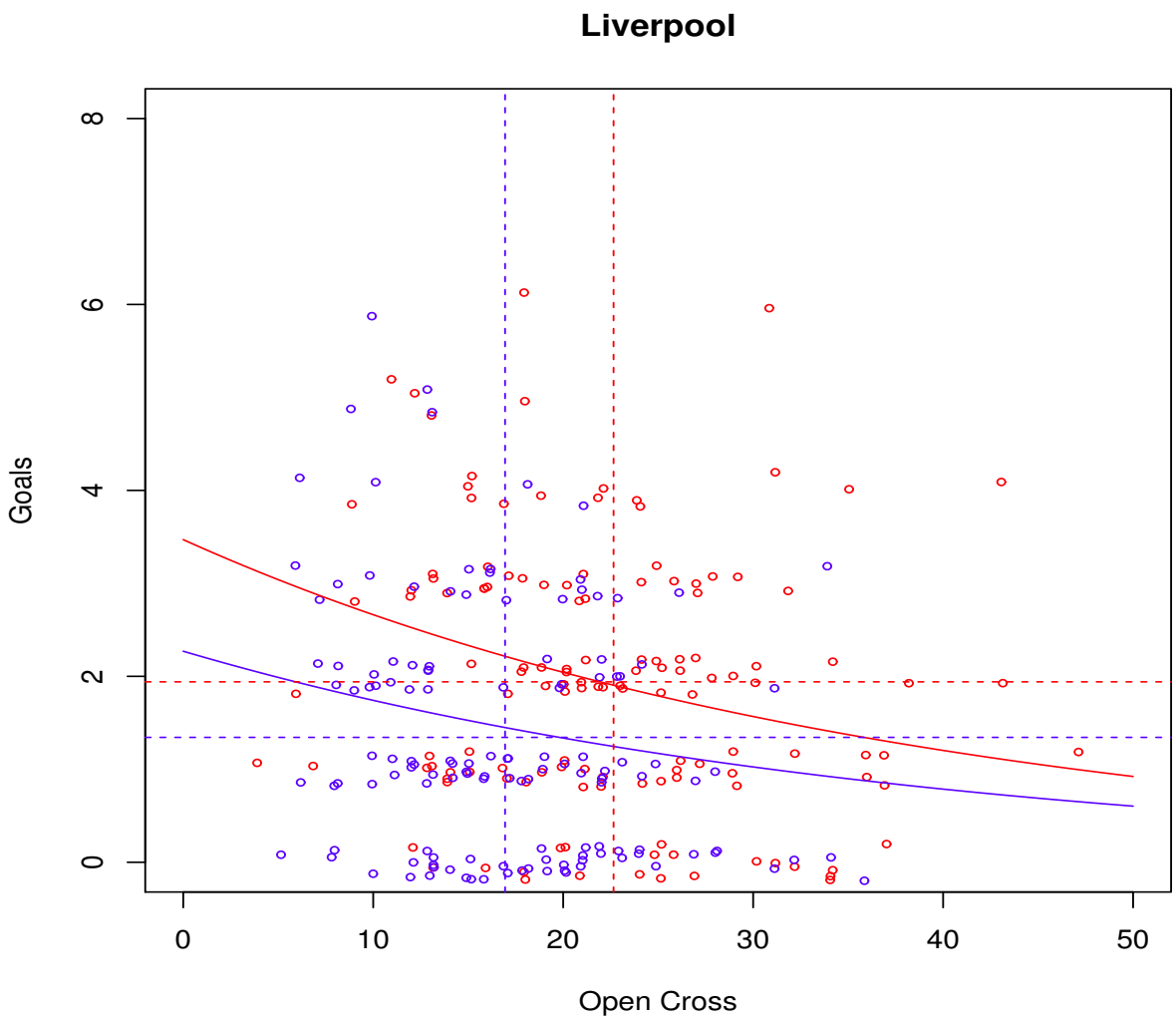
Jan Vecer, Frankfurt School of Finance and Management

Motivation

The Statistical Model

Graphs

Conclusions



Manchester City Attack

Crossing in soccer has a strong negative impact on scoring:

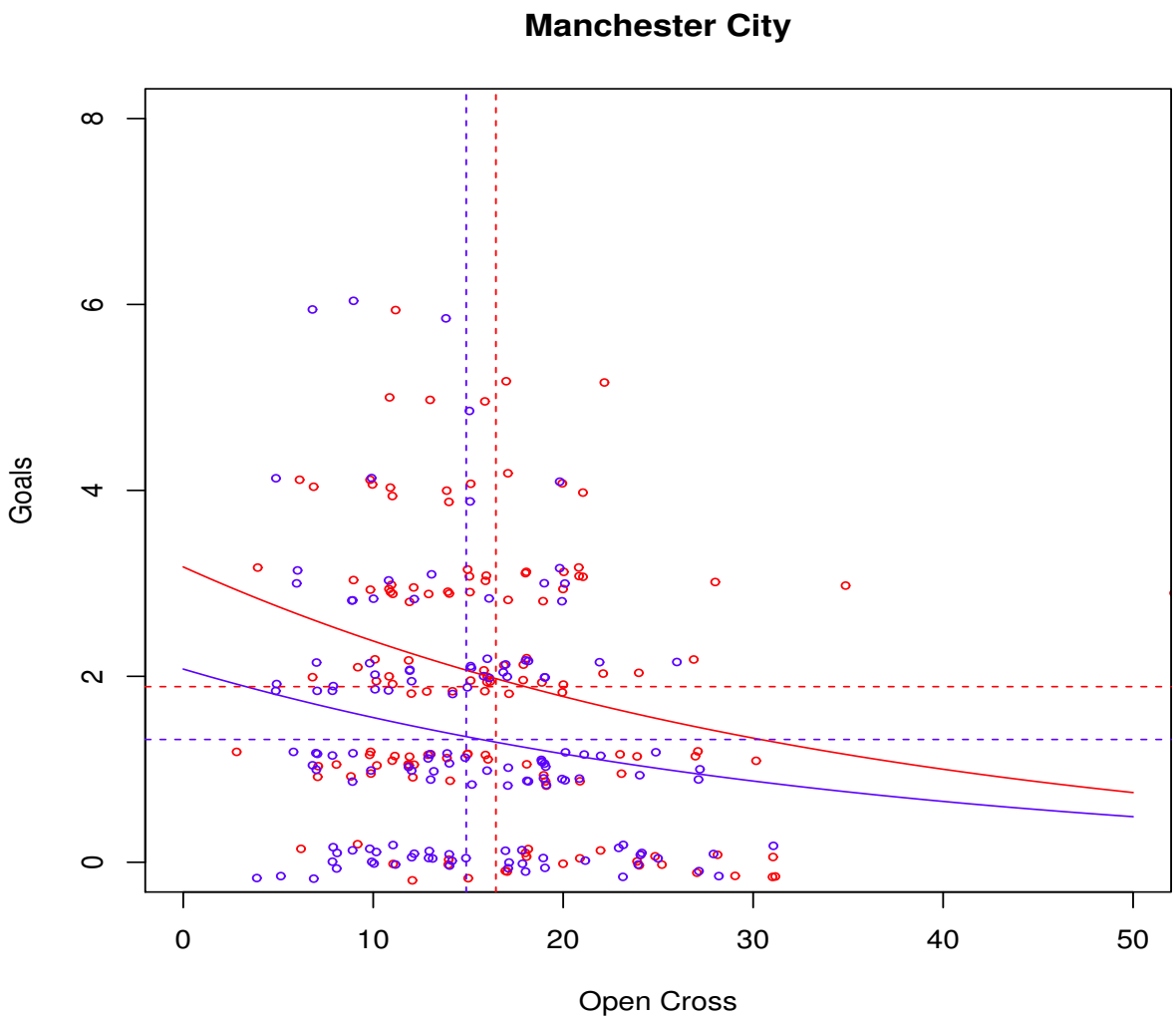
Jan Vecer, Frankfurt School of Finance and Management

Motivation

The Statistical Model

Graphs

Conclusions



Stoke City Attack

Crossing in soccer has a strong negative impact on scoring:

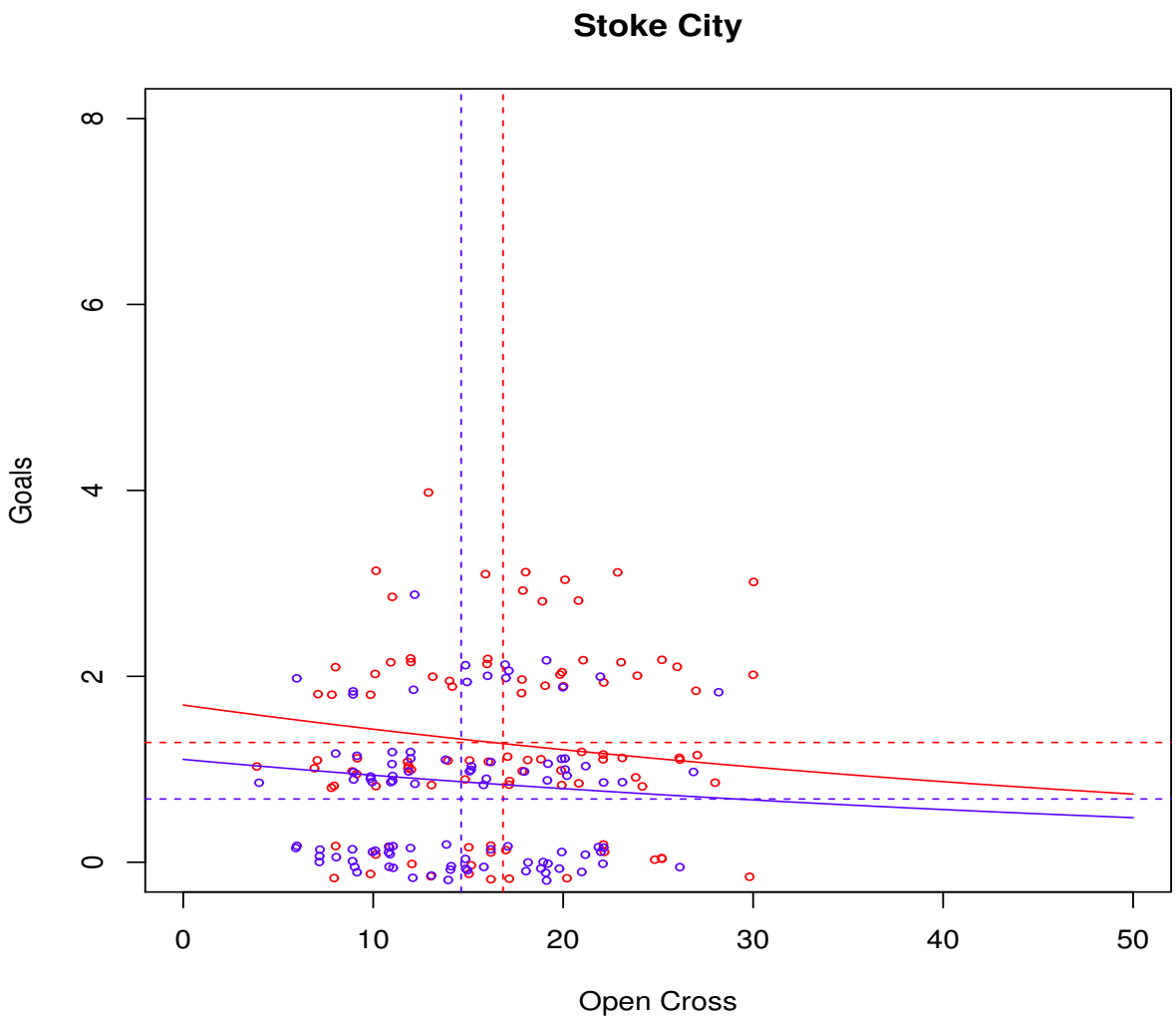
Jan Vecer, Frankfurt School of Finance and Management

Motivation

The Statistical Model

Graphs

Conclusions



FC Bayern Munich Attack

Crossing in soccer has a strong negative impact on scoring:

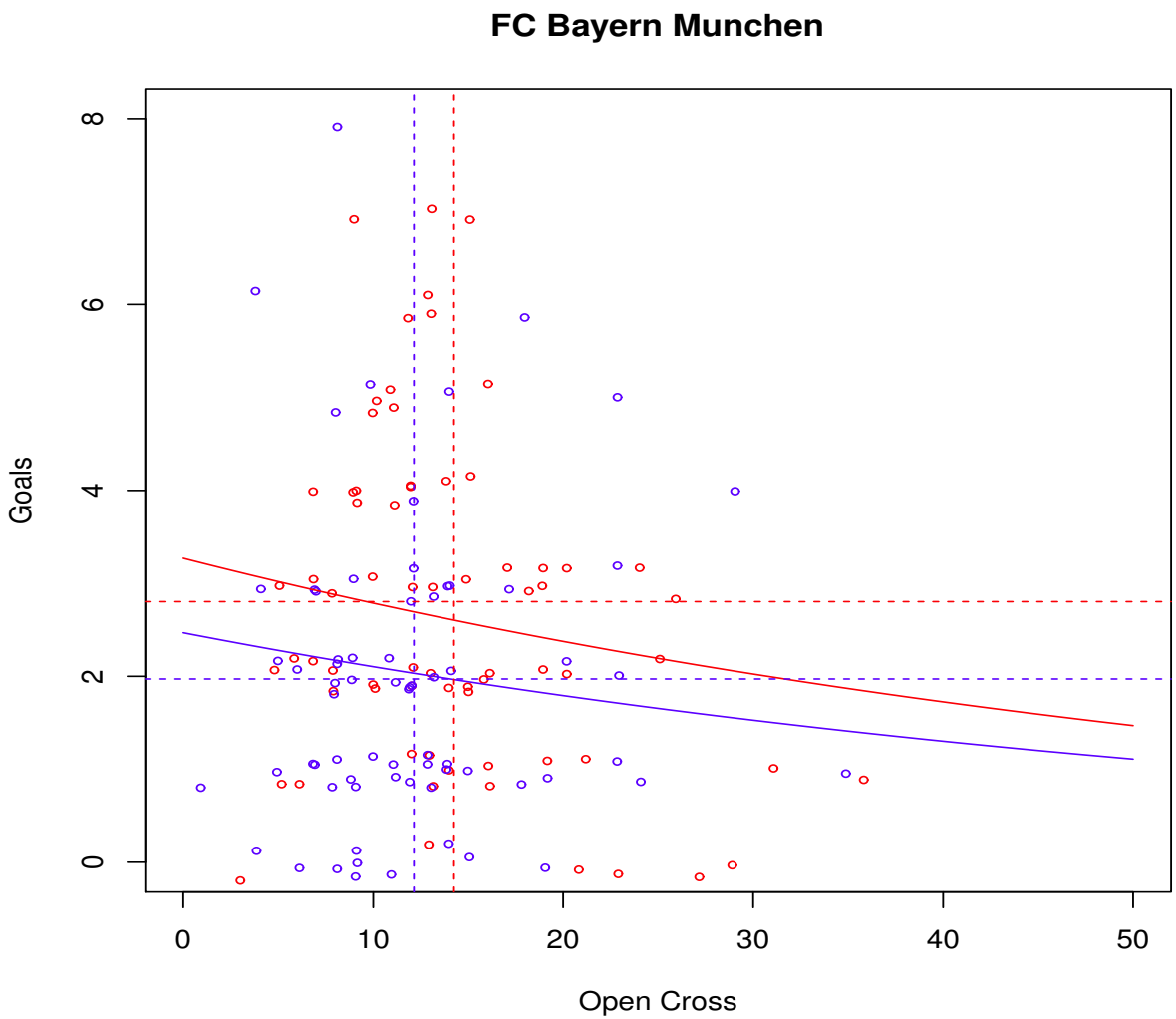
Jan Vecer, Frankfurt School of Finance and Management

Motivation

The Statistical Model

Graphs

Conclusions



Borussia Dortmund Attack

Crossing in soccer has a strong negative impact on scoring:

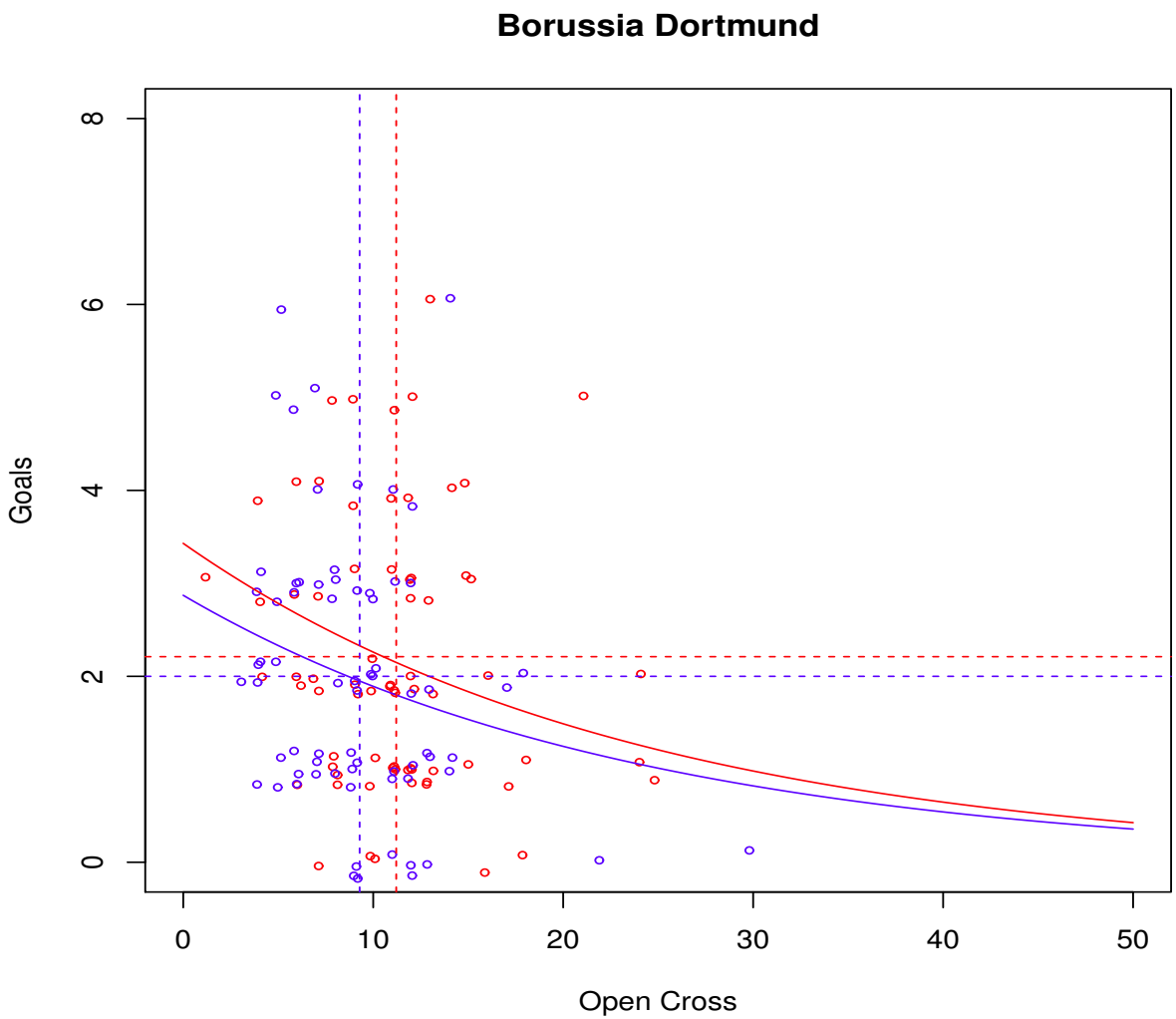
Jan Vecer, Frankfurt School of Finance and Management

Motivation

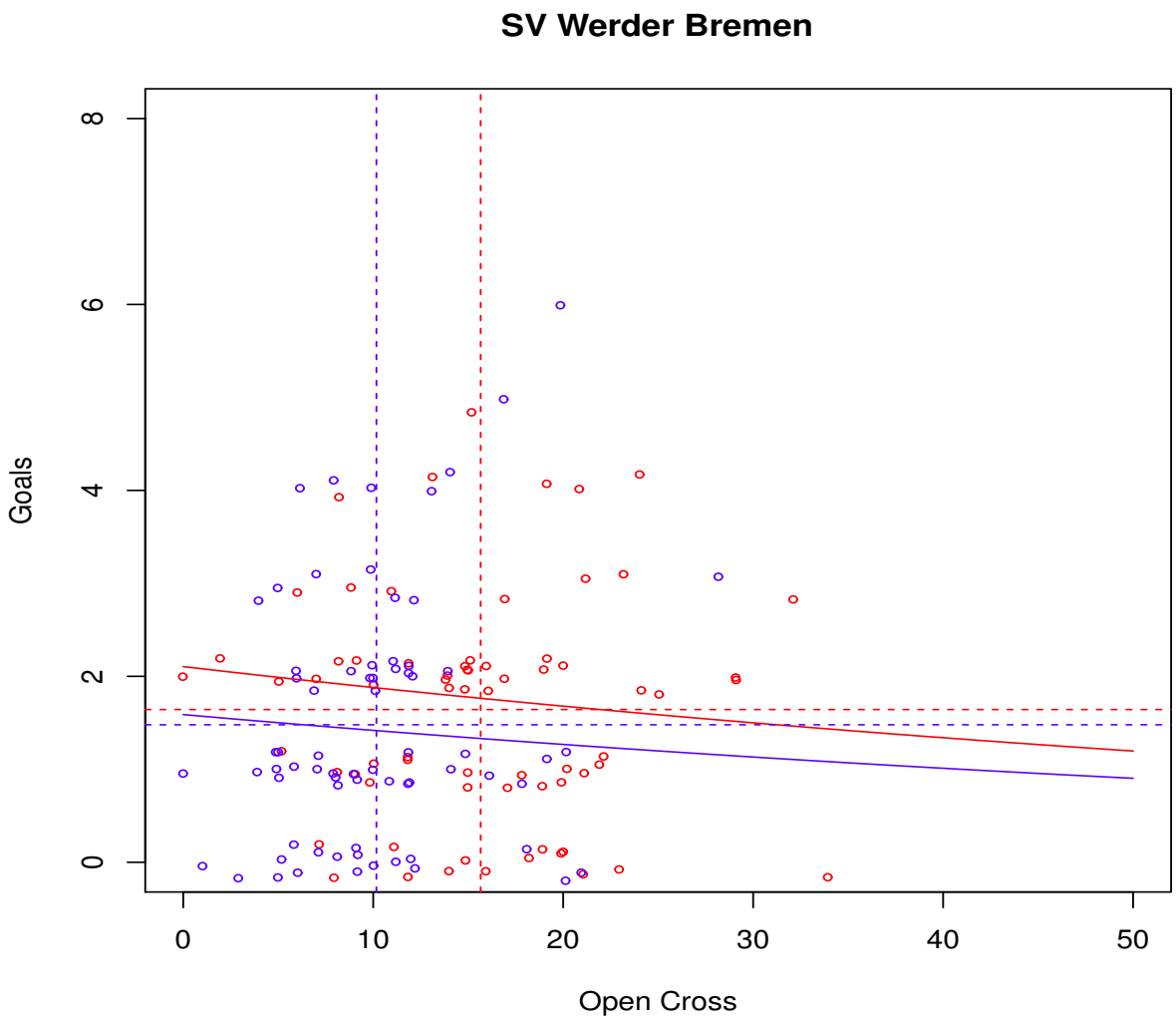
The Statistical Model

Graphs

Conclusions



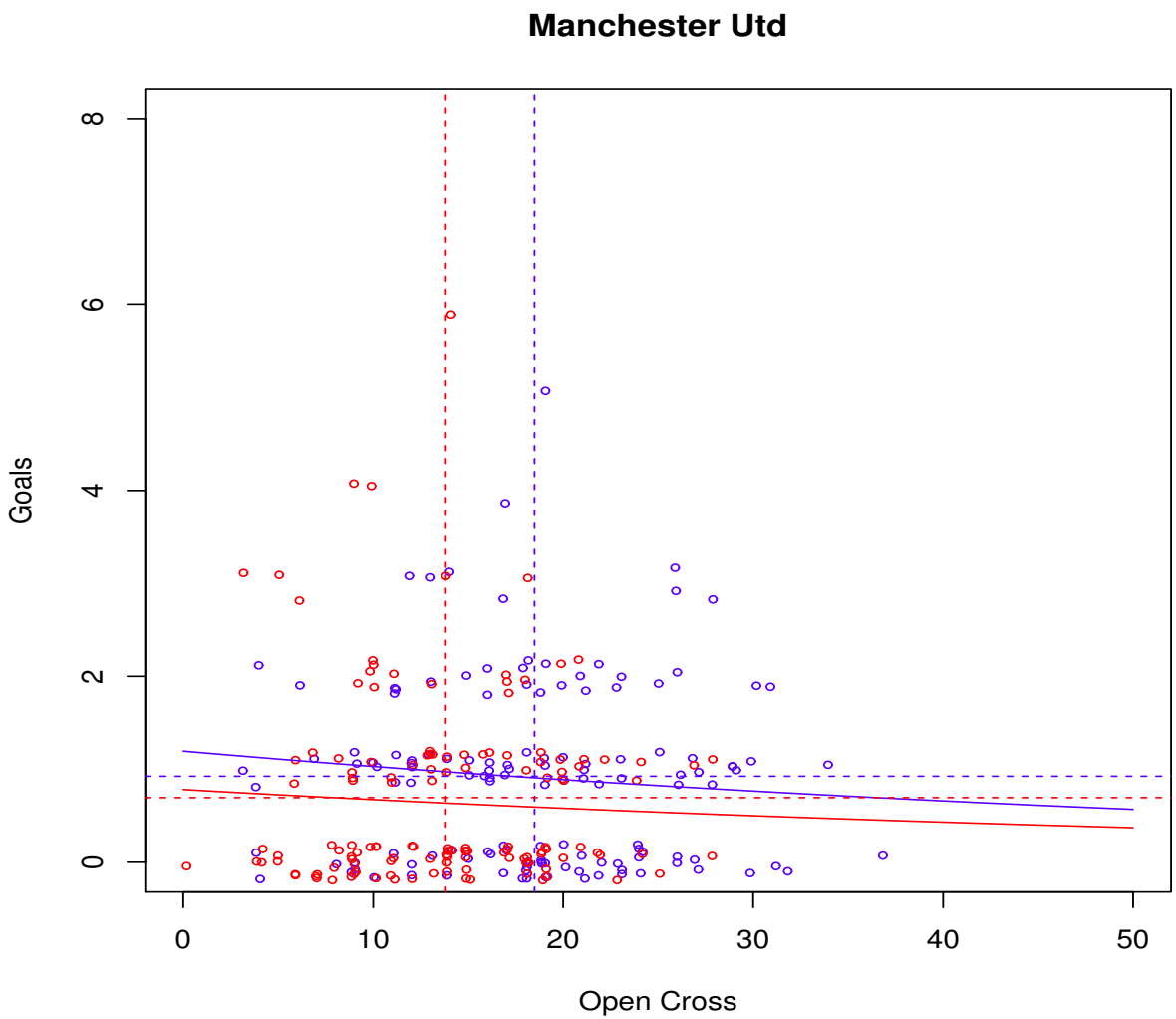
Werder Bremen Attack



Manchester United Defense

Crossing in soccer has a strong negative impact on scoring:

Jan Vecer, Frankfurt School of Finance and Management



Tottenham Defense

Crossing in soccer has a strong negative impact on scoring:

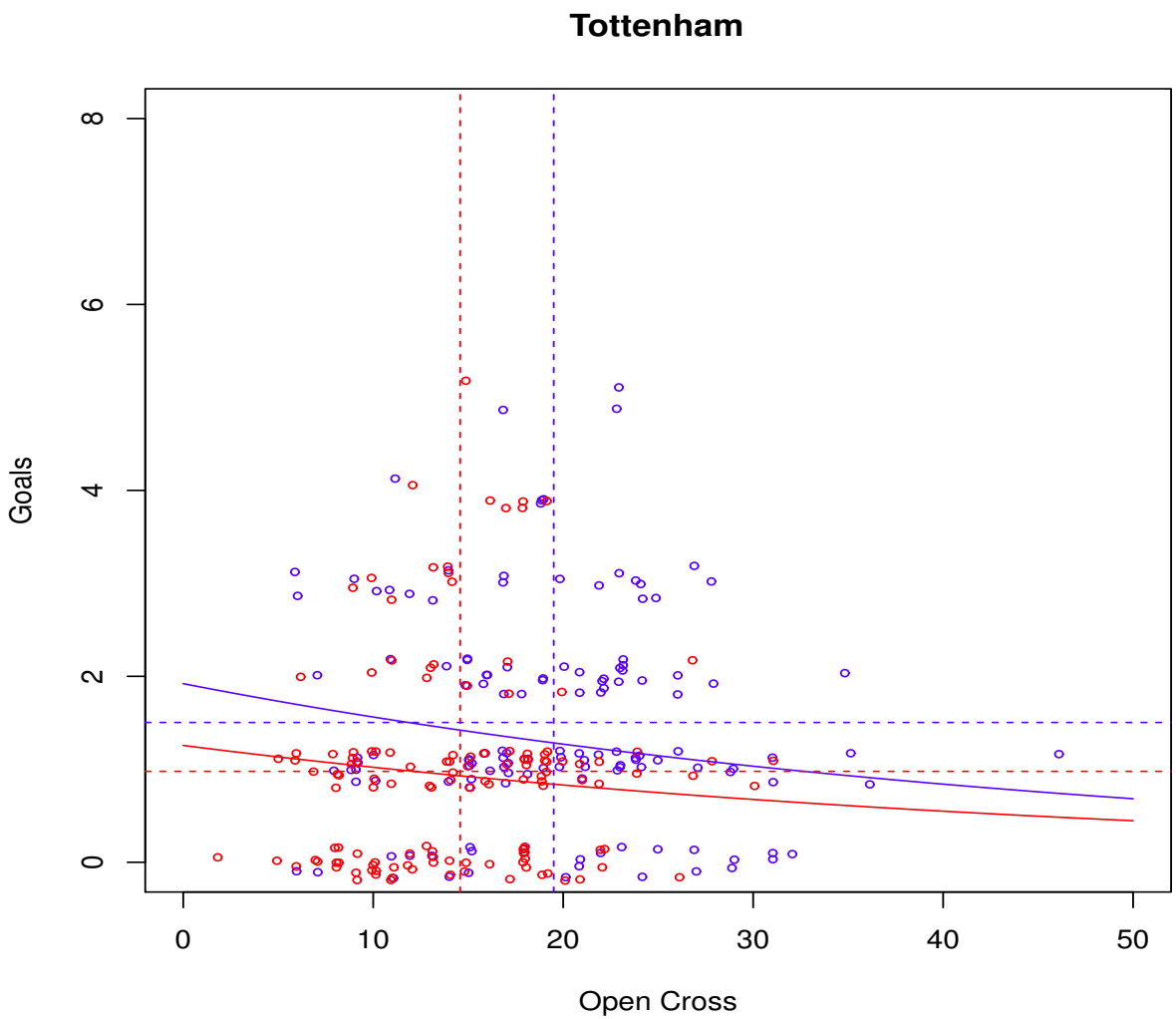
Jan Vecer, Frankfurt School of Finance and Management

Motivation

The Statistical Model

Graphs

Conclusions



Arsenal Defense

Crossing in soccer has a strong negative impact on scoring:

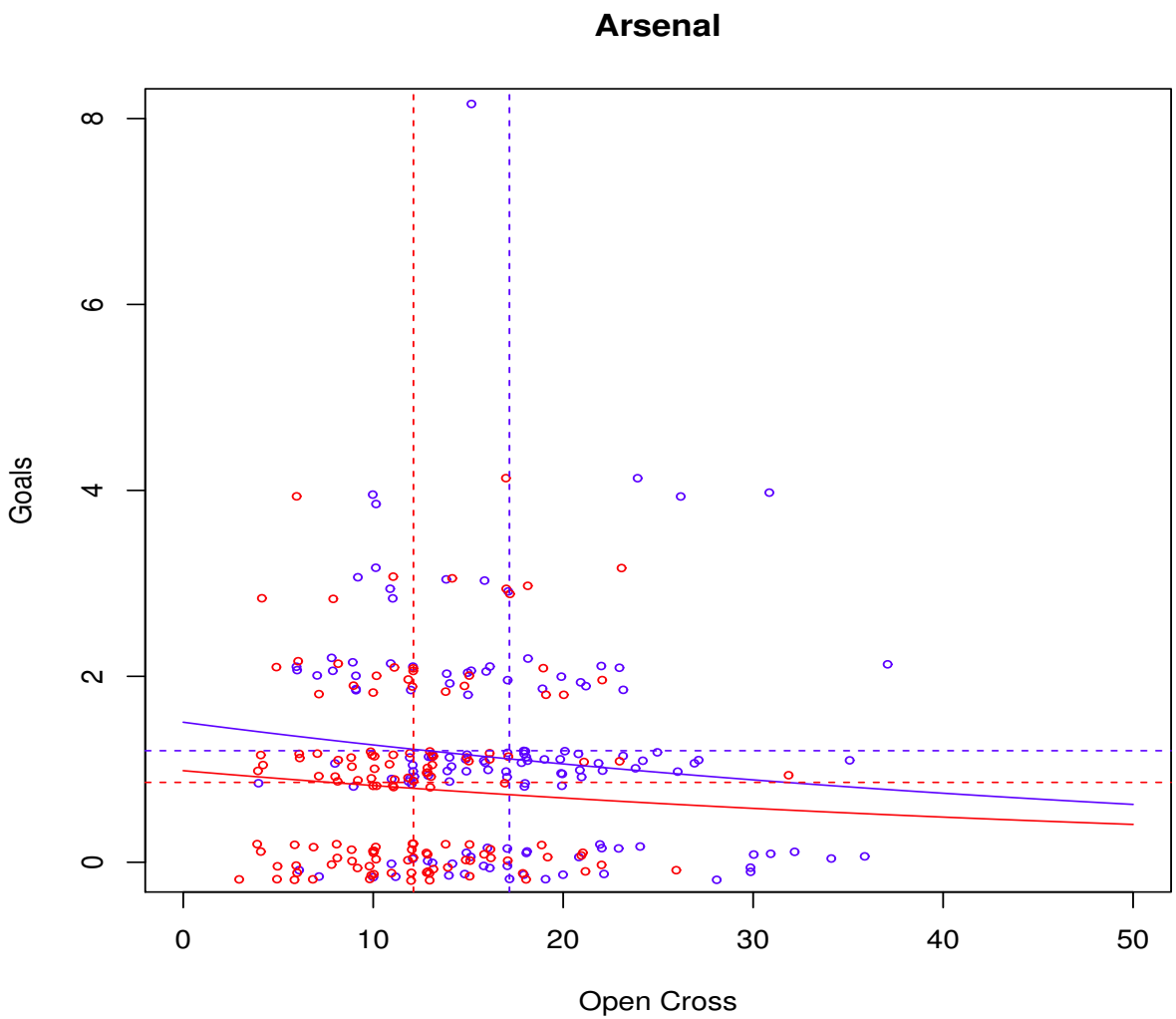
Jan Vecer, Frankfurt School of Finance and Management

Motivation

The Statistical Model

Graphs

Conclusions



Chelsea Defense

Crossing in soccer has a strong negative impact on scoring:

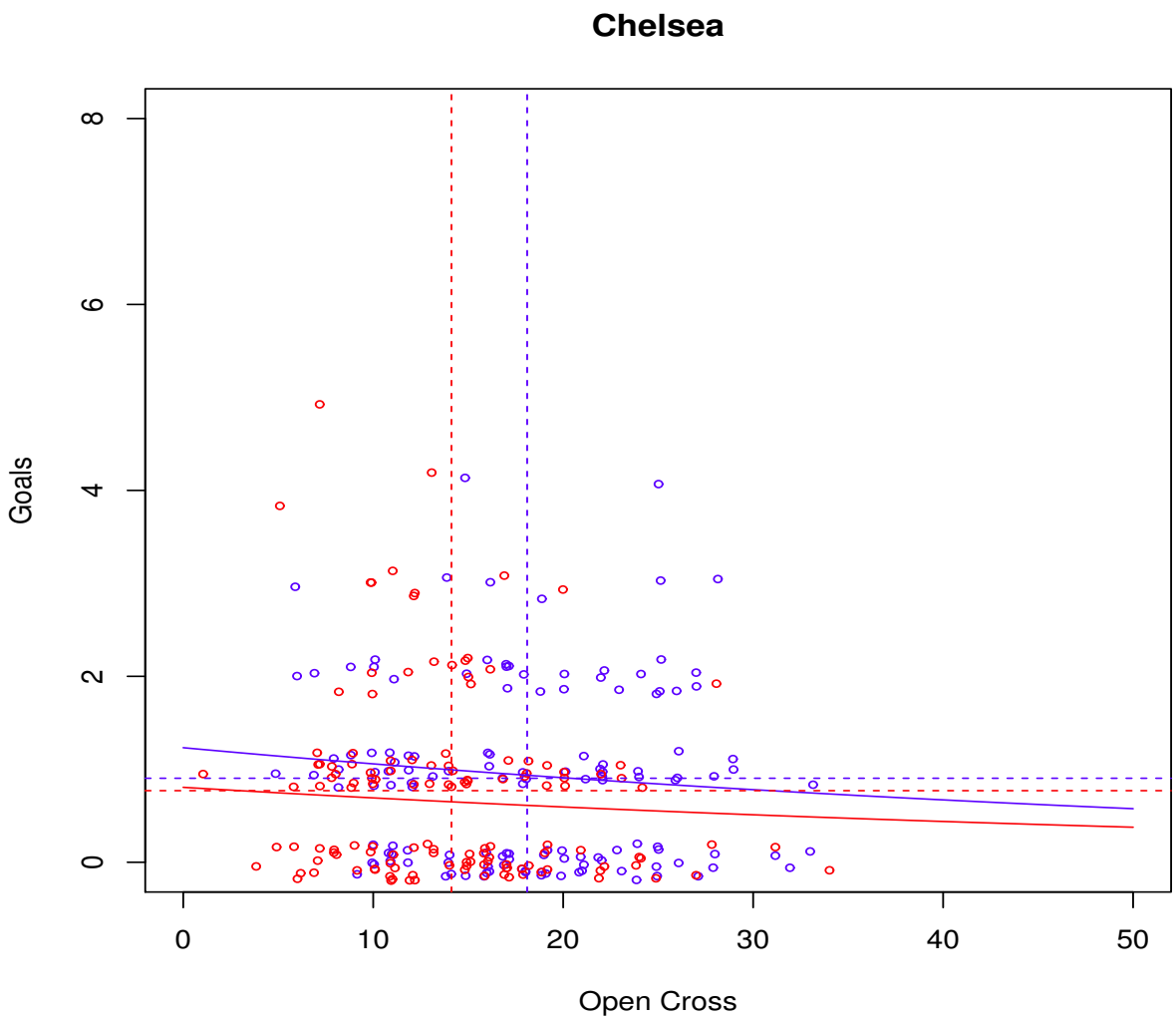
Jan Vecer, Frankfurt School of Finance and Management

Motivation

The Statistical Model

Graphs

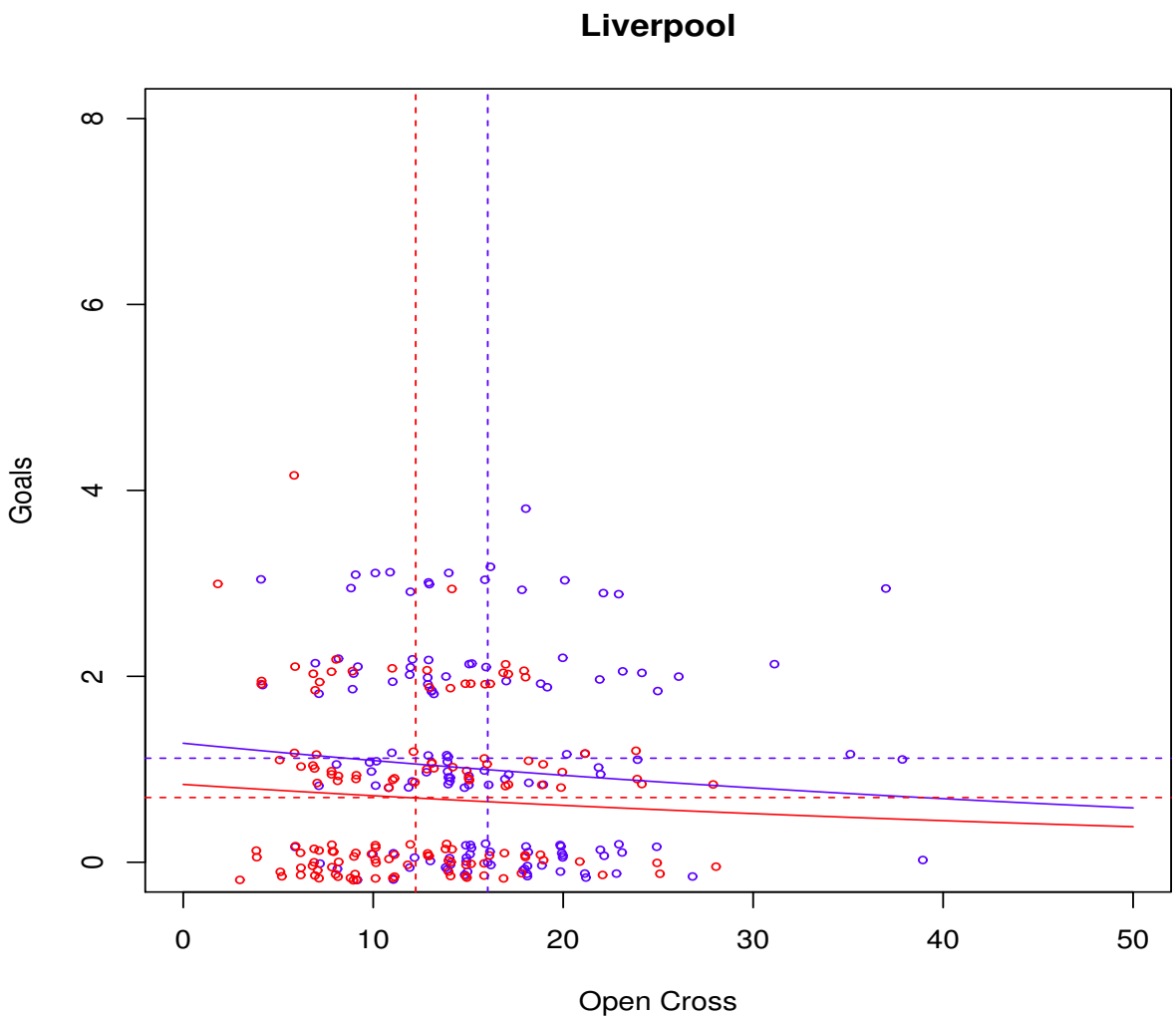
Conclusions



Liverpool Defense

Crossing in soccer has a strong negative impact on scoring:

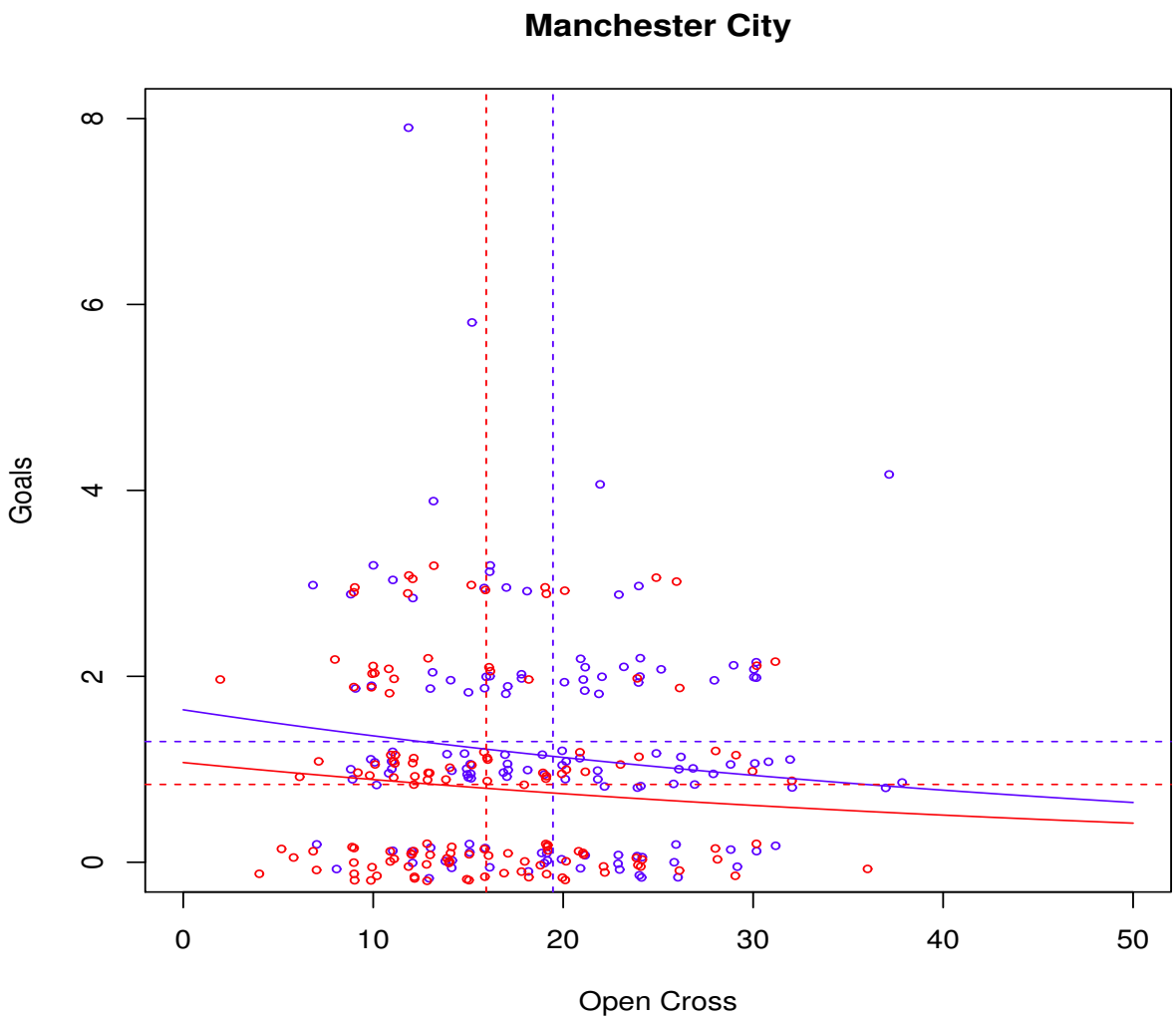
Jan Vecer, Frankfurt School of Finance and Management



Manchester City Defense

Crossing in soccer has a strong negative impact on scoring:

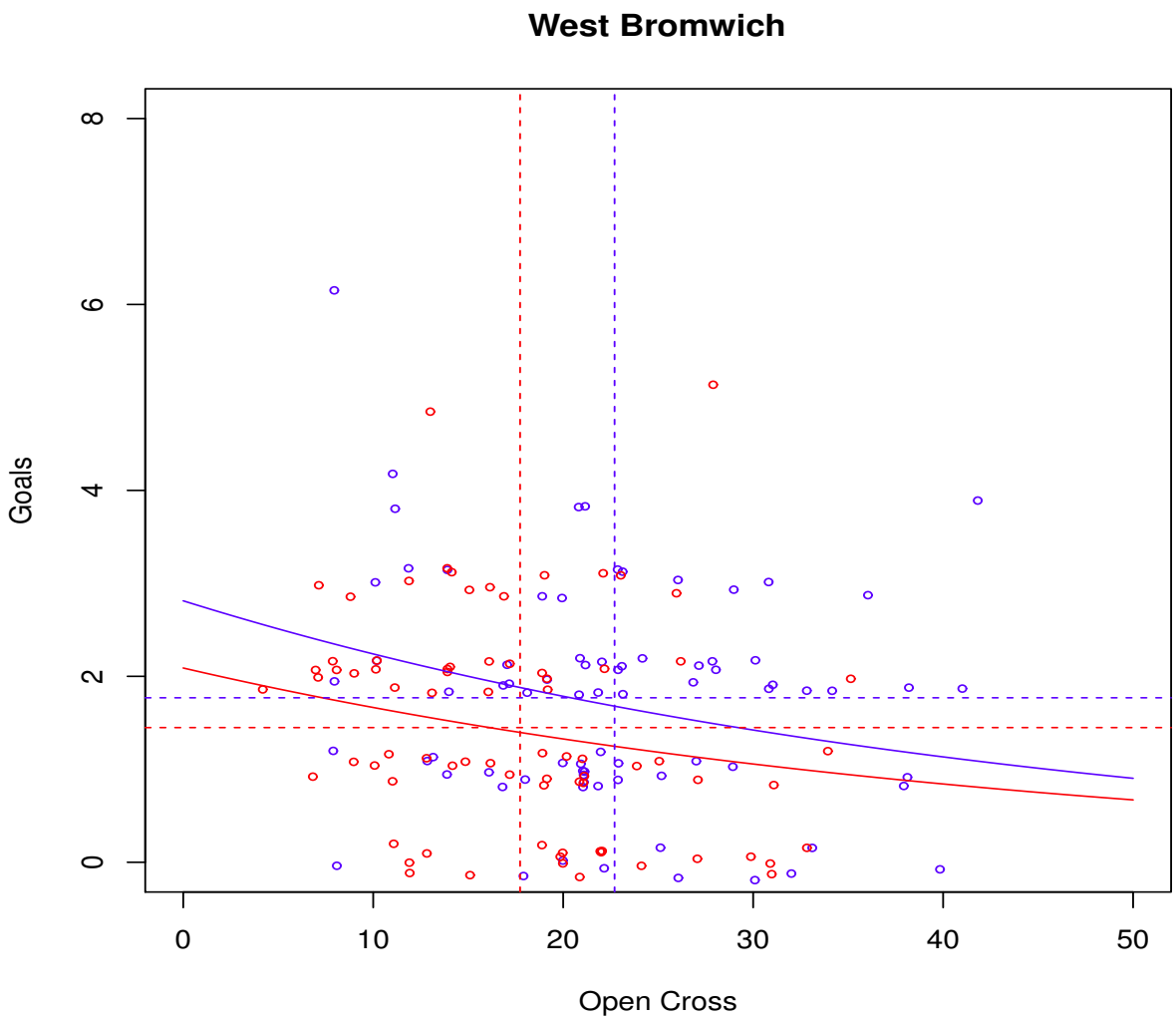
Jan Vecer, Frankfurt School of Finance and Management



West Bromwich Defense

Crossing in soccer has a strong negative impact on scoring:

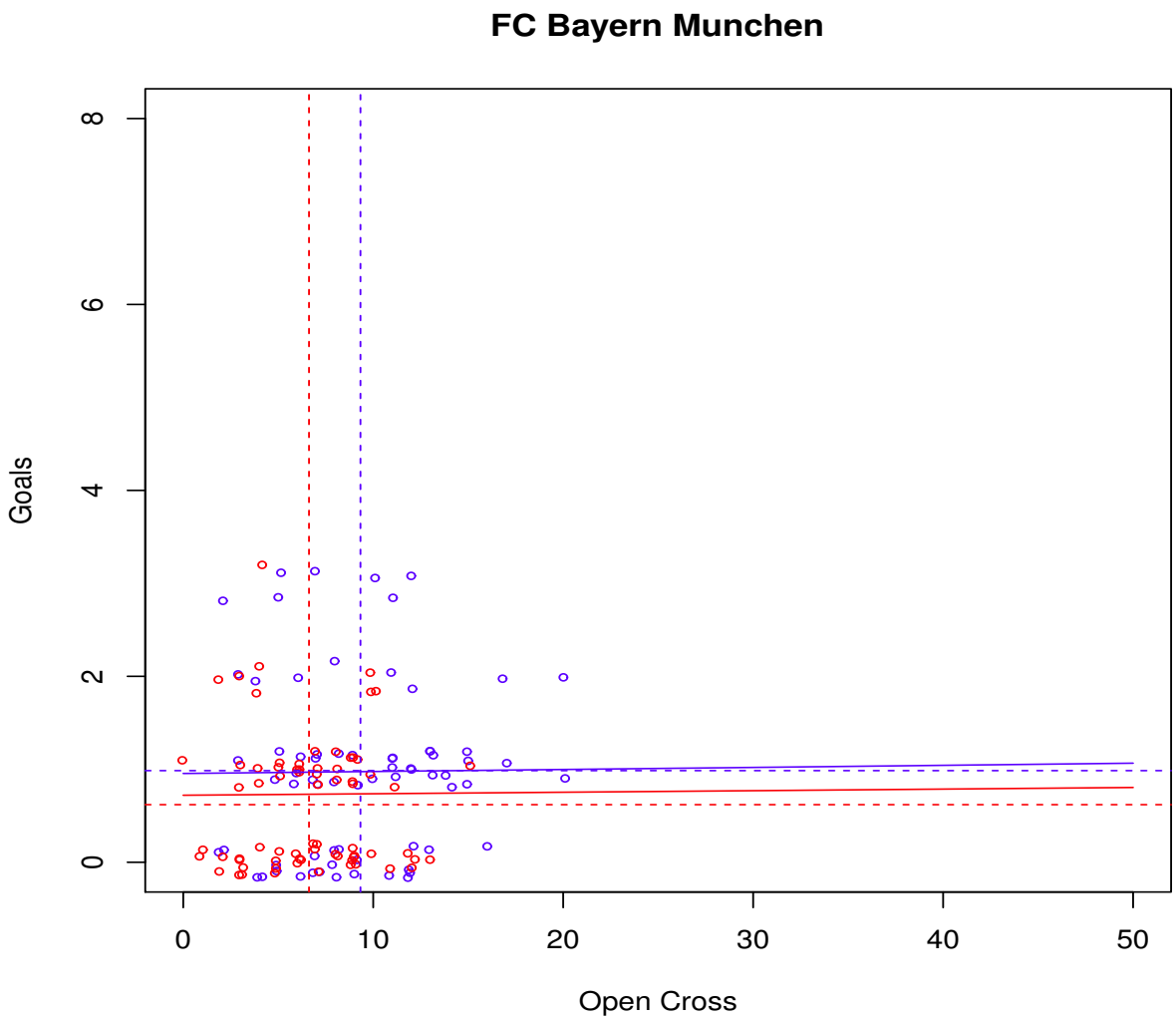
Jan Vecer, Frankfurt School of Finance and Management



FC Bayern Munich Defense

Crossing in soccer has a strong negative impact on scoring:

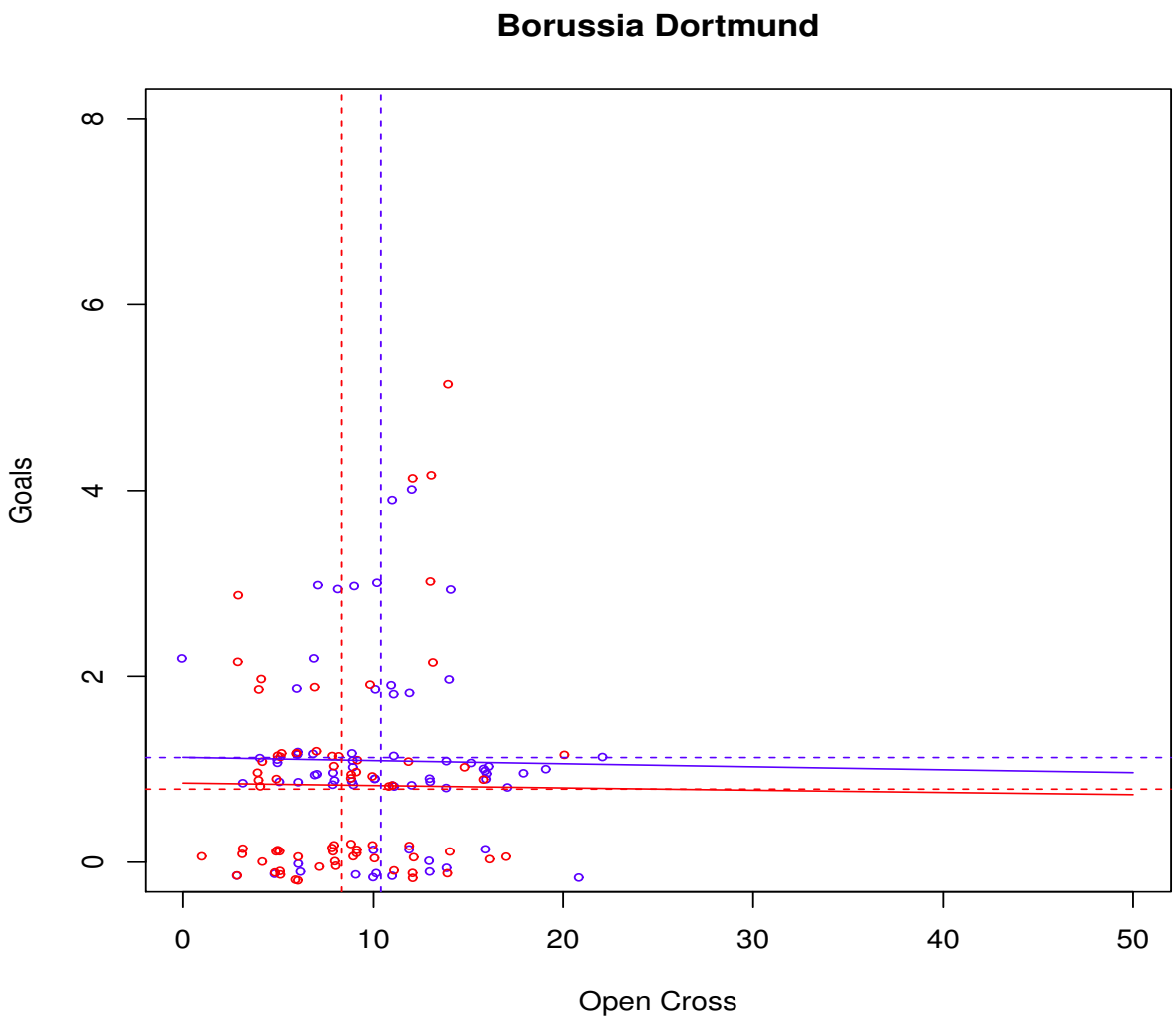
Jan Vecer, Frankfurt School of Finance and Management



Borussia Dortmund Defense

Crossing in soccer has a strong negative impact on scoring:

Jan Vecer, Frankfurt School of Finance and Management



Werder Bremen Defense

Crossing in soccer has a strong negative impact on scoring:

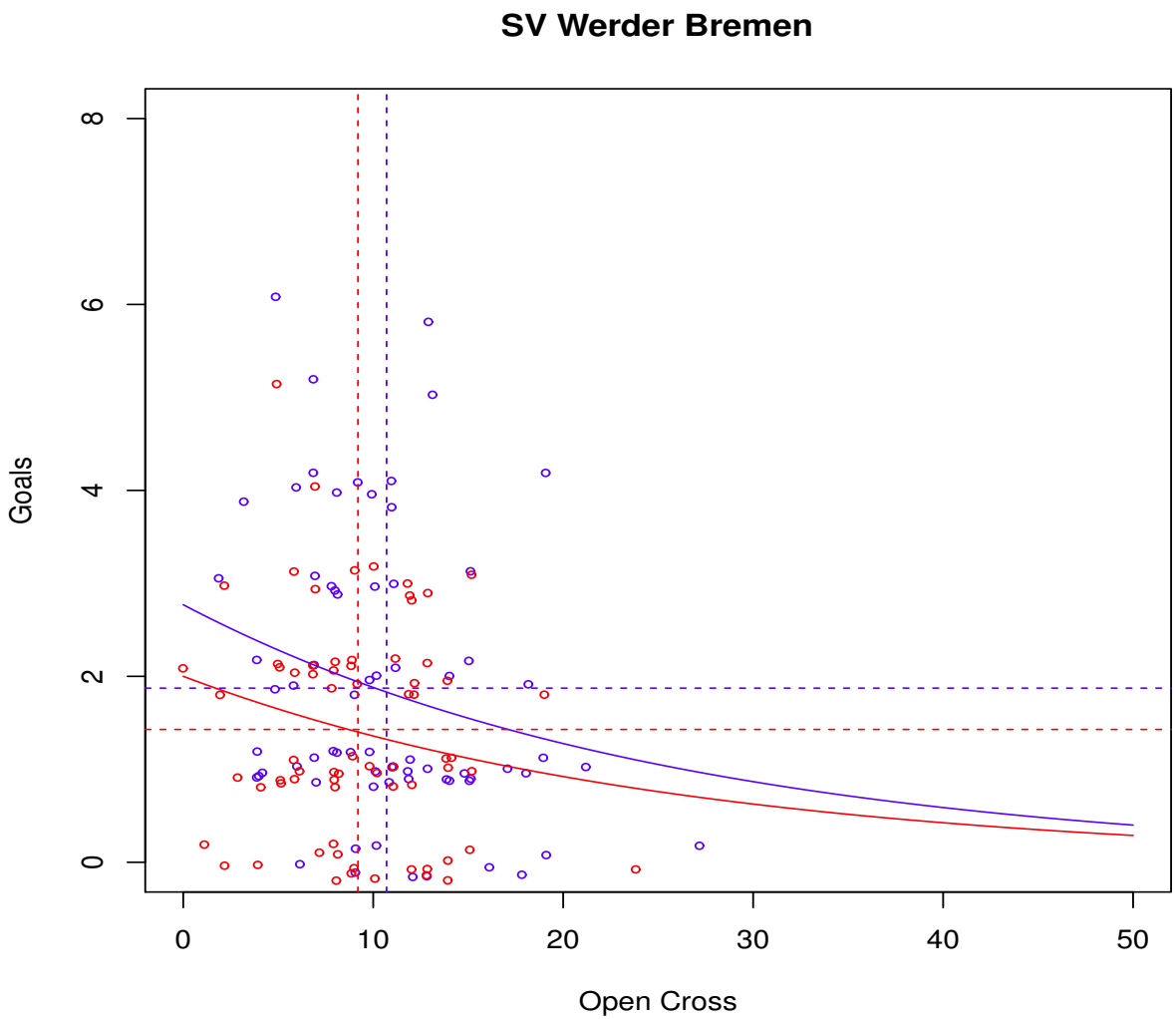
Jan Vecer, Frankfurt School of Finance and Management

Motivation

The Statistical Model

Graphs

Conclusions



Conclusions

Is open crossing dead?

No, but it should be either used by weaker teams playing against stronger teams when the luck plays a more important role, or the stronger teams must improve the crossing quality to the point of Manchester United (43.8 open crosses per goal) to make it neutral. This would need a big improvement, the second best crossing team, Chelsea, needs 62.6 open crosses per goal.

At the present time, the teams seem to overuse open crossing. Its reduction can increase scoring for most of the teams. Some top teams can score 40+ extra goals in seasons by reducing crossing. That's about how many goals scores Messi in his top season.

Conclusions

So what if we see indeed the decrease of open crossing and increase of scoring?

Do not forget to send me a check for such goals or for winning the championship, I am OK with half of Messi's salary.

I do not need the Golden Shoe for the best scorer.

Conclusions

I will post a new version of the paper on www.ssrn.com in the near future. An old version that uses a standard linear regression is available, but the conclusions are pretty much the same.