

What's wrong with *Soccernomics*?

A guide to its potential errors related to European club soccer

North Yard Analytics
July 2013

Executive summary

The book *Soccernomics* by Simon Kuper and Stefan Szymanski contains many potential errors in the interpretation and analysis of data related to professional soccer in Europe. Some of these errors could have costly consequences if acted upon by clubs. This paper offers an explanation of the issues of most interest to professional clubs.

Key observations

- The authors are likely to have overstated the causal relationship between players' wages and league position
- Profitability is not necessarily at cross-purposes with club performance, contrary to the authors' assertion
- Apparently overachieving managers may be less valuable than the authors' analysis would suggest
- Transfers may play a bigger role in clubs' success than the authors admit
- The authors may have underestimated the influence of penalties on the outcome of league games

Contact

Dan Altman, northyardanalytics@gmail.com

Table of contents

Introduction	3
Wages, transfers, and performance	4
The value of soccer managers	9
Financial management of professional clubs	12
City size and club success	16
Miscellany	19

Introduction

Soccernomics, by Simon Kuper and Stefan Szymanski, is a fascinating and often entertaining book. It is also a paradoxical one. The authors' message is that much of the conventional wisdom about global soccer can be upended by analyzing data. Yet it is precisely in the area of data analysis that the authors go astray, often contradicting themselves in their conclusions and even in the central themes of their book.

For example, consider the authors' writing about the appraisal of players. Among the lessons of the data revolution in baseball was that even professional observers might have a mistaken idea of what mattered in the game. Baseball players who looked good doing their jobs weren't necessarily doing their jobs well. The authors make a similar statement about the soccer midfielder Claude Makelele:

p. 153: "If you watched the game, you could miss Makelele. If you looked at the data, there he was."

Yet earlier, the authors seemed to suggest that such misjudgments were unlikely:

p. 93 "Anyone who knew soccer could judge fairly easily how good a player was just by watching him play."

If the latter statement were true, then there would be little reason to use data in evaluating soccer players.

North Yard Analytics believes that data do have an important place in soccer, as long as they are used carefully and correctly. What follows is one possible accounting of the authors' departures from factual, mathematical, and statistical rigor. It focuses on the first half of the book, which deals mainly with professional soccer as played by clubs across Europe – the primary area of interest for North Yard Analytics as well.

This exercise will not come as a surprise to the authors. On p. 90, they write that "reviewers delight in finding errors, no matter how trivial." In this case, some of the errors are not trivial. Professional soccer may not be as big an industry as semiconductors or cars, but millions of dollars, pounds, and euros are still involved in many transactions. The analysis presented here calls into question some of the guidance *Soccernomics* has offered, intentionally or otherwise, for the management of those monies.

Author's note: At times this text may appear overly technical. Readers should feel free to skip over details that do not interest them. These details are provided only for the sake of transparency.

Wages, transfers, and league position

One of the authors' main findings is that spending on players' wages is highly correlated with league position at the end of the season for clubs in England. In fact, they assert that higher wages cause higher finishes. They also assert that spending on wages is much more effective in producing higher finishes in the league than spending on transfers. Much of their argument may be wrong.

* * *

p. 13 “In short, the more you pay your players in wages, the higher you will finish.”

This statement, on which several early passages of the book are based, draws a causal link between wages and team performance. It is unlikely to be true, at least by itself. If you were to take a team's current players and double their salaries, would the team score twice as many goals? Economists would call this an “efficiency wage”, but examples of efficiency wages can be hard to find. If the statement above were true, you could take a bunch of no-hopers, pay them Premier League wages, and they'd be a match for any Premier League team. Clearly that's not the case. More likely, the authors mean something else.

The authors appear to take for granted that teams will always get value for money when they pay high wages. Thus, a team paying higher wages will simply have better players. This is not necessarily true, either, as evidenced by the recent example of Queens Park Rangers, relegated in 2013 despite sky-high salaries. As the authors indicate on p. 17, players may choose to leave a club if they don't feel that they're being paid a fair wage (though they also write on p. 41 that “unknowns accept modest salaries,” apparently independent of their skills). But players are also unlikely to complain if clubs overpay them.

The causation could also run in the opposite direction. In 2013, the payoffs to Premier League clubs differed by more than £14 million between first place and last place. Taking into account the payoffs from qualification for European competitions, the difference could be more than £34 million. Shirt sales from star players and relegation for the bottom clubs open this gap even further. As a result, performance (which is fairly persistent from year to year) could indeed lead to higher salaries. This is a chicken-and-egg problem. **Clubs should not conclude that paying higher wages automatically leads to better finishes.**

* * *

p. 88 “Clearly the market in players' pay was highly efficient: the better a player, the more he earned. And this made sense, because soccer is one of the few markets that indisputably meets the

conditions in which competition can work efficiently: there are large numbers of buyers and sellers, all of whom have plenty of information about the quality of the players being bought and sold.”

This time, the authors are essentially arguing that if high salaries buy results, then high salaries must buy good players; they use the word “efficient” to imply that price always corresponds to quality. But if the causality between salaries and results goes in the other direction, the argument falls apart.

Moreover, several of the authors’ statements here are wrong. For the top players, who command weekly salaries of hundreds of thousands of pounds, there are very few buyers and sellers. How many teams could afford Neymar, and how many Neymars are there in the market at any given time? When only three or four teams are trying to buy the services of three or four top players, it’s much more likely that efficiency will suffer.

The same goes for players scouted from faraway leagues. A club like Arsenal might make the effort to find “unknown” players in the domestic leagues of Sub-Saharan Africa, but how many other clubs are competing for these players’ signatures? More likely, one club may come calling with a deal that the player can’t refuse; he may have no other options outside his home country. In fact, most other clubs in the Premier League may not even know he exists, and he will be undervalued as a result.

It is also likely that players have more information about their future performance than clubs do. Hiring players is a little like buying a used car; you can do a pretty thorough check-up and still end up with a lemon. There are several areas in which player might have a better idea of his future physical, psychological, and logistical aptitude than his prospective employers: does he have a nagging injury that’s getting worse; is he planning to have children; are his relatives ill; will he sign up for a national team that is likely to select him; etc. If clubs are signing players based on a few video clips, as the authors suggest in passing, these potentially important items may elude them.

Finally, clubs may not pay players just to win games. In the past, clubs have allegedly brought players into their squads to sell shirts (if they were famous but over the hill) or to open up new markets (as in the case of some East Asian players). **There are many reasons to suspect that the market for players is far from efficient in translating money into results.**

* * *

p. 43 “Some clubs – notably Barcelona for most of this century – understand the inefficiencies of the transfer market and therefore avoid buying players wherever possible.”

Without making a direct comparison based on data, the authors assert that players' wages are determined efficiently while transfer fees for the same players are not. The example of Barcelona is also a poor one. Barcelona's €50 million purchase of Zlatan Ibrahimovic' in 2009 (which also sent Samuel Eto'o to Inter Milan) may have been an exception to "most of" here. He did score 21 goals in his one full season at the club, but then he was offloaded for only €24 million to AC Milan. This summer, Barcelona have splashed the cash again for Neymar, spending €57 million for a player whom Johan Crujff, who is often lauded by the authors, said the club did not need.

But these expensive purchases were hardly exceptions to the rule. In fact, according to the website *transfermarkt* (which the authors call a "reputable source" on p. 414), Barcelona had net spending on players of more than €380 million from 2001 through the end of the 2010-11 season, which the authors presumably could observe before publication in 2012. **Barcelona certainly did not avoid buying players, and in fact it incurred a deficit of roughly €35 million per year in the transfer market.**

* * *

p. 414 "[W]hen we said that spending on transfers bears little relation to where a club finishes, we were referring to *net* transfer spending – transfer fees paid minus transfer fees received. (We have clarified that point in this new edition.) The net figure is the crucial one, because hardly any clubs can just keep buying players without occasionally selling some to more or less balance the books."

Here the authors dispute the usefulness of the observation that gross transfer fees correlate just as well with league position as spending on wages. First, they argue that gross spending is not important, because most clubs also have to sell players. It's not clear why this notion (partially disproven earlier in this paper) makes the correlation less interesting. If gross spending is correlated with league position but net spending isn't, the conclusion seems to be that clubs that buy and sell more players – or pricier players – finish higher in the league. In other words, if we assign a causal relationship to the correlation, being active in the transfer market might lead to better results, holding net spending constant. The level of activity in the transfer market could indeed be an informative correlate with league position.

The notion that most clubs in the Premier League "more or less balance the books" is also mistaken. In the ten seasons from 2001-02 to 2010-11, to which the authors would presumably have had access, their source *transfermarkt*

reports the following transfer balances for the clubs in the Premier League at the end of that period:

<u>Club</u>	<u>Transfer balance in £, 2001-11</u>
Arsenal	-39,776,000
Aston Villa	-128,911,200
Birmingham City	-76,511,600
Blackburn	10,282,800
Blackpool	-3,182,960*
Bolton	-38,985,320
Chelsea	-525,219,200
Everton	-52,844,880
Fulham	-85,228,000
Liverpool	-160,009,520
Manchester City	-446,445,120
Manchester United	-163,402,800
Newcastle United	-55,248,160
Stoke City	-53,644,800*
Sunderland	-111,225,840
Tottenham	-249,638,400
West Bromwich Albion	-45,964,600
West Ham United	11,347,600
Wigan Athletic	-20,059,600
Wolves	-53,765,800

* may be incomplete

Only two teams had positive balances over the ten seasons, while seven lost at least £10 million a year on transfers. **Most Premier League clubs do not balance their transfer books.**

* * *

p. 110 “In 2010, the last year in our database, United’s share of the Premier League’s wage spending peaked at just over 10 percent. Yet even that wasn’t outsize. Manchester City that year spent about the same proportion, and Chelsea accounted for 14 percent of the Premier League’s total outlay on wages from 2004 through 2010.”

What does “outsize” mean here? If these three clubs indeed accounted for 34 percent of total wages in the Premier League, then the other seventeen would have had an average of less than 4 percent each. **Manchester United’s spending was actually more than 2.5 times the average spending by most of the other clubs in the division.**

* * *

p. 113 “Chelsea was the country’s biggest-spending club for six or seven years until the rise of Manchester City, and Abramovich is entitled to feel disappointed that his money bought him only three league titles.”

Once again, this passage depends on the notion that the line of causation between spending and league position runs in only one direction. As noted above, this notion may be tenuous. But even if it is correct, were three titles in six or seven years such a bad haul?

Even if wages explained 90 percent of the variation in league position, there was probably a pretty good chance that a team would finish one position higher – or lower – in a given year. But in a single season, the relationship is different. On p. 12, the authors put the correlation between league position and wages at 70 percent, or 0.7.

“Correlation” and “percent of variation” explained are two different statistical concepts; in fact, the latter is the square of the former. So if the correlation between league position and spending is 0.7, then the variation in either explained by the other is just 49 percent. That leaves 51 percent of the variation in league position for the errors in prediction described above; with twenty positions in the league, normally distributed errors will have a standard deviation of about 4.2.

This translates into a 45 percent chance that a generic team will finish at least one league position lower than its spending would have predicted in a given season – almost a coin flip. The chance of winning six titles in a row may have been as low as 3 percent, depending on just how much Chelsea spent relative to its nearest competitors. **Most likely, Abramovich can be very pleased that he won three titles in six or seven years as top spender in the Premier League.**

The value of soccer managers

Overall, the authors are fairly suspicious of managers' ability to enhance the results of their clubs. But they do identify several dozen managers whose clubs appear to have performed better than the rest in a consistent way. Additional analysis suggests these managers may simply have been the beneficiaries of random clusters of good results. Nevertheless, sacking managers may still have more effects than the authors admit.

* * *

p. 104 “There are twenty clubs in the Premier League, which means that the average club spends 5 percent of the league’s total wages. If a club spending 5 percent finished in the top half of the table, the manager was overachieving.”

Crediting any deviation from expected performance to the manager may be a mistake in itself. Putting that aside, the “average club” here probably doesn't exist. The average spending among clubs will naturally be 5 percent; no matter what their actual spending is, as a whole they will spend 100 percent of the total divided by twenty clubs. But individually, the middle club in spending – the median club – might spend much more or much less.

As a result, the second statement in the passage is also wrong. Say the top five clubs each account for 8 percent of total spending and the other fifteen each account for 4 percent. One of the latter can only be said to overachieve if it manages to finish in the top five, not just the top half of the table. **Given the lopsided spending shares cited by the authors, with some clubs in double digits, a club spending 5 percent of the total might always finish in the top half.**

* * *

p. 105 “[A]t best 10 percent of the 699 managers we have observed since 1973 are proven overachievers.”

In reality, the authors' analysis may not prove that any managers at all are overachievers. The 10 percent figure is almost exactly what randomness would predict according to the authors' definition of overachieving, suggesting the possibility that no managers actually add value to their clubs.

The authors used a statistical technique called regression analysis to establish a link between wages and league position. Using this link, they were able to predict the league position for each team in each season using its total wage bill. In regression analysis, the errors in this prediction – that is, the differences

between the predicted and actual league positions – should on average be zero. This generally means that the errors are just as likely to be positive as negative. Also, regression analysis will usually predict fractional league positions, such as 4.6 or 13.2. In cases where the size of the error is less than 0.5, rounding the predicted league position will give the actual league position. So a manager can really only outperform expectations if the error is positive and greater than or equal to 0.5.

This may not seem like a high bar, except that the authors insist that the manager should outperform five years in a row to be considered an overachiever. Considering the ninety-two places in the top four divisions over twenty-seven seasons (as the authors do), and the notion that wages account for about 90 percent of the variation in league position (as the authors state on p. 13), the standard deviation of the errors in prediction should be about 8.4. If these errors are normally distributed, then the chance of overachieving by at least half of a league place as a result of randomness is about 48 percent in a given year. To do so randomly for five years in a row, the chance is 2.5 percent.

Of course, managers who work for more than five years have several chances at the feat. Among the top forty managers listed by the authors on p. 106, the average tenure in the league is about 11 years. If the whole sample of 251 were made up of managers with the same tenure as these forty, randomness alone would yield about 39 so-called overachievers. On p. 105, the authors write that “somewhere between forty and seventy of the managers in our sample made a positive difference; that is, they usually overachieved with their teams.” In other words, almost all of the managers singled out by the authors might have been accounted for by randomness.

If the tenure of managers in the rest of the sample was actually shorter than the tenure of the top forty – after all, those who didn't outperform might not have lasted as long in the league – then more managers may indeed have outperformed than randomness alone would have predicted. But the possibility remains: **All of the authors' so-called overachievers could have just been the beneficiaries of random variation in team performance.**

* * *

p. 118 “Any statistician can predict what should happen after a low point: whether or not the club sacks its manager, or changes its brand of teacakes, its performance will probably regress to the mean. Simply put, from a low point you are always likely to improve.”

This statement would be true if nothing ever changed in soccer. Regression to the mean only works when the dynamics of the system are the same. Everything

that matters about the club must stay the same, as must everything that matters about the other clubs in the league.

The authors' hypothesis is that the manager does not matter to the club, but they neglect the possibility that the manager's sacking might coincide with something else that matters to the club's performance in the league. Moreover, they conflate average performance across the league ("the average club earns 1.3 points a match") with a club's own performance ("a club sacks its manager when it averages only 1 point a match"). Only the club's own performance can be expected to regress to the mean, since clubs are all different; for clubs that sack their managers, that performance may fall short of 1.3 points per game.

Moreover, managers are often sacked after the club loses a series of games that it was expected to win. Because the club's schedule does always regress to the mean – it has to play the good teams as well as the bad ones – then one might expect performance to be as bad or worse after a sacking. The authors do mention the possibility that players work hard to impress the new manager, concluding that "on that logic clubs should sack managers even more often." But sacking managers has costs: clubs usually have to pay off the old manager; it takes effort to find a new one; and a coaching staff that's constantly changing may be disruptive to the club as a whole. **Without an accounting of the costs and benefits of sacking managers, the authors' statement has no meaning.**

Financial management of professional clubs

The idea that football clubs do not exist to make a profit is another theme to which the authors often return. Indeed, they suggest that profitability may be at cross-purposes with results on the field. Yet the analytical basis for this assertion is hardly solid.

* * *

p. 67 “Any team that pursued the highest possible profits would probably end up being relegated, because it wouldn’t be spending enough to hire good players. And if the club got relegated, it would lose much of its revenues. So soccer becomes an arms race: every club overspends for fear of the neighbors.”

This passage contains at least two fallacies. First, it assumes that teams make decisions with only one season in mind. Earlier, the authors quote Billy Beane saying that “Wenger runs his football club like he is going to own the club for one hundred years.” Many other clubs engage in rebuilding projects and attempts for European qualification that last several years. Second, the clubs are not overspending if the alternative would be relegation. Imagine an ice cream factory that starts using cheap chemicals instead of eggs to make its product richer. If consumers can tell the difference, it will lose business to its competitors. So it spends more in order to stay in business, earning the “highest possible profits” over a longer period. **This is not overspending; it’s spending justified by competition and the long-term viability of the club.**

* * *

p. 69 “One reason Bundesliga clubs rarely do well anymore in the Champions League is precisely that they make a profit.”

The gist here is that spending more money (presumably on wages and not transfers, following the authors’ earlier points) would lead to better performance for German teams in the Champions League. This may be true, but to say that “Bundesliga clubs rarely do well anymore in the Champions League” is not quite accurate.

The authors also miss the elephant in the room: Compared with England, Italy, and Spain, Germany has fewer teams in the Champions League to begin with. Going back a decade and more, Germany has had three spots in the Champions League compared to four for England, Italy, and Spain. Typically, two German teams went directly into a group stage with thirty-two teams, and another could get there through qualifying. At most, Germany could have less than a tenth of the teams in the group stage.

With this in mind, German performance has been pretty good. In the twenty seasons through 2010-11 (the time of the authors' writing), a German team won the Champions League twice – a tenth of the time. Five German teams appeared in the final out of a potential forty spots – an eighth overall. Since the knockout round became sixteen teams in 2003-04, an average of 1.75 German clubs made it to that stage of the competition out of the three that entered – about a ninth of the sixteen spots per year. Germany did as well as or better than the average Champions League team would have been expected to do.

On p. 168, the authors give a list of the twenty most popular clubs in Europe. Together, these clubs account for all but one of the Champions League wins in the twenty seasons from 1991-92 through 2010-11, the time of the authors' writing. Only one of the twenty clubs, Bayern Munich, is German. The fact that Bayern won twice seems impressive given the relatively small size of German clubs. (To be sure, winning the Champions League might have made other German clubs more popular, but there's a chicken-and-egg problem here.) And of course, Dortmund won in 2013. **Compared to all other countries, "rarely do well" does not give a statistically fair appraisal of Germany's performance in the Champions League.**

* * *

p. 66 "Spanish teams were twelve places *above* their profit-maximizing position over the sample period, but less than half a place *below* their win-maximizing position."

The authors cite a study purporting to show evidence that Spanish clubs targeted results rather than profits. Yet the budget constraint faced by clubs is missing here. Economists always maximize an outcome subject to a constraint. In this case, it's not clear what the constraint is: clubs are maximizing wins subject to what? If the answer is a budget, then there is an implicit assumption that clubs' budgets are fixed. Yet in the very next paragraph, the authors refer to the deep pockets of "magnates like Florentino Pérez and Jesús Gil y Gil," who were "especially prone to blowing what looked like absurd sums of money on players." For these wealthy men, the club's budget was almost unlimited. Moreover, the authors spend many later pages describing how clubs could run up huge debts and then simply walk away from them – another indication that the budget was only a soft constraint. **Without a hard constraint, the term "win-maximizing" is meaningless.**

* * *

p. 67 "No matter how much money Spanish clubs got their hands on, they spent it. In the decade that Stefan and Pedro studied, the

average revenues of a club in the Spanish first division (Primera División) rose nearly fourteenfold, from €4.3 million in 1994 to €59 million in 2004. Yet the share of revenue that clubs spent on player wages didn't drop much throughout the period: in that decade, first-division clubs paid over an average of 62 percent of their revenues to their players. In other words, the clubs weren't able to save all the additional money or do anything else with it, such as build new stadiums or cut ticket prices."

On the contrary, clubs clearly were able to do other things with the extra money – and there was a lot of it in absolute terms. If the share of revenues paid to players stayed the same, then the amount of money not paid to players apparently grew fourteenfold between 1994 and 2004. Some other costs, like taxes, may have risen proportionally as well, but it seems likely that clubs went from having about €1.6 million to having about €22 million to save or spend on things other than wages. **In fact, seven clubs that played in the Primera División during those ten seasons – Espanyol, Extremadura, Huelva, Las Palmas, Murcia, Numancia, and Oviedo – even built new stadiums.**

* * *

p. 200 "During the second, 'unequal' period, total attendance rose from 8 million to 13 million, even though tickets became much more expensive and people had many more choices of how to spend their free time."

Here the authors only consider reason why attendance at soccer games in England might have fallen during the period from 1989 to 2008, a period they say was characterized by unequal performance among the teams in the top division. They fail to account for the many other reasons why attendance may actually have increased.

The Premier League was founded in 1992 when the teams of the old first division decided to break from the rest of the Football League in order to capture new revenues from broadcasts. This deal eventually resulted in many more matches being shown on television in England and around the world, hugely expanding the audience for football and its fan base.

In addition, the inception of the Premier League coincided with a new wave in the globalization of football personnel. As Ollie Williams of the BBC wrote in 2009, "At the start of the 1989-90 season, leading clubs like Arsenal and Manchester United boasted just one or two regular first-team players who were born outside the United Kingdom. Now, Premier League teams have, on average, 13 foreign-born stars within their ranks." The opening of the Premier League's labor market to foreigners greatly increased the size of the talent pool and, as a logical

consequence, the quality of players in the division. This increase in quality may also have attracted more fans.

But the final factor that ought to have been considered by the authors is the increase in stadium size. The change in Arsenal's stadium alone, from 38,419-seat Highbury to the 60,361-seat Emirates in 2006, accounted for a difference in attendance of more than 400,000 over a season's nineteen home games. Newcastle's expansion of its St. James's Park from about 37,000 seats to more than 52,000 added almost 300,000 more to league attendance. Several other clubs also moved into bigger stadiums or expanded their existing stadiums. **In short, the increasing attendance in England's top division between 1989 and 2008 may have had more to do with stadium size and promotion of the product than with any changes in the product itself.**

* * *

p. 80 "[E]conomic historian Les Hannah made a list of the top one hundred global companies in 1912, and researched what had become of them by 1995. Nearly half the companies – forty-nine – had ceased to exist. Five of these had gone bankrupt, six were nationalized, and thirty-eight were taken over by other firms."

The authors use this passage to contrast with the fact (on p. 72) that 97 percent of the clubs in the English league in 1923 still existed at their time of writing. Yet soccer clubs are not like companies. Those that are nationalized or taken over by other firms do not cease to exist; their labor and capital just become part of a new organization, usually doing the same thing as before. A more accurate statement would be that five out of one hundred had ceased to exist, providing their liquidation wasn't followed by a reconstitution of a new business in much the same way the authors describe for soccer clubs. **At least 95 percent of these companies clearly did continue to exist in some form, so there is not necessarily anything special about England's top soccer clubs.**

City size and club success

The authors trace a trend suggesting that successful soccer clubs come from provincial, industrial towns rather than Europe's biggest cities and capitals (though they suggest this may change). Along the way, they make some factual errors and appear to misinterpret some mathematical results. Their conclusions about the future epicenters of soccer success are therefore open to question.

* * *

p. 162 “Provincial towns like Nottingham, Glasgow, Dortmund, Birmingham, and Rotterdam have all won European Cups, while the seven biggest metropolitan areas in Europe – Istanbul, Paris, Moscow, London, Saint Petersburg, Berlin, and Athens – never have.”

This list of the “seven biggest metropolitan areas in Europe” is a peculiar one, not least because it compares population in the 2000s to European soccer success since 1956. Athens doesn't make the top seven in the list of Larger Urban Zones compiled by the European Union's Eurostat agency, which doesn't include Russia. Madrid, on the other hand, ranks fourth. The authors also seem to have dropped the Ruhr, which the EU considers to be a single metropolis. This table lists the top ten metropolitan areas in Europe according to official statistics and the related teams that have won major European competitions in the past twenty years:

<u>City</u>	<u>Population</u>	<u>Data year</u>	<u>UCL</u>	<u>UEFA/Europa</u>
London	12.3 million	2007-09	Chelsea	Chelsea
Paris	11.5 million	2003-06		
Moscow	11.5 million	2010		CSKA
Istanbul	11.0 million	2003-06		Galatasaray
Madrid	6.3 million	2007-09	Real	Atlético
Ruhr	5.2 million	2010-12	Dortmund	Schalke
Berlin	5.1 million	2010-12		
St. Petersburg	4.8 million	2010		Zenit
Barcelona	4.4 million	2007-09	Barcelona	
Athens	4.1 million	2007-09		

Chelsea's Champions League victory in 2012 probably came too late for the authors, and indeed their next passage is on “why Arsenal and Chelsea haven't won the Champions League.” It should also be noted that Berlin was a divided city from the 1940s until 1990, and this may have hampered the development of a dominant soccer team; in the pre-war years, Hertha BSC Berlin was perhaps the second-best team in Germany.

These are small points, however, compared to the main problem: 31 of 58 winners of the European Cup or Champions League have come from the biggest cities in their respective countries, and 47 of 58 have come from England, Italy, Spain, the Netherlands, and Portugal. Though Russia, France, and Turkey may have big cities, they don't have the best soccer teams; the quality of their leagues today is evidence of this as well. Shanghai, Mumbai, Cairo, and Lagos don't have the best soccer teams, either. They also don't have the fortune to be in Europe and are thus ineligible for the Champions League. If they had been in Europe, would they have been legitimate points of comparison? In soccer terms, this seems to be a case of apples-to-oranges. **The posited relationship between city size and European success seems to be based on faulty or at least selective data.**

* * *

p. 167 “Few would dispute that this top twenty includes most of Europe’s best-supported clubs. And there is something remarkable about this list: the biggest clubs are not in the biggest cities.”

It is easy to check this assertion using the table above. Half of the top twenty most popular clubs are from metropolitan areas that appear in the top ten above: Barcelona, Real Madrid, Chelsea, Zenit St Petersburg, Arsenal, CSKA Moscow, Galatasaray, Spartak Moscow, Fenerbahçe, and Dynamo Moscow. This list also includes four of the five most popular teams. Adding the ten next-biggest metropolitan areas, four more of the biggest clubs would be included: Manchester United, AC Milan, Bayern Munich, and Inter Milan. **The top twenty metropolitan areas in Europe are actually home to fourteen of the twenty biggest soccer clubs.**

* * *

p. 179 “Nonetheless, Kwok Tong Soo from the London School of Economics has shown that if you measure European metropolitan districts, Zipf’s law works well in eight out of nine European countries.”

The authors use Soo’s result to argue that the set of megacities in Europe is likely to stay small and constant; they then assert that these cities – Moscow, Istanbul, Paris, and London – will come to dominate European football. This is a partial misrepresentation of the results in Soo’s paper, “Zipf’s Law for cities: a cross-country investigation” (*Regional Science and Urban Economics*, Volume 35, Issue 3, May 2005, pp. 239–263). It appears to be based on a single panel of Table 2 in the paper, which uses a statistical technique called a Hill estimator. It is worth noting that Soo treats all European countries separately, looking at the

biggest cities in each country rather than the biggest cities across all of Europe; the authors do not do this in their assessment of Champions League winners.

Soo essentially rejects this result because of low statistical significance. He concludes that “the claim that Zipf’s Law holds for urban agglomerations is strongly rejected for our sample of countries.” **The dominance of a few big cities in European soccer may be further off than the authors suggest.**

Miscellany

More questionable statements and interpretations of data are scattered throughout the first half of *Soccernomics*, though only some of them may concern professional clubs. Here are a few of the most pertinent.

* * *

p. 7 “AC Milan’s in-house medical outfit found that just by studying a player’s jump, it could predict with 70 percent accuracy whether he would get injured.”

This statistic sounds impressive, but it’s meaningless without some notion of time. Could the medics predict with 70 percent accuracy whether a player would be injured in a specific game, in a season, or in an entire career? **Predictions become more difficult the longer you look into the future, so the missing time horizon matters here.**

* * *

p. 94 “Quite soon, enough clubs were hiring blacks that black players came to be statistically overrepresented in soccer. Only about 1.6 percent of people in the 1991 British census described themselves as black. Yet in the early 1990s about 10 percent of all players in English professional soccer were black.”

The authors are comparing the population of professional soccer players to a much bigger and very different pool. For starters, soccer players only come from a small age band of the population whose racial makeup may not reflect the population as a whole. Among men in England aged 18-34, 2.5 percent identified themselves as black in the 1991 census.

More importantly, many people in the population don’t play soccer; soccer is likely to be much more common, for example, in urban areas where there are enough players to form a league. In London, the share of men aged 18-34 who identified themselves as black was 9.3 percent, almost the same as the share of professional soccer players cited by the authors.

The urban British pool is too small, however. It only makes sense to compare the population of players to the pool from which they were drawn: young men who played soccer in all the countries that contributed players to the English league. Black players began arriving from Africa and the Caribbean in the 1990s, too. **The authors’ is an apples-and-oranges comparison.**

* * *

p. 126 “Match results appear to be the same with or without penalties. Penalties do not matter.”

The authors make this categorical statement after a comparison of games where a penalty was awarded with games where no penalty was awarded. However, in their analysis they fail to make a distinction between games where one side is awarded more penalties than the other and games where both sides are awarded the same number of penalties. They place all of these games into the same group with total penalties greater than zero.

This may be an important mistake. Penalty awards that cancel each other out – one or more for each team – are a rare but known phenomenon. Some observers have even accused referees of awarding a soft penalty to one side after making a poor decision in awarding a penalty to the other. When both sides receive the same number of penalties, the scope for bias or for penalties to affect the result of a game may therefore be narrower. **Without separate consideration of the subset of matches where one team is awarded more penalties than the other, the authors’ statement cannot be proven.**