EXPLAINING INTERNATIONAL FOOTBALLER SELECTION THROUGH POISSON MODELLING

David Butler, Robert Butler, Justin Doran, and Sean O'Connor
Department of Economics, Cork University Business School, University College Cork

Paper published as: Butler, D., R. Butler, J. Doran, and S. O'Connor (2018) Explaining international footballer selection through Poisson modelling, Journal of Economic Studies, Vol. 45, Iss. 2
http://www.emeraldinsight.com/journal/jes

ABSTRACT

Purpose: Growing evidence suggests regional economic factors impact on individual outcomes, such as life expectancy and well-being. This article investigates the impact that player-specific and regional differences have on the number of senior international appearances football players accumulate over their careers, for six UEFA member countries, from 1993 to 2014.

Design/methodology/approach: The research employs a Poisson regression model to analyse the impact of individual and regional factors on the number of senior international caps a footballer receives over the course of their career.

Originality/Value: This is the first study to analyse a Pan-European dataset, using an increasingly adopted econometric method to understanding regional economic development - Poisson modelling.

Findings: The results indicate that both individual and regional variables can explain the number of caps a player receives over the course of their career. We find that an individual's career length positively influences the number of international caps accrued. Players born in wealthier and more populous regions accumulate a greater number of international appearances. Distance from the capital has no effect, however, the number of youth academies in the player's region of birth has a significant positive effect.

Practical Contribution: The results can provide insights for local football authorities and policy makers concerned with regional characteristics and those interested in the development of elite talent.

Keywords: Association football, regional development.
JEL: R1, Z2

1. INTRODUCTION

An extensive literature now exists which demonstrates the benefits of living in a country or region with higher levels of income per person. These benefits can range from longer life expectancy (Carlson, 2005; Chetty *et al*., 2016) to increased life satisfaction (Okulicz-Kozaryn, 2012). This line of research has now extended to sporting achievement, with the relationship between regional development and sporting success receiving increased attention over the past number of years (Mourão, 2010). In light of these recent advances, we develop the connection between spatial factors and sporting performance further by considering whether regional variables influence a footballer's potential to play international football for their country of birth. In doing so, we contribute not only to a nascent literature concerning regional determinants of performance in football, but also offer economic insights to those concerned with environmental factors associated with talent development.

The primary motivation for this study derives from two contrasting intuitions. This first suggests that individuals born in wealthier European regions are more likely to accumulate higher numbers of international caps as they develop their skills in locations that have greater investment in footballing infrastructure. Examples of this include improved playing surfaces, floodlit pitches, indoor training facilities, and investment in human capital which increases coaching expertise. A second contrasting intuition exists. One could argue that economically depressed regions have historically produced superior players. The argument here is premised on lower incomes bringing about greater practice levels. The costs of practice for football are low in comparison to other sports and leisure activities. Informal or 'street' environments, epitomised by the phrase *jumpers for goalposts,* allow children to hone their skills without fulfilling the formal requirements associated with the fully-fledged version of the sport. For example, Kuper and Szymanski (2009) explain how football success was predominantly associated with working class regions. These regions possessed lower incomes per capita. Capital cities tended to have a variety of amenities in comparison to provincial cities or regions, where sport, particularly football, played a more central role in society. Many examples come to mind including cities such as Glasgow, Hamburg, Liverpool, Nottingham and Porto, all of which have produced teams to win the European cup, whereas capitals such as Berlin, Lisbon, Paris or Rome never have. Although these winning cities are major urban locations, and are not always significantly depressed relative to primary municipalities, capital cities may yet hold specific features which cause individuals to substitute away from football.

To question if development, and thereafter elite performance, is a function of one's region of birth, we collect data on individual footballers for six European countries (England, France, Germany, Italy, Netherlands and Spain) from 1993 to 2014. Our measure of elite footballer performance is the number of senior international caps received by the player over their career and is confined to those footballers selected to represent their country at a major international tournament. In sum, we assess regional factors underpinning 12,330 senior international appearances. Utilising information on both individual specific and regional characteristics, we assess the extent to which these explain an individual's performance. Our economic data is sourced from Eurostat at the NUTS2 level[1]. This allows us to control for

regional impacts on player's international performance. In addition, we control for the spatial distribution of elite academies across the six countries, allowing us to take account of the location and number of talent development centres across Europe. The use of regional and individual characteristics in our regression model is an innovative development in this field of research and allows us to investigate this relationship at the individual and regional level. In doing so, we add to a general literature that explores the economic and demographic determinants of international football success across national boundaries (Hoffmann., *et al* 2002; Hoffmann., *et al* 2006; Gelade and Dobson, 2007). Specifically, we add to an underdeveloped literature that considers regional variations associated with success in football (Dobson and Goddard, 1996; Mourão, 2010).

As an applied study, this research has several advantages. Firstly, we adopt a sophisticated econometric technique by employing Poisson regression models to analyse the number of senior caps players receive. Secondly, while past studies have focused on regional effects at a club level, we analyse a pan-European dataset. This approach allows us to merge both *regional* and *individual* characteristics for sub-national districts across Europe. As is understood, an analysis at this level is yet to be conducted. Thirdly, as we focus on regional development and performance, our results can provide insights for local football authorities and policy makers concerned with regional characteristics and development of elite talent. As we outline in the next section, these findings can also inform those concerned with the talent development process.

This paper is proceeds as follows. Section 2 conceptualises the issue of player development by referring to the talent formation literature from sport science. This provides contextual detail on the player development process and allows us to highlight where economic insights can add value. Section 3 briefly considers the related empirical literature specific to the intersection of regional characteristics and sport. As this literature is relatively underdeveloped we consider regional level contributions across all sports. This is intended to provide an insight to the few works that have considered sport from a regional perspective. Section 4 describes the data accessed for this study, provides descriptive statistics and explains the Poisson regression method. The results are presented in Section 5 along with a discussion of their relevance. Section 6 concludes the paper.

2. CONCEPTUAL FRAMEWORK

Achieving the status of an elite footballer through a talent developmental process can be thought of as a product of three interactions; *intrapersonal characteristics*, *environmental factors* and *chance* (Vaeyens *et al*, 2008). Our analysis in this study can be conceptualised at the nexus of the latter two components in the development process, assuming birth location is stochastic. In light of our intuitions regarding talent development presented in the Introduction, we consider whether regional economic factors influence a footballer's potential to play international football.

In contrast to the *environmental-chance* interaction, a wealth of literature has considered each of the three domains related to player development. For instance, Reilly *et al* (2000)

document intrapersonal characteristics associated with player development, addressing the physical and cognitive demands of becoming an elite player. More recently, le Gall *et al* (2013) have shown superior anthropometric and fitness characteristics within youth players (e.g. height, body mass and sprinting ability) play a key role in the likelihood of a youth player reaching elite level. A developed strand of literature also exists for the last criterion - *chance* events important to talent identification and development. For example, an abundance of studies have shown how the (mostly) fortuitous event of birth-date can lead to relative age effects, with elite underage performers more likely to be born at the start of a selection period (Musch and Grondin, 2001; Helsen, Van Winckel and Williams, 2005; Butler and Butler, 2015).

Unravelling the *environmental factors* fundamental to talent development in football is a complex task. By definition the environment in which footballers develop will differ and will likely be multifaceted. To unpick this complexity, explicit environmental factors are often studied in isolation. These analyses have usually concentrated on football academies, the primary developmental environment for young footballers. Many of these investigation have questioned the role of environmental factors operating *within* academies, evaluating dynamics such as the importance of organizational culture or practice and learning environments. Typical variables which are deemed important to the developmental process include player lifestyle, relationships with academy staff and school performance. Unsurprisingly, recent qualitative evidence suggests that appropriate surroundings are essential for reaching professional status (Larsen *et al*, 2013; Mills *et al*, 2014).

For the purposes of this work, we focus on economic characteristics *external* to the academy environment. A detailed analysis of regionally specific economic factors associated with a player's success has been scarcely considered in the context of European football. Internal (or club-specific) environmental conditions appear to matter for optimal player development. Do local economic conditions influence the outcome too? While past studies have highlighted the significance of spatial factors relating to elite performance, and identified the merit of this approach as an avenue for investigation, they have generally focussed on population level characteristics, rather than employing a deeper examination of economic variables (Côté *et al*, 2006; MacDonald *et al*, 2009; Baker *et al*, 2009). The absence of investigation here presents an opportunity to add economic insights and inform those interested in the player development process with findings on any sporting externalities associated with regional economic development. As intuition suggests, spatial factors could act as a barrier or support to reaching an elite level.

3. RELATED EMPIRICAL LITERATURE – SPORT & REGIONS

A growing number of researchers are highlighting the importance of economic factors, at both regional and country level, in determining sporting success. Nearly two decades ago, Downward and Dawson (2000) demonstrated the impact of unemployment rates on professional team performance across sports, and the consequences this had for clubs excelling at a national level. This was followed by Bernard and Busse (2004) who examined the determinants of Olympic success at a country level and report that while population size was an important explanatory variable in determining the number of medals won, "Real GDP

[was] the best single predictor of a country's Olympic performance (Bernard and Busse, 2004, p. 413). This finding is supported by Rathke and Woitek (2007), with the authors reporting the existence of a positive and statistically significant relationship between national income, population levels and sporting achievement. Some counterbalancing evidence does exist. Mourão (2014) examines the regional determinants of competitiveness in volleyball, basketball and handball and finds that variables related to the sports environment can explain competition performance. A centralised location is not always critical to reach an elite level of performance. The results presented "support the possibility that non-central locations can rediscover their potential as athletic centres" (Mourão, 2014, p. 1484) and suggests that higher levels of local income are not always necessary for producing elite sporting talent.

As is understood, the analysis of explicit regional (as opposed to transnational) level economic factors effecting football success remains largely unexplored. Simple analyses of the regional distribution of where European footballers are born has been left to the media (Smith, 2016). Past academic research in the area has focused on how regional level economic activity correlates with both the demand for domestic football, and competitive balance within leagues. One of the earliest examinations in the area was conducted by Dobson and Goddard (1996). While the investigation did not examine team performance, it did apply regional measures to analyse the demand for football in both English and Welsh regions in the short run and long run. The authors' reported that a team's performance, ticket prices and the number of goals scored in matches all influenced demand in the short-run. Long-run demand was influenced by team performance and the unemployment rate in a particular region. England is not the only country to be examined. Leeds and Leeds (2009) considered international performance and find that hosting the FIFA World Cup, population size and income level all improve performance. The regional factors effecting club performance, have been examined across Europe. Using data from 1990 to 2006 Mourão (2010) finds strong evidence between regional economic development and elite club performance, showing that densely populated regions with higher GDP per capita are more likely to host competitive professional teams.

4. DATA AND METHODS

3.1 Data and Descriptive Statistics

Data is collected for competitive matches in elite international competitions only. The sample consists of all 274 players selected in the national team squads of England, France, Germany, Italy, Netherlands and Spain from the FIFA World Cup Finals in 2006 to the FIFA World Cup Finals in 2014, inclusive. This covers five competitions; three FIFA World Cup Finals in 2006, 2010 and 2014, and two UEFA European Champions Finals in 2008 and 2012.

This timeframe is dictated by access to economic data. The earliest year any member of these squads received a cap was 1993 (Santiago Cañizares – Spain). Hence, we analyse players who received caps between 1993 to 2014. Accordingly, the distribution of players by nationality is 49 (England), 47 (France), 46 (Germany), 51 (Italy), 40 (Netherlands) and 41(Spain). These players are chosen as they represent the most consistent performers at international level amongst European nations, in terms of qualifying for elite international

tournaments and progressing to the latter stages of major competitions. These nations have a strong football culture and the sport is valued nationally, ensuring that accurate and detailed biographic information is available on the early years of these footballers lives.
While 274 observations may appear a relatively small sample given the frequency of international matches, this sampling strategy is chosen to ensure that only leading footballers are analysed. Ultimately, this procedure is premised on the revealed preferences of managers, who are assumed to include elite footballers in tournament squads. Furthermore, many squads between international competitions do not experience comprehensive turnover in personnel. For instance, it is not uncommon that half of an international squad would straddle two World Cup competitions, maintaining their position over a four-year period. The player data is retrieved from multiple online sources for triangulation purposes. The primary database accessed for performance statistics is available online at World Football (2016). This website provides comprehensive career statistics on an individual player basis. Data is gathered on a player's international career length (in years), whether they performed for their country at a U-21 level and their number of senior international caps.

It is assumed that a player's place of birth is a proximate measure for the region where their skills are developed. We believe that this is a reasonable assumption as a player's earliest development generally occurs in the same location of their birth. Knowledge of player ability is also concentrated in local regions. We assume that that players are most likely to be recruited locally. Two measures are adopted to ensure the validity of this assumption. Firstly, players born in alternative countries to their international affiliations are removed from the sample[2]. Secondly, we assess biographic information for each player in an effort to verify their background and ensure that their birthplace reflects the region of their development. Biographic information existed given that the sample includes elite internationals. A key indicator that allows us to question whether a player developed outside of their region of birth involves identifying the location of each player's primary schooling.

Figures 1 to 6 provide a visualisation of birthplace distribution across the six countries[3].

*****Figure 1 about here***** *****Figure 2 about here*****

*****Figure 3 about here***** *****Figure 4 about here*****

*****Figure 5 about here***** *****Figure 6 about here*****

As player development commonly occurs through the academy structure of professional clubs, we collect data on the spatial distribution of 118 elite academies across the six countries. These academies are frequently attached to elite professional clubs that would have existed historically during the development period of the players under analysis. Regional development centres funded and structured by official national associations are also included. This data is accessed via online resources and is verified using official club addresses accessible via official club websites and Google maps. In cases where a specific academy address could not be located, the location of the home stadium of the club is used as a

proximate measure of the academy location. Regularly, academies are located in close proximity to a club's stadium.

At the NUTS2 level, we collate a dataset on wealth and population levels for European regions to match specific players. Additionally, we measure each individual's place of birth relative to the capital city using great circle distance calculations from the centroid of the individual's birth region to the centre of the capital city of the country[4]. The distance from a player's birthplace to the capital is measured in kilometres. For the purpose of the empirical estimations, we utilise the inverse of distance from the capital city. As some individuals are born in the capital city region, these individuals receive a value of 1, indicating no distance from the capital city region. Table 1 provides a list of the variables under analysis, including descriptive statistics.

*****Table 1 about here*****

The average number of caps for the sample is just under 45, with a standard deviation of 33.53. Four players have a single cap. The distribution of caps is presented in Figure 7. As selection to the senior international squad is usually not instantaneous, we utilise a binary variable to measure if a player has represented their country at U-21 level. In total, nearly 80% of the sample received a youth cap. At the time of writing, 51% of the sample are now retired from international football, recording an average career length of 7.44 years. The career span of retired players varied, ranging from 1 year to 15 years. A mean distance from the capital of approximately 339km is reported with a standard deviation of 287km. Eighteen players are born in their respective capital city.

*****Figure 7 about here*****

Regional GDP per capita at Purchasing Power Standard (PPS) and population act as measurements for regional wealth and size[5]. Mean values of circa €28,000 and 3,845,000, along with substantial standard deviations for both variables indicate our sample consists of varied regions in terms of size and economic prosperity. *Campania* in Italy, a predominantly agricultural economy, has the smallest GDP per capita, while *Inner London*, has the largest value. France hosts both the largest and smallest regions in our sample (in terms of population), represented by Corse and Ile de France (region of Paris).

3.2 Model Specification

In our methodology the aim is to understand the factors which determine the number of caps a player receives. The purpose of this work is not to explain club or country performance but rather to examine *individual* performance across regions, evaluating external environmental conditions important to player development. In order to conduct our analysis, we employ a multivariate Poisson model. As the dependent variable in our case is a count of the number

of caps a player receives, this variable will not follow a standard normal distribution. This means that estimation via ordinary least squares (OLS) would produce biased results. Therefore, the estimation strategy adopted is a Poisson regression (Fingleton, 2006). The model to be estimated is presented in equation (1).

$$Y = X\beta + \mu \quad (1)$$

Where Y is the dependent variable (the number of caps a player receives). X is a matrix of independent variables (the full list of which is outlined above in Table 1) which is N*k, where N is the number of observations in the data and k is the number of variables entered in the regression. β is a k*1 vector of the coefficients to be estimated. We note that a constant is included in X and β. μ is a N*1 vector containing the error terms.

Included in our independent variable matrix are the length of career of the player, the number of youth caps the player received, the wealth and population of the region the player was born in, the distance from the region of birth to the capital city of the country, and the number of youth academies in a player's region of birth.

When estimating our model, a further issue arises due to the repeated sampling within clusters. Moulton (1990) notes that the inclusion of micro level data from repeatedly sampled areas raises potential implications for the standard errors of our model. Even the slightest level of correlation within groups in the error term can cause serious downward bias in the estimated standard errors which has implications for hypothesis testing. This is also noted in recent work by Canton (2009), Baum and Mitchell (2010) and Doran and Fingleton (2013; 2015; 2016). These point out that it is likely that observations will be correlated within regions as region specific elements may be impacting on all the people within that region. Therefore, since we expect to find intra-regional correlation within our regression model, we use the standard modification for intra-group dependence which produces larger than otherwise standard errors (and adjusts the variance-covariance matrix), avoiding upwardly biased t-ratios. This is important as Moulton (1990) points out that failing to make this adjustment can result in incorrect inferences regarding the significance of our variables and we would be more likely to incorrectly reject a null hypothesis of insignificance.

5. RESULTS

Table 2 presents the results of our analysis. The estimation utilises individual level data and all regional variables.

The length of a player's career has a significant positive effect on the number of caps a player receives. While this result is expected, it is necessary to include this measure in our estimations. Only elite footballers (a small proportion of the general footballer population) reach international player status. Even fewer are selected for the national squads at major tournaments. Many players have long careers without receiving a cap and some exceptional young players are constrained by the number of caps they receive given the brevity of their

career. For others, their opportunity is constrained by injury. Consequently, this measure is an important determinant of success. We also find no statistically significant relationship between a player represented their country at a U-21 level and their subsequent number of senior caps.

We note that three the four regional variables included in this model are statistically significant.
The inverse of distance from the capital city has no significant effect on the number of caps a player receives. While this coefficient is negative, implying that as the distance between the region in which a player is born and the capital city of that country increases, the number of caps received by the player increases, a statistically significant result is not reported.

The population density of a region is found to have a positive effect on the number of caps a player receives. This suggests that players born in more densely populated regions receive more caps. It may be that more populous regions have better football infrastructure, or at a general level, which may result in better player identification or development. This is a common finding in existing literature (see, Leeds and Leeds (2009) as an example).

Players born in wealthier regions accumulate more caps over the course of their careers as do players born in regions with more youth academies. It would appear that regional wealth, rather than disadvantage, is driving talent selection at elite international level across the six countries considered. While the *jumpers for goalpost* intuition may be romantic, like many aspects of modern day football, investment is a key determinant of success.

In order to compare the relative importance of each variable in our model we follow the approach adopted by Fingleton (2006). This involves re-estimating the model but dropping each variable one at a time from the full model to identify the impact of that variable on the model's R^2. The results of this are displayed in Appendix 1. We note that the log of career length appears to be the most important variable at explaining the variation in the number of caps a player will receive. However, this does not mean that the other variables are unimportant. Indeed, the results in Table 2 point to many variables having a statistically significant impact on the number of caps received.

*****Table 2 about here*****

What are the implications of these results for policy makers? In light of the intuitions introduced at the start of this paper, and the process of talent development, we believe that policy makers should take heed of the finding that regional wealth is an important external economic factors when seeking to develop elite talent. Policy makers encouraging local investment in football academies should find solace in these results. Investment in local academies and providing increasingly more opportunity to access greater numbers of these facilitates can improve the likelihood of producing the most elite talent.

6. CONCLUSION

Collecting data on international footballers over twenty-one years and using a Poisson approach to capture regional variation across regions and players, we find that both individual and regional variables can explain the number of caps a player receives over the course of their career. Concerning individual characteristics, the duration of a player's career positively affects the number of caps they accumulate over the course of their career. A player's number of caps at a U-21 level does not influence the number they receive at an elite level. Analysing regional characteristics, we find that the wealth and population of a region has significant explanatory power. There is no statistically significant effect between the numbers of caps received by players born further from capital cities.

These insights build on the work of Mourão (2010; 2014) and adds to our understanding of regional dynamics in football under a new area of investigation – international selection. Our findings open up various avenues for future research. Firstly, due to data limitations, our analysis is limited to regional variations within economically developed states. It would be interesting to test whether the correlation between relative regional development and international success exists in less developed countries. Naturally, comparison across continents would likely allow for a wider distribution of regional economic development. Secondly, we only address international football in this study and cannot comment on the generality of our findings across genders or sports. Applying the ideas suggested in this study to other sports can provide a nuanced view of the power of local economic conditions on talent development.

REFERENCES

Baker, J., Schorer, J., Cobley, S., Schimmer, G., & Wattie, N. (2009). Circumstantial development and athletic excellence: The role of date of birth and birthplace. *European Journal of Sport Science*, Vol. 9, No.6, pp. 329-339.

Baum, S., & Mitchell, W. F. (2010). People, space and place: A multidimensional analysis of unemployment in metropolitan labour markets. *Geographical Research*, Vol. 48, No. 1, pp. 13-23.

Bernard, A and Busse, M. (2004), "Who wins at the Olympic Games: Economic resources and medal totals", *The Review of Economics and Statistics*, Vol. 86, No. 1, pp. 413-417.

Butler, D and Butler, R. "The Relative Age Effect and Under-21 Irish Association Football: A Natural Experiment and Policy Recommendations." *The Economic and Social Review*, Vol. 46, No. 4, pp. 511-519.

Canton, (2009) Human capital externalities and proximity: evidence from repeated cross-sectional data, *De Economist*, Vol. 157, pp. 79–105.

Carlson, P. (2005), "Relatively poor, absolutely ill? A study of regional income inequality in Russia and its possible health consequences", *Journal of Epidemiology and Community Health*, Vol. 59, No. 5, pp. 389-394.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016), "The Association Between Income and Life Expectancy in the United States, 2001-2014", *The Journal of the American Medical Association*, Vol. 315, No. 16, pp. 1750-1766.

Côté, J., Macdonald, D. J., Baker, J., & Abernethy, B. (2006). When "where" is more important than "when": Birthplace and birthdate effects on the achievement of sporting expertise. *Journal of Sports Sciences*, Vol 24, No.10, pp. 1065-1073.

Downward, P., and Dawson, A. (2000), *The economics of professional team sports*, Routledge, New York, NY.

Dobson, S. M., and Goddard, J. A. (1996), "The demand for football in the regions of England and Wales", *Regional Studies*, Vol. 30, No. 5, pp. 443-453.

Doran, J., & Fingleton, B. (2013). US metropolitan area resilience: insights from dynamic spatial panel estimation. In *RSAI (British and Irish Section) Congress, Univ. of Cambridge, August*.

Doran, J., & Fingleton, B. (2015). Resilience from the micro perspective. *Cambridge Journal of Regions, Economy and Society*, Vol. 8, No.2, pp. 205-223.

Doran, J., & Fingleton, B. (2016). Employment resilience in Europe and the 2008 economic crisis: insights from micro-level data. *Regional Studies*, Vol 50, No. 4, pp. 644-656.

EUROSTAT. (2016a), "Gross domestic product indicators – ESA2010", available at http://ec.europa.eu/eurostat/web/regions/data/database (accessed 15th September 2015).

EUROSTAT. (2016b), "Population and Area", available at http://ec.europa.eu/eurostat/web/regions/data/database (accessed 15th September 2015).

Fingleton, B. (2006). A cross-sectional analysis of residential property prices: the effects of income, commuting, schooling, the housing stock and spatial interaction in the English regions. *Papers in Regional Science*, Vol 85, No. 3, pp 339-361.

Gelade, G. A. and Dobson, P. (2007), "Predicting the comparative strengths of national football teams", *Social Science Quarterly*, Vol. 88, No. 1, pp. 244-258.

Gerrard, B. and Dobson, S. (2000), "Testing for monopoly rents in the market for playing talent-Evidence from English professional football", *Journal of Economic Studies*, Vol. 27, No. 3, pp.142-164.

Helsen,W., Van Winckel,J. & Williams,M. (2005). The relative age effect in youth soccer across Europe. *Journal of Sports Sciences*, Vol 23, No.6, pp. 629-636.

Hoffmann, R., Ging, L. C., and Ramasamy, B. (2002), "The socio-economic determinants of international soccer performance", *Journal of Applied Economics*, Vol. 5, No.2, pp. 253-272.

Hoffmann, R., Chew Ging, L., Matheson, V., and Ramasamy, B. (2006), "International women's football and gender inequality", *Applied Economics Letters*, Vol. 13, No. 15, pp. 999-1001.

Kuper, S. and Szymanski, S. (2009), *Soccernomics: Why England Lose, Why Germany and Brazil Win and Why the U.S., Japan, Australia, Turkey – and Even Iraq – are Destined to Become the Kings of the World's Most Popular Game*. Nation Books, New York, NY.

Larsen, C. H., Alfermann, D., Henriksen, K., & Christensen, M. K. (2013). Successful talent development in soccer: The characteristics of the environment. *Sport, Exercise, and Performance Psychology*, Vol 2, No.3, pp.190-207

Le Gall, F., Carling, C., Williams, M., & Reilly, T. (2010). Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy. *Journal of Science and Medicine in Sport*, Vol 1,3 No.1, pp. 90-95.

Leeds, E. and Leeds, M. (2009), "International soccer success and national institutions", *Journal of Sports Economics*, Vol. 10, No. 4, pp. 369-390.

MacDonald, D. J., Cheung, M., Côté, J., & Abernethy, B. (2009). Place but not date of birth influences the development and emergence of athletic talent in American football. *Journal of Applied Sport Psychology*, Vol 21, No.1, pp.80-90.

Mills, A., Butt, J., Maynard, I., & Harwood, C. (2014). Toward an understanding of optimal development environments within elite English soccer academies. *The Sport Psychologist*, Vol 28, No. 2, pp. 137-150.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The review of Economics and Statistics*, Vol 72, No.2, pp. 334-338.

Mourão, P. J. R. (2010), "Regional determinants of competitiveness: the case of European soccer teams", *International Journal of Sport Finance*, Vol. 5, No. 3, pp 222.

Mourão, P. R. (2014), "Does European regional competitiveness influence sports? An analysis of three sports", *Applied Economics*, Vo. 46, No. 13, pp. 1476-1489.

Musch, J., & Grondin, S. (2001). Unequal competition as an impediment to personal development: A review of the relative age effect in sport. *Developmental Review*, Vol 21, No.2, pp. 147-167.

Okulicz-Kozaryn, A. (2012), "Income and Well-being Across European Provinces", *Social Indictors Research*, Vol. 106, No. 2, pp. 371-392.

Rathke, A. and Woitek, U. (2007), "Economics and Olympics. An efficiency analysis", working paper No. 313, Institute for Empirical Research in Economics, University of Zurich, Zurich, January.

Reilly, T., Bangsbo, J., & Franks, A. (2000). Anthropometric and physiological predispositions for elite soccer. *Journal of Sports Sciences*, Vol 18, No.9, pp. 669-683.

Smith, A. (2016), "Where are England international players born?", Sky Sports, available at http://www.skysports.com/football/news/18804/10649832/where-are-england-international-players-born (accessed 4th April 2017)

Sulloway, F. J., & Zweigenhaft, R. L. (2010). Birth order and risk taking in athletics: A meta-analysis and study of major league baseball. *Personality and Social Psychology Review*, Vol 14, No. 4, pp. 402-416.

Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2008). Talent identification and development programmes in sport. *Sports Medicine*, Vol 38, No. 9, pp. 703-714.

World Football (2016a), "England National Team: England » Players and coaches from A-Z", available at http://www.worldfootball.net/teams/england-team/10/ (accessed 14th September 2015).

World Football (2016b), "France National Team: England » Players and coaches from A-Z", available at http://www.worldfootball.net/teams/frankreich-team/10/ (accessed 14th September 2015).

World Football (2016c), "Germany National Team: England » Fixtures and Results", available at http://www.worldfootball.net/teams/deutschland-team/2016/3/ (accessed 14th September 2015).

World Football (2016d), "Italy National Team: England » Players and coaches from A-Z", available at http://www.worldfootball.net/teams/italien-team/10/ (accessed 14th September 2015).

World Football (2016e), "Netherlands National Team: England » Players and coaches from A-Z", available at http://www.worldfootball.net/teams/niederlande-team/10/ (accessed 14th September 2015).

World Football (2016f), "Spain National Team: England » Players and coaches from A-Z", available at http://www.worldfootball.net/teams/spanien-team/10/ (accessed 14th September 2015).

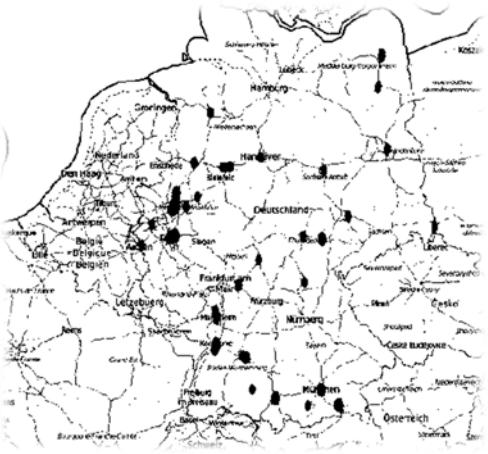
Table 1: Variables & Descriptive Statistics

Variable	Description	Obs.	Mean	SD	Min	Max
Dependent Variable						
No. of Senior Caps	Inclusion for competitive matches in elite international competitions	274	44.7	33.5	1	160
Player Characteristics						
No. of Youth Caps	An indicator of developmental progress - equals 1 if a player attained underage international appearances, 0 otherwise	274	0.7	0.4	0	1
Career Length (yrs)	An indicator of the opportunity available to a player (injury control)	274	7.4	3.7	0	15
Regional Characteristics						
Academy Structure	A control for the spatial distribution of professional development centres	118	-	-	-	-
Distance to Capital (km)	A measure to consider the influence of capital city births	274	339	287	0	2,028
Regional Wealth	A measure of regionally-specific economic development	274	€ 28,239	€ 10,757	€ 15,688	€ 79,988
Regional Population	A measure of regionally-specific population	274	3,845,271	2,454,716	301,719	11,600,000

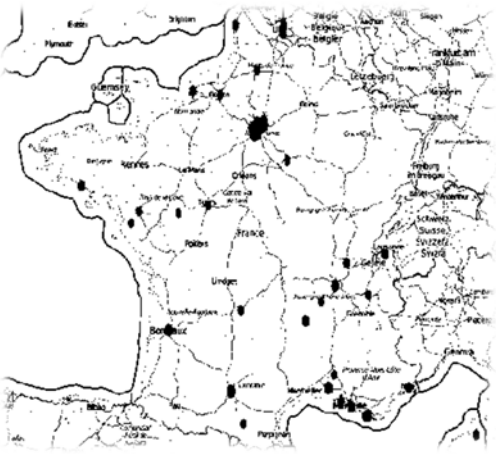
Sources: World Football (2016); Eurostat (2016); Google Maps; Official Club/Association Websites

Figure 1-6: Birthplace Distribution by Country

Germany (n=46)

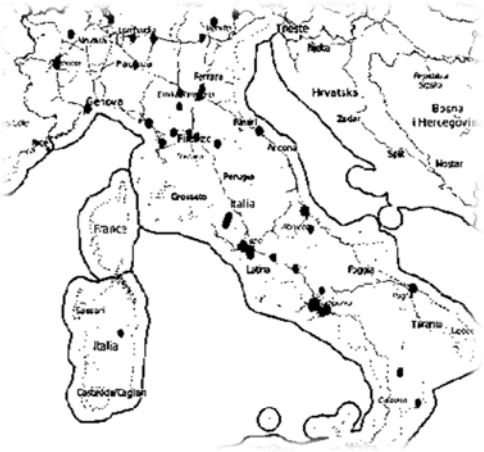


France (n=47)

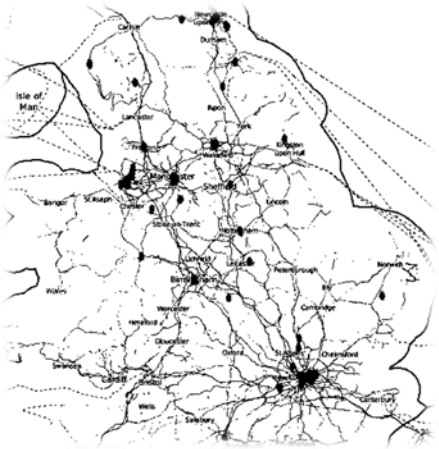


Spain (n=41)

Italy (n =51)



England (n =49)



Netherlands (n=40)

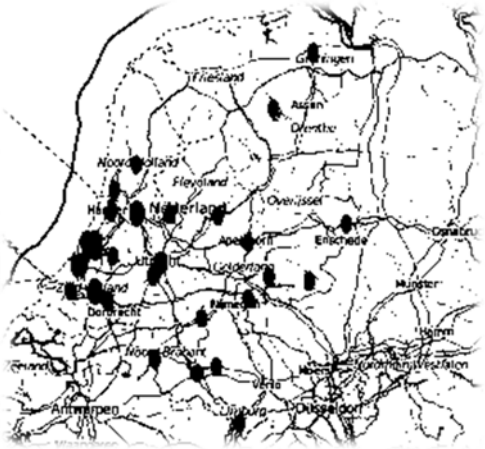


Figure 7: Distribution of Number of Caps

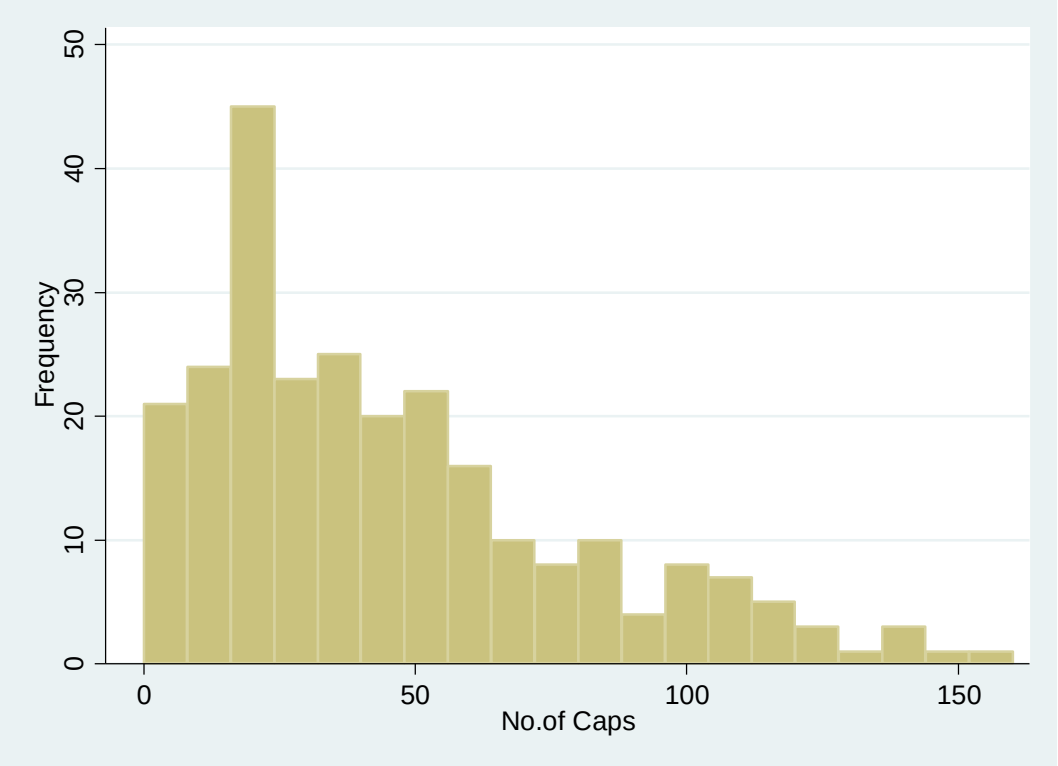


Table 2: Results

Variable	Equation (1)
Constant	-0.4282
	(1.0269)
Log of Career Length	1.0556***
	(0.0554)
No of Youth Caps	0.0954
	(0.0754)
Log of Income	0.1343**
	(0.0668)
Log of Population	0.0557**
	(0.0253)
Inverse Distance (1/distance)	-0.0179
	(0.0173)
No of Youth Academies	0.0415***
	(0.0145)
No of Obs	274
Pseudo R2	0.4898
Log Likelihood	-2101.84

Note 1: ***, **, and * indicate significance at the 99, 95 and 90 percent level.

Appendix 1: Comparison of R^2

Variable	**Log-Likelihood**	**Pseudo R^2**
Full Model	-1663.6753	0.4963
Log of Career Length	-3027.7449	0.0833
Youth Caps	-1669.486	0.4946
Log of Income	-1678.7681	0.4918
Log of Population	-1665.4043	0.4958
Inverse Distance	-1813.385	0.4745
No. Youth Academies	-1680.5826	0.4912

END NOTES

[1] NUTS stands for Nomenclature of Territorial Units for Statistics.

[2] France and Germany both have eight players born outside their county. Two players are born outside of Spain, Italy and the Netherlands. England has one player born outside of the country.

[3] These maps are generated using open heat map software the code for which is all licensed under the Gnu Public License version 3 and map data which is available under a creative commons licence.

[4] For the purposes of this measure, Amsterdam is considered the capital of the Netherlands.

[5] Given limitations in regional data, we derive an average value of GDP per capita for NUTS2 regions and population. This is 2004-2011 for GDP per capita, and 2004-2012 for population.