Birthplace diversity and team performance

Enzo Brox[1]   Tommy Krieger[2]

July 9, 2022

Abstract

Using hand-collected and web-scraped data on 7,208 matches and 3,266 players from the highest division of German male football, this paper examines how birthplace diversity affects team performance. The results of two instrumental variable analyses suggest that birthplace diversity has a hump-shaped effect on team performance. To explain our result, we argue that nationally diverse teams have a wider range of skills and face greater communication barriers.

---

[1]University of St. Gallen, Chair of Econometrics, Varnbüelstraße 14, 9000 St. Gallen, Switzerland. Email: enzo.brox@unisg.ch

[2]ZEW – Leibniz-Center for European Economic Research, Department of Corporate Taxation and Public Finance, L7 1, 68161 Mannheim, Germany. Email: tommy.krieger@zew.de

1 Introduction

The importance of team-based production increased notably over the last three decades (see e.g. Deming, 2017, Jones, 2009, Wuchty et al., 2007). Understanding which factors determine successful collaboration is thus at the very heart of various studies in labor, business, and organizational economics. In particular, a large body of literature studies whether working groups whose members differ in their sociodemographic characteristics are more or less productive than homogeneous working groups.

Making an unambiguous prediction about how diversity in sociodemographic worker characteristics affects the performance of a team is virtually impossible (see e.g. Prat, 2002, van Knippenberg et al., 2004, Williams and O'Reilly, 1998). The reason for this ambiguity is that diversity can have both positive and negative effects. For instance, on the one hand, diversity may broaden the range of skills, which in turn may raise team performance (Cox et al., 1991). However, on the other hand, diversity may lower team performance because of lower group cohesion or higher risk of conflict (O'Reilly et al., 1989). In addition, since various characteristics exist in which members of a team can differ from each other (e.g. gender, tenure, religion), the relevance of the positive and negative aspects of diversity may depend on the characteristic. It is also possible that surrounding factors moderate the effects of diversity (van Knippenberg et al., 2004).

This paper complements the existing literature by providing causal evidence on the relationship between birthplace diversity and team performance. We choose this type of diversity for three main reasons. First, the number of multinational working groups increased considerably in the last decades (see e.g. Freeman and Huang, 2015, Kahane et al., 2013). Second, compared to other sociodemographic worker characteristics, only little research exists on the effects of birthplace diversity on team performance. Finally, managers are interested in this type of diversity and differ in their beliefs regarding its effects (see e.g. Stahl-Rolf et al., 2018).

From an empirical perspective, studying the relationship between birthplace diversity and group performance is interesting because theoretical considerations do not allow a clear statement. More specifically, birthplace diversity may promote team performance because workers from different countries are educated under different systems and thus have different skills and experiences (see Alesina et al., 2016, Docquier et al., 2020). By contrast, birthplace diversity may create communication problems because of language barriers and different cultural habits (see Lang, 1986, Lazear, 1999). Furthermore, inter-group biases and internal conflicts are more likely to exist if individuals from various countries collaborate with each other (see Earley and Mosakowski, 2000, Alesina and La Ferrara, 2005).

Providing causal evidence on the effect of birthplace diversity on team performance is challenging for a number of reasons. A main challenge is to find an environment where working teams are clearly identifiable and people from different countries collaborate. Another difficulty is to satisfy the data requirements for such an empirical analysis. In

particular, we need objective measures of performance and data on workers' abilities. Finally, we must develop an identification strategy that is appropriate for establishing causality.

We exploit hand-collected and web-scraped data from the highest division of German male football (*Bundesliga*) to address the challenges that arise when investigating the relationship between birthplace diversity and team performance.[1] The *Bundesliga* is an attractive environment for our purposes since we clearly observe who collaborates with whom and have objective measures of team performance. In addition, freely accessible sources provide detailed information about football players, including their country of birth. Especially remarkable in this regard are time-varying measures of quality that we obtain from the video game series *FIFA* and the web platform *Transfermarkt*. In total, our dataset consists of 7,038 *Bundesliga* matches and 3,266 players, coming from almost 100 countries.

The most common approach when investigating how diversity affects team (or firm) performance is to estimate two-way fixed effect regression models. In these models, the measures of diversity are typically based on ex-post observable information about the composition of the team (see Dezsö and Ross, 2012, Haas and Nüesch, 2012, Kahane et al., 2013, Prinz and Wicker, 2016, among others). An objection against this popular approach is that it neglects compositional changes that happen during the production process. This neglect is problematic since the performance of a team may influence the extent to which the composition of a team changes over time. For instance, imagine a research unit of three junior scientists that has limited time to develop a new software package. Suppose further that these junior researchers realize after three-fourth of the time that they are stuck and thus ask a senior scientist to join the team. The ex-post observable information on the composition of the team therefore provides an inaccurate picture in this example and hides important dynamic adjustments.

In our specific case, such adjustments are a severe issue. In particular, we illustrate that the level of birthplace diversity increases throughout a match if a team does not perform well (in the case above, it would be the tenure level of the research group). We explain this pattern with the fact that team managers replace defensive with offensive players if their team is behind. These performance-based substitutions increase the level of birthplace diversity since the share of foreigners in *Bundesliga* clubs is often larger among offensive than among defensive players. Consequently, when applying a two-way fixed effect approach in our setting, we underestimate the optimal level of birthplace diversity.

To address the aforementioned and other endogeneity problems, we use instrumental variable approaches.[2] The first of these approaches exploits predicted starting line-ups

---

[1] Due to great data quality, using data from the professional sports industry is a popular approach in economics. Examples of studies that exploit such data include: Alvarez et al. (2011), Arcidiacono et al. (2017), Bryson and Chevalier (2015), Dohmen (2008a,b), Dohmen and Sauermann (2016), Frick (2011), Guryan et al. (2009), Kleven et al. (2013), Krumer and Lechner (2017, 2018), Lichter et al. (2017), Papps et al. (2011), and Price and Wolfers (2010).

[2] Unobserved time-varying manager characteristics are another notable source of bias in our fixed effect

published by the leading football magazine *Kicker*. More specifically, we compute how diverse the players in these predicted starting line-ups are and use this measure as an instrument for the diversity of the fielded players. Our second approach resembles the approach by Bettinger and Long (2005) who exploit unexpected absences of university lectures to study the importance of role models for study choice. In our case, we use unexpected replacements in the starting line-ups. Conditional on our large set of fixed effects and control variables, we believe that these changes produce plausible exogenous variation in team structures. In a robustness check, we only exploit those unexpected replacements for which we know that they are caused by an injury.

Our instrumental variable regressions provide evidence for a non-linear relationship between birthplace diversity and team performance. In particular, we observe that the effect of birthplace diversity on team performance increases up to a certain level and becomes smaller beyond this threshold. Put differently, our estimates imply that an intermediate level of birthplace diversity maximizes team performance. A variety of robustness checks substantiates this view. We also present estimates suggesting that the optimal level of birthplace diversity varies across tasks and depends on the importance of interpersonal communication.

This paper contributes to the literature that links birthplace diversity to economic performance. At the macro level, most existing studies have identified positive effects of birthplace diversity (for country-level studies, see e.g. Alesina et al., 2016 & Bove and Elia, 2017; for regional-level studies, see e.g. Ager and Brueckner, 2013, Docquier et al., 2020, Hornung, 2014, Ottaviano and Peri, 2005 & Suedekum et al., 2014). The common explanations are complementary skills and business specializations. However, the macro studies remain inconclusive about situations in which diverse individuals are forced to directly collaborate At the micro level, two types of studies exist. The first consists of studies that conduct (field or laboratory) experiments. The results of these experiments are mixed. For instance, Earley and Mosakowski (2000) conclude that the relationship between birthplace diversity and team performance is U-shaped, whereas Lyons (2017) finds negative effects. To explain her result, Lyons (2017) provides evidence, suggesting that nationally diverse teams suffer from communication problems. The second type of studies analyzes observable data with fixed effect approaches. In many cases, this data comes from the sports industry. An exception is Freeman and Huang (2015) who study research groups and find that birthplace diversity enhances performance. By contrast, most of the studies that exploit sports data conclude that birthplace diversity reduces team performance (see e.g. Addesa et al., 2022, Haas and Nüesch, 2012, Kahane et al., 2013, Maderer et al., 2014).[3] We show results from instrumental variable regressions, suggesting that the relationship between birthplace diversity and team performance is hump-shaped. We also illustrate that the classical fixed effect approach produces much

---

regressions.

[3] A notable exception is the study by Balsmeier et al. (2019) who show that the increase in birthplace diversity that was caused by the Bosman ruling enhanced team performance. Their result is consistent with our results since prior to this ruling birthplace diversity was much lower than the optimal level of birthplace diversity.

lower estimates for the optimal level of birthplace diversity. More specifically, we show that neglecting performance-based substitutions creates results that are consistent with prior studies, considering comparable circumstances. To the best of our knowledge, our instrumental variable procedures have not been used in other studies that examine how birthplace diversity affects team performance.

More generally, we contribute to the literature that analyzes how diversity in socio-demographic worker characteristics affects team performance. This literature consists of numerous studies, considering several types of diversity and applying various empirical methods (for detailed reviews, see Williams and O'Reilly, 1998, Horwitz and Horwitz, 2007, Joshi and Roh, 2009, van Knippenberg and Mell, 2016, Guillaume et al., 2017).[4] Apart from providing causal evidence on the role of one specific type of diversity, our paper has three contributions.[5] First, we illustrate that regression results can be severely biased if an empirical approach neglects that the composition of a team changes during the production process, depending on how it performs. Since such changes can occur in many industries, we think that such an illustration is important to further improve the awareness for this endogeneity problem. Second, our paper establishes instrumental variable approaches. Since variants of our methods can be applied for other types of diversity and in other institutional environments, our paper may help other scholars to address their problems. Finally, we support the view that external factors moderate the effect of workplace diversity on team performance by showing that the optimal level of diversity depends on the importance of interpersonal communication. In particular, we exploit the division of tasks on the pitch between forwards and defenders and show that the optimal level of birthplace diversity is higher for offensive players than for defensive players. We consider this difference as notable because offensive players have to solve creative tasks, whereas defensive players need to rely more on communication skills to fulfill their duties.

This paper proceeds as follows. Section 2 informs about the *Bundesliga* and our data. Section 3 describes our empirical approaches. Section 4 presents our estimation results. Section 5 concludes.

2   Institutional framework and data

The first challenge when examining how birthplace diversity affects team performance is to find an appropriate institutional environment. Kahane et al. (2013) suggest that such an environment has to meet four requirements. First, individuals from various countries

---

[4]Several studies provide evidence for a non-linear relationship between diversity and team performance. Examples include Staats et al. (2012), Kesavan et al. (2014), Tan and Netessine (2014, 2019), as well as Schwab et al. (2016).

[5]Compared with the huge number of studies that exploit observable data to analyze how diversity affects performance, relatively few studies apply methods that allow for a causal interpretation of the results. Examples include Delis et al. (2017), Kesavan et al. (2014), and Tan and Netessine (2019). All these studies apply an instrumental variable approach. Hoogendoorn et al. (2013) design a randomized field experiments to establish causality (see also Lyons, 2017).

must collaborate with each other. Second, we need to observe who works together with whom. Third, information regarding workers' origins, skills, and experiences have to be available. Fourth, the performance of each team must be objectively measurable. This section shows that these four requirements are satisfied when using the *Bundesliga* as institutional environment.

2.1 Institutional background

The *Bundesliga* is the first division of German male football. It includes 18 clubs and is organized as a double round-robin system. Each club therefore plays 34 matches per season, 17 of these matches at home. Every match day consists of 9 matches, whereby each club participates in one of them. Matches are assigned to match days prior to the start of the season. Usually, matches take place on a weekend.[6] Between the first round (August – December) and the second round (January – May) of every season, a winter break of four or five weeks without matches is made.[7] In both of the rounds, every club plays against all others. The home field advantage changes between the first and second round.

A soccer match lasts 90 minutes and consists of two halves. Prior to each match, the team manager nominates eleven starting players (one goal keeper, ten field players) and seven substitutes. During the match, the manager can substitute up to three players. At the end of the match, the winner obtains three ranking points, while the loser gets no points. In case of a draw, both clubs obtain one point. At the end of a season, the total number of ranking points determines the position in the table. If clubs have the same number of points, the difference between the goals scored and goals conceded serves as decision criterion.

The starting players and the substitutes are chosen out of the squad, which in turn consists of all players hired by the club. Typically, the squad is compiled by the club managers. As in most other football leagues, the club managers can change the squad twice per year. The first transfer period is from July to August, while the second is in January. There are no rules determining the size of the squad, salary caps do not exist, and budgets differ across clubs. The club budget depends on various factors, including ticket and fan article sales, transfer revenues, sponsorship contracts, TV revenues, and monetary rewards for participation in a European club tournament.

The rule governing the fielding of foreign players has only changed twice since the introduction of the *Bundesliga* in 1963. The initial regulatory scheme lasted until the middle of the 1995/96 season and allowed the fielding of three foreign players. In late 1995, the European Court of Justice declared the initial form of the *three-players rule* illegal on the grounds that it is incompatible with the treaties of the European Union.[8]

---

[6]For organizational reasons, some match days have to take place during midweek days (Tuesday and Wednesday). For details, see Krumer and Lechner (2018).

[7]In some seasons, the first match day of the second round takes place before the winter break starts.

[8]This landmark decision is known as *Bosman ruling* and is named after the Belgian Jean-Marc Bosman who sued his club, RFC Liège, because of contractual disputes. For details, see Dobson and Goddard

Figure 1 Trends in birthplace diversity (Bundesliga, 1995/96 – 2017/18).

Afterwards, the *three-players rule* only applied to non-European players.[9] As of to the 2004/05 season, also this restriction was abolished.
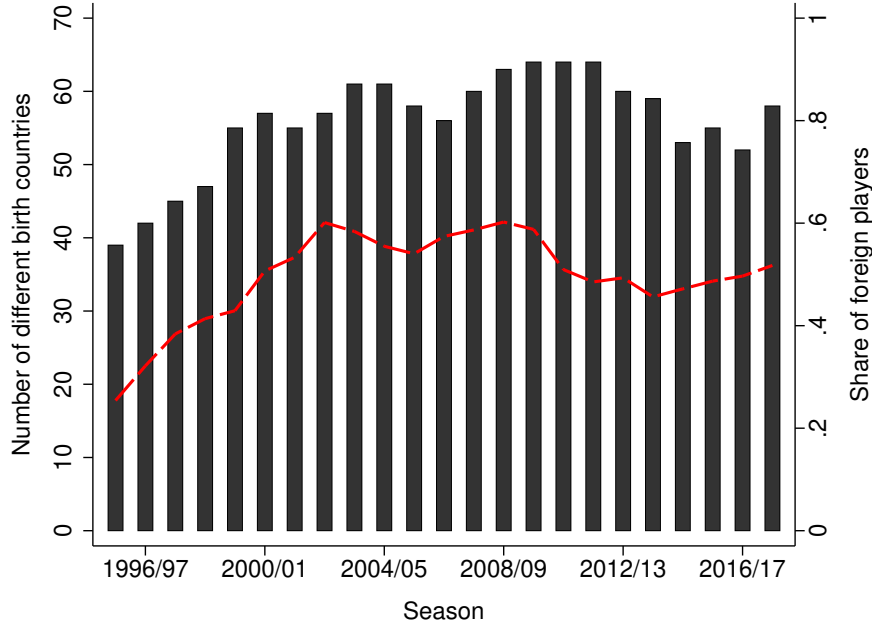
Figure 1 illustrates that the *Bundesliga* is an environment where people from many countries collaborate. The dashed red line in Figure 1 shows how the share of foreign players changed over time (right scale). In the 1995/96 season, only about one quarter of the *Bundesliga* players was not German-born. Caused by the relaxation of the *three-players rule*, the share of foreign players started to increase in the 1996/97 season and peaked at 60 percent in the 2002/03 season. Since the 2010/11 season, about half of the *Bundesliga* players are German-born.

Figure 1 also includes a bar chart, suggesting that foreign *Bundesliga* players come from various countries (left scale). The lengths of the bars reflect the total number of birth countries. Between the 1995/96 season and the 2003/04 season, this number grew from 39 to 61. Since then, the number of birth countries varies between 52 and 64.

(2011), Kleven et al. (2013), and Simmons (1997).

[9]This rule had, however, rather little practical relevance since most of the non-Europeans had a dual citizenship of their country of birth and an European country.

2.2 Data

For our analysis, we create a unique data set, including information on all *Bundesliga* matches from the 1995/96 season up to the 2017/18 season.[10] For each of these 7,038 matches, we know the date, the participating clubs, the final result, and the place at which the match took place. We also observe the starting line-ups, have information on substitutions, and know the ranking position of each club. Furthermore, we compile the names of the team managers and several manager characteristics (e.g. age, tenure, birth country). Finally, we produce dummies that indicate whether a club participated in a match of a European club tournament or the national cup in the week just before or immediately after a *Bundesliga* match. Our key source of information was the German football magazine *Kicker* and its homepage (www.kicker.de).

In total, 3,266 players participated in the 7,038 matches that are included in our database. For each of them, we know the country of birth. The total number of birth countries is 98 (for a full list, see Table D.1). Apart from Germany, the most frequent countries are Brazil, France, Poland, and Denmark. For all players, we also know their date of birth, the date when they were hired by a *Bundesliga* club, and the number of matches that they have played prior to a match in (i) the *Bundesliga*, (ii) the highest league in France, England, Italy, and Spain, (iii) European club tournaments, and (iv) European and world championships.[11] All information are collected by hand and come from the website of *Kicker* and the web platforms *Transfermarkt* (www.transfermarkt.com) and *World Football* (www.worldfootball.net). Lastly, we complement our objective information with two expert-based measures of players' quality. The first is the market value of the player as reported by *Transfermarkt*. Our second measure comes from the video game *FIFA* which is released by the video game company *Electronic Arts*. In this very popular video game, each player has a playing strength ranging from 0 to 100. We develop a PHP script to download the ratings from the *FIFA* web platform (www.fifa-index.com). A common feature of our quality measures is that they are updated at the beginning of each season and thus time-variant. Unfortunately, both measures are only available for the 13 latest seasons (2005/06 – 2017/18).[12]

---

[10] We began the data collection with the 1995/96 season for three reasons: (i) prior to this season, it was prohibited to field more than three foreigners, (ii) as of this season, the winner of a match got three ranking points rather than two, and (iii) data quality.

[11] We choose the leagues of France, England, Italy, and Spain because the level of play in these league is similar to the level of play in Germany.

[12] A question may be whether our the market values reported by *Transfermarkt* and the *FIFA* ratings constitute meaningful measures of players' quality. To provide some evidence on this matter, we check how our two measures correlate with individual performance outcomes, such as minutes played, goals scored, and assists made. The results are shown in Figures C.1 and C.2. We find a strong positive correlation.

3 Empirical framework

3.1 Fixed effect approach

Our analysis of the impact of birthplace diversity (B) on team performance (Y) begins with the following regression model:

$$Y_{isrmd} = \mathfrak{F}(B_{isrmd}) + \gamma \cdot \mathbf{X}_{isrmd} + \Lambda_{isrm} + \Theta_d + \varepsilon_{isrmd} \tag{1}$$

where i denotes a club, s a season, m a manager, and d a match day. The subscript $r \in \{1, 2\}$ indicates whether a match took place in the first or the second round of a season. $\mathbf{X}$ is the set of controls, Λ the set of club-by-season-by-round-by-manager fixed effects, and Θ the set of match day fixed effects.

In contrast to most other sectors, team performance can be directly and objectively measured when using data from professional sports (Kahane et al., 2013). As in most studies that exploit such data, we apply two measures of team performance. The first performance measure is the difference between the goals scored and goals allowed. Our second measure of team performance is the total number of ranking points that a club obtained at the end of a match.

Our key explanatory variable is the birthplace diversity of a team. As Alesina et al. (2016), we define birthplace diversity as the likelihood that two randomly drawn team members were born in two different countries. Consequently, we measure the birthplace diversity of a team with the following fractionalization index:

$$B = 1 - \frac{1}{n^2} \sum_{l=1}^{n} \sum_{k=1}^{n} b_{lk} \tag{2}$$

where b_{lk} denotes a dummy that equals 1 when players l and k have the same country of birth, and 0 otherwise. The parameter n reflects the total number of players that participate in a match. In theory, our measure of birthplace diversity can range from 0 (homogeneous) to 1 (completely diverse).[13] In Figure 2, we present a histogram that illustrates how the level of birthplace diversity is distributed in *Bundesliga* teams. We observe that homogeneous teams are rather rare and that most teams have a level of birthplace diversity between 0.55 and 0.85.

As sketched in Section 1, birthplace diversity may have both positive and negative aspects. The effect of birthplace diversity on team performance might therefore be non-linear.[14] To test for non-linearity, we assume a quadratic functional form in our main analysis:

$$\mathfrak{F}(B_{isrmd}) = \beta_1 \cdot B_{isrmd} + \beta_2 \cdot B^2_{isrmd}. \tag{3}$$

---

[13]As combinations with $l = k$ are not excluded when calculating B (see Bossert et al., 2011), $B = 1$ is possible only if $n \to \infty$. In our case, the maximum value of B is $\frac{13}{14}$.

[14]In Appendix A, we present a simple theoretical framework that illustrates why the this effect may be non-linear and concave.

Figure 2 Birthplace diversity in *Bundesliga* teams.
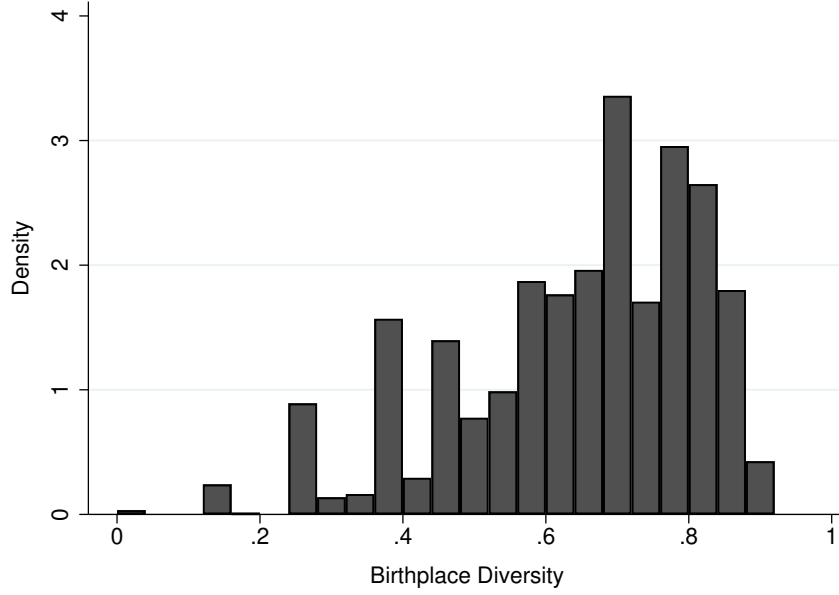
The optimal level of birthplace diversity is then:

$$B^* = -\frac{\beta_1}{2 \cdot \beta_2} \quad \text{with} \quad \beta_1 \in [0, -2 \cdot \beta_2] \quad \text{and} \quad \beta_2 < 0. \tag{4}$$

To alleviate potential concerns regarding our procedure, we relax the functional form assumptions in our robustness checks. Of course, we also present estimates from linear regressions.

When estimating (1), we exploit variation within clubs, while controlling for a large number of potential confounders. First, our cross-sectional fixed effects (Λ) capture all factors that are specific to a club in a round of a particular season if a specific team manager is in charge. Among others, such factors are the abilities of the manager and other staff members, the composition of the squad, the budget and prestige of a club, cultural differences between manager and players, the quality of training facilities and youth sections, and the fan base. Our fixed effects also ensure that our results are not biased due to general time trends. In addition, to control for team characteristics that vary between matches, we add variables to our regression models that reflect players' experience and quality (for details, see Section 2.2). Lastly, we control for a variety of time-varying manager and opponent characteristics, and the home field advantage. In Table D.2, we list all match-specific control variables.

3.2 Instrumental variable approaches

Our OLS estimates of β_1, β_2, and B^* might be biased for two major reasons. First, the composition of a team changes endogenously during a match due to substitutions. In particular, as shown in Appendix B with within-match data, poorly (well) performing teams increase (decrease) their birthplace diversity in the course of a match. The key reason for this change is that the managers replace defensive (offensive) with offensive (defensive) players if their team is behind (in front). In the *Bundesliga*, this (rational) behavior reduces (increases) birthplace diversity because the share of foreign offensive players is much larger than the share of foreign defensive players. From an empirical perspective, this development is problematic since it biases our OLS estimates of B^* downwards. The second potential source of bias is that our comprehensive sets of fixed effects and control variables do not suffice to control for all confounding factors. For instance, it might be that the relationship between the manager and the players (or at least some of them) changes between two games, which in turn might affect both the line-up and the performance of the team. It might also be the case that a manager has more knowledge about specific opponents, for instance, because he or one of his staff members worked for this club in the past. The direction in which an unobserved confounder biases our OLS estimates is unclear.

In our empirical analysis, we use two instrumental variable approaches to address the aforementioned endogeneity problems. Below, we first describe them (see Section 3.2.1 and 3.2.2) and then explain why they complement each other (see Section 3.2.3).

3.2.1 Diversity in predicted starting line-ups

In the literature that investigates how diversity affects team performance, only a few studies apply an instrumental variable approach to allay endogeneity concerns (see e.g. Delis et al., 2017, Kesavan et al., 2014, Tan and Netessine, 2014, 2019). Among these studies, a popular approach is to exploit a lagged value of the variable of interest as instrument. Obviously, when using this approach, the regression coefficients cannot be biased due to reverse causality. In addition, the first-stage relationship is likely to be strong in this case because working groups often change slowly over time. However, a general objection against using a lagged value as an instrument is that the exclusion restriction might not be satisfied (Angrist and Pischke, 2009, 2010, Bazzi and Clemens, 2013). In our case, the risk that the exclusion restriction is violated when the level of birthplace diversity in the previous match is used as instrument is non-negligible. For instance, as outlined above, we are worried about unobserved manager characteristics. Since the previous line-up is usually made by the same manager and thus affected by these characteristics, we think that it is legitimate to question whether using a lagged value of our main variable (B_{isrmd}) addresses this endogeneity concern in a convincing manner.

Our alternative to the standard instrumental variable approach is a procedure from

which we think that it shares the strength of the standard approach and decreases the likelihood that the exclusion restriction does not hold. More specifically, we exploit the birthplace diversity in predicted line-ups published by Germany's best selling football magazine (*Kicker*) at the day before the first match of a match day starts. To realize our approach, we first manually digitize all 7,956 expected starting line-ups that were published between 07/2005 and 06/2018.[15] In the second step, we compute the level of birthplace diversity in all predicted starting line-ups. Finally, we estimate the first-stage equations:

$$B_{isrmd} = \rho_1 \cdot K_{isrmd} + \rho_2 \cdot K^2_{isrmd} + \alpha \cdot \mathbf{X}_{isrmd} + \Lambda_{isrm} + \Theta_d + \eta_{isrmd} \quad (5)$$

$$B^2_{isrmd} = \delta_1 \cdot K_{isrmd} + \delta_2 \cdot K^2_{isrmd} + \lambda \cdot \mathbf{X}_{isrmd} + \Lambda_{isrm} + \Theta_d + \mu_{isrmd} \quad (6)$$

where K denotes the level of birthplace diversity in the starting line-up predicted by *Kicker*.[16]

In Figure 3, we inform about the relationship between the starting line-ups and the line-ups predicted by *Kicker*. The upper graph illustrates that 87 percent of all starting players also belong to the predicted starting line-up. The share of correctly predicted German players differs only slightly from the share of correctly predicted non-German players. The strong overall overlap is reassuring because it suggests a strong first-stage relationship. A concern might be that *Kicker* correctly predicts the starting line-up in most cases and makes serious mistakes in a few specific cases. In the middle graph of Figure 3, we try to allay this concern. In particular, we illustrate that *Kicker* correctly predicts the line-up in only 23 percent of the cases. We also show that the number of incorrect predictions per starting line-up is usually small. The lower graph of Figure 3 highlights which consequences the incorrect predictions have for the predicted levels of birthplace diversity. In 48 percent of the cases, the level of birthplace diversity is the same in the actual and the predicted starting line-up. Great discrepancies are relatively rare.

3.2.2 Unexpected replacements

Our second instrumental variable approach is inspired by several studies in education economics, especially Bettinger and Long (2005) and Herrmann and Rockoff (2012). The common feature of these influential studies is that they exploit unexpected absences of instructors. We adapt their idea and thus aim to use the extent to which the level of birthplace diversity changes due to unexpected replacements as an instrument for the actual level of birthplace diversity among the fielded players.

To implement our second instrumental variable approach, we proceed as follows. In a first step, we define three different groups of players (for an example, see Figure C.3).

---

[15] We do not expand the sample to the earlier seasons because of limited data availability and low data quality.

[16] A potential question is why we exploit the starting line-up predicted by *Kicker* instead of the actual starting line-up to build our instruments. The advantage of the former is that it is made by journalists rather than the manager. Therefore, our estimates are less likely to be confounded by an unobserved manager characteristic when using the birthplace diversity in the line-up predicted by *Kicker*.

Figure 3 Accuracy of the predicted starting line-up.
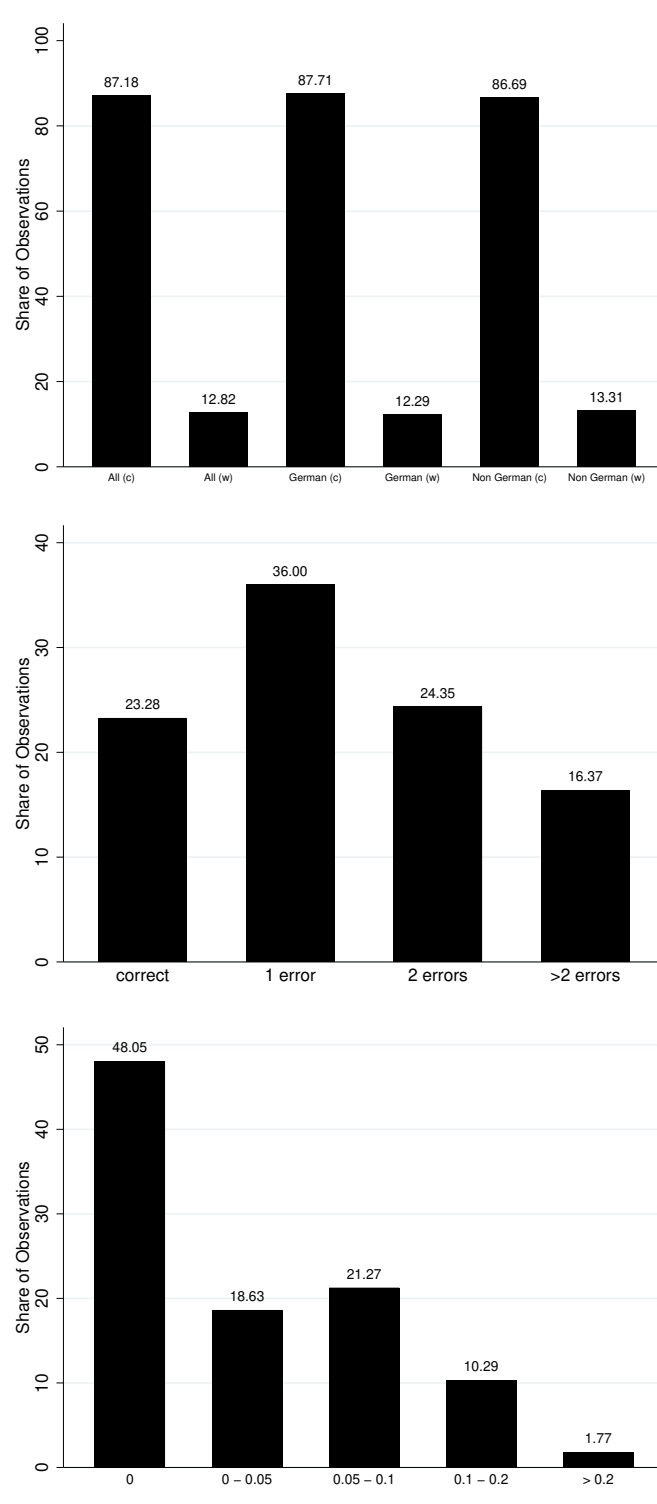
Notes: The upper graph shows how likely it is that a player who is part of the line-up predicted by *Kicker* also belongs to actual starting line-up. The middle graph shows how likely it is that there are no (one, two, or more than two) discrepancies between the starting line-up predicted by *Kicker* and the actual starting line-up. The lower graph shows how much the level of birthplace diversity in the line-up predicted by *Kicker* deviates from the level of birthplace diversity in the starting line-up.

Group $\mathcal{A}$ consists of all players that are not starting players but belong to the line-up predicted by *Kicker*. Group $\mathcal{B}$ includes starting players that are not part of the expected starting line-up. Group $\mathcal{C}$ includes all remaining starting players. In a second step, we calculate the dissimilarity between the players in $\mathcal{C}$ and $\mathcal{A}$ ($\mathcal{B}$):

$$\Delta(\mathcal{A},\mathcal{C}) = \frac{1}{|\mathcal{A}|} \cdot \frac{1}{|\mathcal{C}|} \cdot \sum_{j \in \mathcal{A}} \sum_{k \in \mathcal{C}} (1 - s_{jk}) \tag{7}$$

$$\Delta(\mathcal{B},\mathcal{C}) = \frac{1}{|\mathcal{B}|} \cdot \frac{1}{|\mathcal{C}|} \cdot \sum_{j \in \mathcal{B}} \sum_{k \in \mathcal{C}} (1 - s_{jk}) \tag{8}$$

where s_{jk} denotes a dummy that is equal to 1 if players j and k come from the same country, and 0 otherwise. In the final step, we create the instrument $Z \in [-1, 1]$ as the difference of two dissimilarity scores:

$$Z = \Delta(\mathcal{B},\mathcal{C}) - \Delta(\mathcal{A},\mathcal{C}). \tag{9}$$

Our instrumental variable only differs from 0 if there is an unexpected change in the starting line-up and if the players in $\mathcal{A}$ and $\mathcal{B}$ were not born in the same countries. In particular, Z is positive only if the dissimilarity between the players that unexpectedly belong to the starting line-up and the other starting players ($\Delta(\mathcal{B},\mathcal{C})$) is larger than the dissimilarity between the players that unexpectedly dropped out from the starting line-up and the players that belonged to both the actual and the expected starting line-up ($\Delta(\mathcal{A},\mathcal{C})$). As intended, we observe a positive correlation between Z and the birthplace diversity of the fielded players (see Figure C.5).

As mentioned in Section 3.1, we test for a hump-shaped relationship between team performance and birthplace diversity in our analysis. To this end, we need instrumental variables for both the level of birthplace diversity and its squared term. As Ashraf and Galor (2013), we use a three-stage approach to meet this challenge.[17] First, we perform a zero-stage regression in which our measure of birthplace diversity is regressed on Z and the controls of the second-stage equation:

$$B_{isrmd} = \phi \cdot Z_{isrmd} + \pi \cdot \mathbf{X}_{isrmd} + \Lambda_{isrm} + \Theta_d + \zeta_{isrmd}. \tag{10}$$

Afterwards, we use the point estimates that we get when estimating (10) to compute predicted values of birthplace diversity ($\widehat{B}$). Finally, we use these predicted values and their squared terms as our instrumental variables. Therefore, the first-stage equations

---

[17]For details about the econometric foundation of the approach used by Ashraf and Galor (2013), see Angrist and Pischke (2009) and Wooldridge (2010). We need a zero-stage regression since Z^2 and B^2 are only weakly correlated. Two reasons exist for this weak correlation: first, $Z \in [-1, 1]$ and $B \in [0, 1]$ have different domains, and second, squaring is a non-monotonic transformation if the domain is $[-1, 1]$, while it is a monotonic transformation if the domain is $[0, 1]$.

are:[18]

$$B_{isrmd} = \rho_1 \cdot \widehat{B}_{isrmd} + \rho_2 \cdot \widehat{B}^2_{isrmd} + \alpha \cdot \mathbf{X}_{isrmd} + \Lambda_{isrm} + \Theta_d + \eta_{isrmd} \qquad (11)$$

$$B^2_{isrmd} = \delta_1 \cdot \widehat{B}_{isrmd} + \delta_2 \cdot \widehat{B}^2_{isrmd} + \alpha \cdot \mathbf{X}_{isrmd} + \Lambda_{isrm} + \Theta_d + \mu_{isrmd}. \qquad (12)$$

3.2.3 Discussion

At the beginning of this section, we explain that our OLS estimates on the effects of birthplace diversity on team performance are likely to be biased for two main reasons. The first reason is reverse causality because the way of how the composition of teams changes in the course of the match is determined by their performance (for details, see also Appendix B). We can allay this endogeneity issue with both of our instrumental variable approaches because when producing our instruments, we only use information that is already available prior to the kick-off of a match. The performance during the match thus plays no role for the instrumental variables.

The second source of bias in the OLS estimates is unobserved confounding factors. In particular, we are concerned about time-varying manager characteristics. With our two instrumental variable approaches, we can allay this concern if (i) our instruments affect (conditional on our fixed effects and controls) team performance only via the level of birthplace diversity and (ii) the first-stage relationships are sufficiently strong. In our analysis, we report first-stage F-statistics to show that the second condition is satisfied. Regarding the exclusion restriction, concerns may exist with respect to each of our two strategies.[19] For instance, a legitimate concern about our second instrumental variable procedure is that differences between the starting line-up predicted by *Kicker* and the actual starting line-up have multiple reasons and that only some of them (e.g. injuries, late arrivals after international matches, birth of a child) are plausibly exogenous. Put differently, it is plausible to wonder whether some of the discrepancies are due to an unobserved manager characteristic. If so, the exclusion restriction may not hold in the second instrumental variable approach. Importantly, our first approach is unaffected by manager characteristics that cause last-minute changes in the starting line-up because *Kicker* exploits publicly available information and releases its predictions one to three days before a match starts. Of course, this fact does not imply that using the predicted line-ups is necessarily the superior approach. For instance, since *Kicker* journalists are experts, it may be that they have all information that determines the decisions of the manager, while our controls and fixed effects fail to capture all determinants. If so, the exclusion restriction is violated when applying our first approach. However, if journalists are perfectly informed, they are able to anticipate the last-minutes changes that are not

---

[18]It is irrelevant for the second-stage estimates of β_1, β_2, and B^* whether we use the variable Z or the predicted value $\widehat{B}$ in the first-stage equations. We choose $\widehat{B}$ because it eases the interpretation of the reduced-form estimates.

[19]From a methodological perspective, our second instrumental variable approach has a clearer data generating process than the first approach. Therefore, it is easier for the second approach to evaluate whether the exclusion restriction holds. A key advantage of the first approach is that its implementation is much more straightforward.

caused by injuries or other unforeseeable events. Consequently, the differences between the line-up predicted by *Kicker* and the true line-up can then be fully explained with such plausibly exogenous factors, which in turn validates our second instrumental variable approach.

Taken together, we think that both of our instrumental variable strategies allay the problem of reverse causality, while they may violate the exclusion restriction. However, the reasons for why our two approaches might not satisfy the exclusion restriction are different and we are thus convinced that they nicely complement each other. In Section 4.2.1, we present an additional way to address concerns against our two instrumental variable approaches. Basically, our idea is to exploit only those unexpected changes for which we have evidence that they are due to injuries. While the exclusion restriction is more likely to be met when applying this extended version of our second instrumental variable approach, the first-stage relationship weakens. Because of the large number of studies that warn against weak instruments, we consider this extension as a robustness check for our two baseline approaches rather than using it as our primary identification strategy.

4   Results

4.1   Main results

4.1.1   Fixed effect estimates

Table 1 shows the main results of our empirical analysis. Columns 1 and 2 report our fixed effect estimates, while Columns 3 and 4 present results from instrumental variable regressions. In all specifications, we use the goal difference at the end of the match as measure of team performance and cluster the standard errors at the club-by-season-by-round-by-manager level. In Panel A, we test for a linear relationship between birthplace diversity and team performance. Panel B investigates whether this relationship is non-linear. When testing for non-linearity, we assume a quadratic functional form. To keep Table 1 manageable, we only show the estimates of β_1, β_2, and B*. In parentheses, we report the p-values of our estimates.

When making a quadratic functional form assumption, the common way to examine whether a relationship is hump-shaped is to check whether the quadratic term has a negative and statistically significant impact on the outcome variable. Lind and Mehlum (2010) argue that this conventional procedure creates misleading results if the optimal level is not within the lower and upper bound of the data range (see also Simonsohn, 2018). More specifically, the estimate of β_2 indicates whether the relationship between birthplace diversity and team performance is concave, but it does not show whether the optimal level of birthplace diversity is between 0 and 1. Lind and Mehlum (2010) also present a statistical test that addresses this issue. We apply this test and thus structure Panel B in the following manner. The upper part shows the results of the Lind-Mehlum-Test,

consisting of an estimate of the optimal level of birthplace diversity (B*) and a p-value that reveals the presence of a hump-shaped relationship within the data range. In the lower part, we report the estimates of β_1 and β_2 and their p-values.

Column 1 uses a sample that consists of 7,038 *Bundesliga* matches to examine the relationship between birthplace diversity and team performance (for summary statistics, see Table D.3). In the linear regression, we find evidence for a negative and statistically significant correlation between birthplace diversity and team performance (see Panel A). The non-linear model suggests that the optimal level of birthplace diversity is 0.148. However, the Lind-Mehlum-Test suggests that B* is not statistically different from 0 (p-value: 0.39). According to our first fixed effect model, we therefore need to reject the hypothesis that the relationship between birthplace diversity and team performance is hump-shaped.

A concern regarding our first fixed effect analysis might be that we do not adequately control for the quality of the teams. This problem might exist since our two measures of players' quality (market value, strength in the video game *FIFA*) are only available for the 13 latest *Bundesliga* seasons. Column 2 of Table 1 thus analyzes a reduced sample, including 3,978 matches (for summary statistics, see Table D.4). Our estimates hardly change compared to the estimates reported in Column 1. More specifically, in Panel B, the estimated optimal level of birthplace diversity is still fairly small ($\hat{B}^* = 0.129$) and not statistically different from 0 (p-value: 0.44). In Panel A, the estimate of β_1 remains negative and statistically significant at conventional levels.

In sum, our fixed effect analyses provide evidence for a negative rather than a hump-shaped relationship between birthplace diversity and team performance. Our results are therefore consistent with other studies that use data from the sports industry and fixed effect approaches (see e.g. Addesa et al., 2022, Haas and Nüesch, 2012, Kahane et al., 2013, Maderer et al., 2014). These studies also suggest explanations for why nationally diverse teams perform less well than nationally homogeneous teams. The most common arguments are that heterogeneous teams suffer from cultural and/or linguistic barriers, have lower group coherence, and are more prone to inter-group biases. A weak spot of these explanations is that they leave open why the actual levels of birthplace diversity differ from 0. For instance, the average level of birthplace diversity is 0.67 in our full sample (for the distribution, see Figure 2). Put differently, it remains unclear why the managers field nationally diverse teams if birthplace diversity has a (strong) negative impact on team performance.[20] One potential answer to this question is that managers might heavily overestimate (underestimate) the positive (negative) aspects of birthplace diversity. However, as the managers and their staff members are experts and frequently analyze the performance of their team, we have some doubts whether such gross mis-judgment is plausible.[21]

---

[20]The fixed effect estimates reported in Columns 1 and 2 of Table 1 suggest that the goal difference decreases by almost one goal per match if a nationally homogeneous team will be replaced by a highly heterogeneous team.

[21]Another potential explanation might be that foreign players are substatially cheaper than domestic

Table 1 Birthplace diversity and team performance (main results).

|	(1)	(2)	(3)	(4)
Panel A: Linear regression model				
Regression coefficients				
Birthplace divers. ($\hat{\beta}_1$)	-0.903***	-0.883**	0.059	0.296
	(0.002)	(0.020)	(0.956)	(0.858)
First-stage F-statistic	-	-	640.5	272.80
Panel B: Non-linear regression model				
Lind-Mehlum-Test				
Optimal level of birthplace diversity ($\hat{B}^*$)	0.148	0.129	0.608*	0.605**
	(0.387)	(0.436)	(0.073)	(0.023)
Regression coefficients				
Birthplace divers. ($\hat{\beta}_1$)	0.311	0.238	5.372	41.924**
	(0.774)	(0.873)	(0.103)	(0.044)
Birthplace divers. sq. ($\hat{\beta}_2$)	-1.048	-0.927	-4.420*	-34.670**
	(0.234)	(0.435)	(0.094)	(0.043)
First-stage F-statistic (B)	-	-	126.3	13.7
First-stage F-statistic (B^2)	-	-	286.5	14.3
Estimation technique	OLS	OLS	IV	IV
IV approach	-	-	1st	2nd
Outcome variable	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Observations	14,076	7,956	7,956	7,956
Seasons	23	13	13	13
Fixed effects	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes
Quality controls	No	Yes	Yes	Yes
Sample mean (B):	0.669	0.699	0.699	0.699

Notes: The table reports results from fixed effect and instrumental variable regressions. We cluster standard errors at the club-by-season-by-round-by-manager level and show p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

4.1.2 Instrumental variable estimates

We propose another explanation for why conventional two-way fixed effect approaches create estimates of the optimal level of birthplace diversity that are much smaller than the average birthplace diversity of the fielded players. Our key explanation is that these approaches do not account for endogenous changes in the team structure that happen during a match. The consequence of this problem is that we underestimate the optimal level of birthplace diversity. More specifically, as shown in detail in our Appendix B, we argue that managers replace defensive with offensive players during the match if their team performs poorly. In the *Bundesliga*, this rational behavior increases the level of birthplace diversity since the share of foreign offensive players is much larger than the share of foreign defensive players. Put differently, we are convinced that our fixed effect estimates are biased due to reverse causality. As explained in Section 3.2, we use two instrumental variable approaches to address this and other endogeneity concerns. The first exploits the birthplace diversity in the line-up predicted by the football magazine

players and that the benefits of increasing team quality thus outweigh the potential adverse effects of diversity.

Figure 4 Birthplace diversity and team performance.
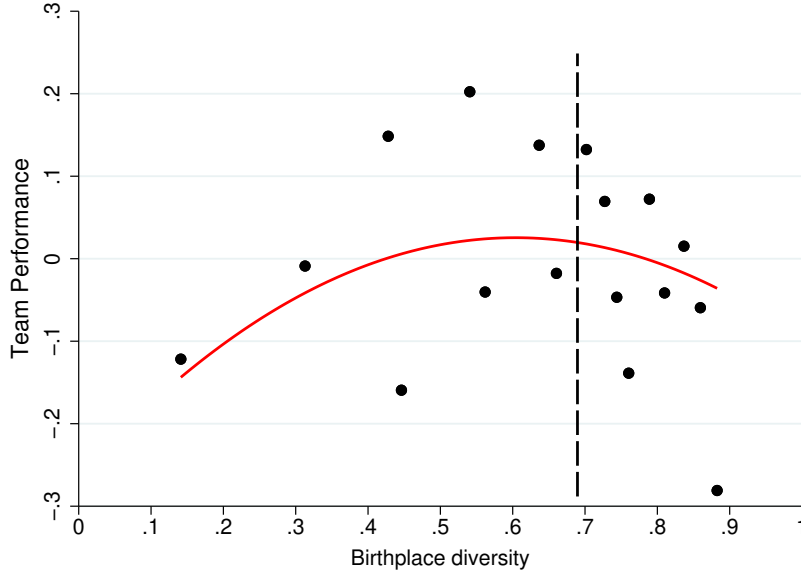
Notes: This figure presents a graphical illustration of the results of our first instrumental variable approach. The red solid line shows the predicted relationship between birthplace diversity and team performance. The dashed line illustrates the average level of diversity. For the sake of vividness, we use a binned scatterplot to present the underlying raw data.

Kicker as instrumental variable. Our second approach uses unexpected replacements in the starting line-up as source of plausibly exogenous variation.

Column 3 of Table 1 presents the results of our first instrumental variable approach (first-stage and reduced-form estimates can be found in Tables D.5 and D.6). Compared with the results of our fixed effect models, we observe four key differences. First, the estimate of β_2 is now statistically significant at the 10 percent level. Second, the implied optimal level of birthplace diversity is 0.609 and thus much larger than the fixed effect estimates of B*. Third, the results of the Lind-Mehlum-Test provide statistical evidence for the presence of a hump-shaped relationship between birthplace diversity and team performance (for a graphical illustration of this relationship, see Figure 4). Finally, the linear model produces an estimate of β_1 that does not differ from 0 in a statistically significant manner (p-value: 0.95). This model thus implies that birthplace diversity has no impact on team performance. By contrast, our non-linear model shows that changing the birthplace diversity by 10 percentage points in either direction at the optimal level reduces the goal difference by 0.044, which is about 3.15 percent of the sample mean. Projected to the season as a whole, such a change amounts to a decrease in the goal difference by more than one goal. Of course, when interpreting this result, some caution is necessary because of the strong functional form assumptions. In Sections 4.2.10 and 4.2.11, we provide some supporting evidence for this assumption.

Column 4 of Table 1 shows that our second instrumental variable approach produces similar results as our first approach. A notable difference is that the estimates of β_2 and B* are now statistically significant at the five rather than the ten percent level. Another

difference is that the strength of the first-stage relationship considerably declines when applying the second approach. A weak-instrument bias is nonetheless unlikely since the first-stage F-statistics are well above 10.

Taken together, our two instrumental variable approaches provide empirical evidence for a causal non-linear effect of birthplace diversity on team performance.[22] Our result is consistent with economic models implying that teams benefit from diversity up to a certain degree but that diversity beyond this level has no further benefits or negative effects due to increased communication problems (see e.g. Ashraf and Galor, 2013, van Knippenberg et al., 2004).[23] Interestingly, we also find that both instrumental variable approaches identify an optimal level of birthplace diversity (≈ 0.61) that resembles the average birthplace diversity of the fielded players (≈ 0.70).[24]

4.2 Robustness Checks

4.2.1 Using Injuries as a source of variation

A concern regarding our second instrumental variable strategy may be that some of the unexpected changes are not caused by exogeneous events, but strategic decisions by the manager. Our estimates might thus be biased due to unobserved time-varying manager characteristics. Ideally, we would like to produce an indicator that reflects whether a discrepancy between the line-up predicted by *Kicker* and the actual starting line-up is caused by a truly exogenous event or the result of a manager decision. Unfortunately, building such an indicator is impossible due to data availability reasons. An alternative approach is to focus on one plausibly exogenous reason for unexpected changes and to identify which discrepancies happen due to this reason. We pursue this idea and try to figure out which discrepancies can be traced back to injuries. To obtain the necessary information, we read numerous match reports and use the web platform *Transfermarkt*. We observe that injuries explain at least 30 percent of the cases where *Kicker* makes a wrong prediction. Importantly, the true share is likely to be higher since reporting on injuries is incomplete, especially if it is a small injury.

To exploit the unexpected replacements that are caused by injuries as our source of plausibly exogenous variation, we must make some small adjustments to the three-step procedure described in Section 3.2.2. These changes concern the first step in which we divide the players into groups. Group $\mathcal{A}$ includes players that belonged to the starting line-up predicted by *Kicker*, but did not participate in the match because of an injury. Group $\mathcal{B}$ consists of those starting players that were not part of the expected starting

---

[22]In Figures C.7 and C.8, we report results of leave-one-out analyses in which we iteratively drop one season or club. We do not find that our estimates are driven by a particular season or club.

[23]In Section 4.3, we will present qualitative and quantitative evidence, suggesting that misunderstandings that are caused by linguistic barriers indeed play an important role.

[24]We must leave open why the average level of birthplace diversity exceeds the predicted optimal level. A potential reason is that managers overestimate (underestimate) the positive (negative) aspects of birthplace diversity. Another explanation is that there are internal constraints (e.g. budget restrictions) or market frictions (e.g. limited supply of appropriate players) that prevent managers from fielding a team with the optimal level of birthplace diversity.

line-up and replaced injured players. Group $\mathcal{C}$ includes all other starting players (for an example, see Figure C.4).

The other steps of the procedure are the same as in Section 3.2.2. The second step calculates the average dissimilarity between the players in $\mathcal{C}$ and $\mathcal{A}$ ($\mathcal{B}$):

$$\Delta(\mathcal{A},\mathcal{C}) = \frac{1}{|\mathcal{A}|} \cdot \frac{1}{|\mathcal{C}|} \cdot \sum_{j \in \mathcal{A}} \sum_{k \in \mathcal{C}} (1 - s_{jk}) \tag{13}$$

$$\Delta(\mathcal{B},\mathcal{C}) = \frac{1}{|\mathcal{B}|} \cdot \frac{1}{|\mathcal{C}|} \cdot \sum_{j \in \mathcal{B}} \sum_{k \in \mathcal{C}} (1 - s_{jk}) \tag{14}$$

where s_{jk} denotes a dummy that is equal to 1 only if the players j and k come from the same country. The last step generates the instrumental variable $Z \in [-1, 1]$ as the difference of two dissimilarity scores:

$$Z = \Delta(\mathcal{B},\mathcal{C}) - \Delta(\mathcal{A},\mathcal{C}). \tag{15}$$

As expected, we find a positive correlation between our instrumental variable and the birthplace diversity of the fielded players (see Figure C.6). However, compared to the baseline approach, this correlation is weaker. This result is not surprising because the main approach uses all unexpected replacements, while our augmented approach only exploits a subset.

Column 1 of Table D.7 reports the results from the instrumental variable regression that exploits injuries. We see only small changes when comparing these results with the baseline results. For example, the estimated optimal level of birthplace diversity is now 0.631 (p-value: 0.049) rather than 0.605 (p-value: 0.023) and the p-value of $\hat{\beta}_2$ increases from 0.043 to 0.093. However, caution is needed when interpreting these results as the first-stage F-statistics are only around 10. Nevertheless, we believe that this robustness check confirms the presence of a hump-shaped relationship between birthplace diversity and team performance.

4.2.2 Alternative measure of team performance

As many related studies, our baseline analysis uses the difference between goals scored and goals allowed as measure of team performance. An objection against this approach might be that maximizing the goal difference is only a secondary objective because the ranking position is predominately determined by the number of ranking points. The goal difference is only of relevance if two (or more) clubs have the same number of ranking points. Put differently, a legitimate question is why we use the goal difference instead of the number of ranking points to measure the performance of teams. We prefer the goal difference due to its greater discriminating power.[25]

---

[25]Frequently, the number of ranking points does not indicate performance differences even if they exist. For example, if a team loses a match zero to one, it receives the same number of ranking points as it would obtain if it loses zero to five. Due to this relatively low distinctiveness, we argue that the number of ranking points is a less appropriate measure of team performance than the goal difference.

To allay the concern that our findings are driven by our measure of performance, we rerun our main analyses with the number of ranking points as dependent variable. We observe that the results of these robustness checks differ only little from our baseline regression results (see Table D.9). In particular, we find that both of our instrumental variable approaches provide statistical evidence for a hump-shaped effect of birthplace diversity on team performance. The optimal level of birthplace diversity is 0.58 in both cases. However, compared with our main analyses, the estimates of β_2 and B* are on average slightly less statistically significant. We consider this change as fairly plausible because the number of ranking points is a less sensitive measure of team performance than the goal difference.

Another concern may be that our measures of team performance are noisy since the outcome of a match depends (at least to some extent) on luck. Although this fact does not constitute a problem for our identification strategy, we believe that it needs to be discussed. To this end, we pay attention on seasonal performance because luck plays a much smaller role for aggregated outcomes. More specifically, Figure C.9 suggests that teams that deviate (on average) less from the optimal level of birthplace diversity have more ranking points and a better ranking position at the end of a season. We also find that the level of birthplace diversity in high quality teams differs less from the optimal level of birthplace diversity than in medium or low quality teams (see Figure C.10).[26]

4.2.3 Alternative definition of team

In our baseline analyses, a team consists of all players who participate in a match for at least one minute. This approach might be criticized since it is debatable whether players who are fielded for a few minutes actually have a notable effect on the performance of a team. As a robustness check, we thus count only those players as team members who played for at least 30 minutes. Columns 2 and 3 of Table D.7 show that our results hardly change if we use this alternative definition of "team". We also find similar results if we weight players according to the minutes that they played (see Columns 4 and 5 of Table D.7).[27]

4.2.4 Alternative procedure for measuring unexpected changes

In our second instrumental variable approach, we differentiate three groups of football players and apply a three-stage procedure to create an instrumental variable (Z) that reflects the extent to which the level of birthplace diversity changes unexpectedly (for details, see Section 3.2.2). An alternative is to calculate the difference of the birthplace diversity in the actual line-up and the birthplace diversity in the line-up predicted by

---

[26]We classify clubs into high, medium, and low quality teams based on their long-run performance. Table D.13 reports which clubs belong to which category.

[27]To produce the weighted measures of birthplace diversity, we first divide the match into 90 1-minute intervals and calculate the level of birthplace diversity in each interval. In the second step, we compute the average over these 90 interval-level measures to obtain a weighted overall measure of birthplace diversity.

Kicker. Column 6 of Table D.7 shows that this alternative approach produces almost the same results as our basic approach. We prefer the latter for two reasons: first, the first-stage F-statistics are slightly higher, and second, it is easier to restrict the analysis to those unexpected replacements that are caused by injuries.

4.2.5   Past performance

Another concern may be that we do not adequately control for past performance of a team. Controlling for the performance in previous matches might be important because it affects the self-confidence of the players. Furthermore, the likelihood that a manager unexpectedly changes his starting line-up might depend on the past performance of his team. To rule out that this omitted factor causes a violation of the exclusion restriction and thus biases our estimates, we augment our regression models by the first two lags of the dependent variable. For our first approach, we find that the estimates of β_1, β_2 and B^* become statistically more significant if we control for teams' past performance (see Column 1 of Table D.8). The results that we obtain when using our second instrumental variable approach remain virtually unchanged (see Column 2 of Table D.8).

4.2.6   Migration during childhood

An argument for why birthplace diversity might be conducive to team performance is that people from different countries were trained under different education systems and thus have different skills (Alesina et al., 2016, Freeman and Huang, 2015). Obviously, this argument can only apply if workers receive their basic training in their country of birth. Among others, in our case, this key condition does not necessarily hold because clubs might scout and hire players at very young ages.

To investigate whether migration during childhood indeed gives cause for concern, we proceed in two steps. In the first step, we exploit the online platform *Transfermarkt* to identify for each player the country where he lived during his youth. For 222 out of our 3,266 players (6.8 percent), this county differs from the country of birth. About half of them lived in Germany during their childhood. For almost 100 of the players who are not German-born but grew up in Germany, we have biographical information that suggests why they moved to Germany. In virtually all cases, the parents of the players migrated to Germany in order to work there or since they fled from their home country for political reasons. For only three players, we find some evidence suggesting that they moved to Germany due to an offer from a football club.[28]

To illustrate that our baseline results are not driven by players who moved from one country to another in their childhood, we rerun our main analyses using the country in which the players lived during their youth rather than the country of birth. Columns 3

---

[28]For example, the parents of the former German national team players Miroslaw Klose and Lukas Podolski were labor migrants. The father of the former German national team player Gerald Asamoah was a political refugee. The current captain of the Austrian national team, Julian Baumgartlinger, was hired by a German football club (*TSV 1860 München*) at the age of 13.

and 4 of Table D.8 suggest that our instrumental variable estimates are robust to this methodological change.

4.2.7 Alternative measure of birthplace diversity

From a conceptual perspective, it is straightforward that two individuals either have the same country of birth or were born in two different countries. As a consequence, most studies assume a binary similarity matrix when measuring birthplace diversity (see e.g. Alesina et al., 2016). In our baseline analyses, we follow these studies and treat all the country differences as equivalent. A potential concern regarding this common way for measuring birthplace diversity is that some countries are perceived to be more similar to each other than others and that neglecting these differences in the degree of similarity might create a systematic bias. To address this issue, Bossert et al. (2011) propose a fractionalization index that takes into account different degrees of similarities. The key problem when using this more general fractionalization index is to operationalize the similarity matrix. Put differently, while probably most people agree that the similarities between German and Swiss players are larger than the similarities between players from Japan and Serbia, creating an uncontroversial measure for the extent to which German and Swiss players are more similar is nearly impossible.

Columns 5 and 6 of Table D.8 present results from instrumental variable regressions in which we use a more distinctive approach for measuring the similarity of players from different countries. More specifically, as robustness check, we exploit data on linguistic distance from Spolaore and Wacziarg (2016) to operationalize the degree of similarity.[29] Compared with our baseline estimates, we observe only minor changes. If anything, we obtain slightly smaller estimates of B^* (≈ 0.56). We believe that the robustness of our results is not surprising since our alternative measure of birthplace diversity is highly correlated with our baseline measure.

4.2.8 Dynamic assessment of player's quality

Our basic regression models include two control variables that reflect the quality of the fielded players. The first is the market value as indicated by the web database *Transfermarkt*. The second measure is the playing strength in the popular video game *FIFA*. A concern about both measures might be that they are only updated once a year at the beginning of a season. Thus, at the team level, our quality controls change only if the line-ups change, but not due to form fluctuations. For our analysis, this fact might be problematic because form fluctuations are a reason for changes in the line-up.

To produce an index that reflects the current form of the fielded players, we exploit individual-level information provided by *Kicker*. After each match, *Kicker* grades the

---

[29]Leaving aside that the measurement of linguistic distances is a controversially debated issue, we think for two major reasons that using linguistic distance as measure for similarity is not ideal. The first is that we treat players whose countries of birth differ but have the same official language as equal. The second key weak point is that we put too much emphasis on one particular channel (communication) through which birthplace diversity might affect team performance.

individual performance of all players that played at least 30 minutes. The grade scaling ranges from 1.0 (very good) to 6.0 (very bad) in increments of 0.5.

We proceed in three steps to measure player's current form. The first is to manually digitize the 163,971 grades. In the second step, we compute for each player (p) and each match (q) the difference between his grade (G) and his average grade during the season ($\bar{G}$):

$$\Delta_{p,q} = G_{p,q} - \bar{G}_p \tag{16}$$

In the third step, we calculate a team-level measure of the extent to which the fielded players deviated from their normal form in the previous match:

$$F_{i,q} = \frac{1}{n_{i,q}} \sum_{p=1}^{n_{i,q}} \Delta_{p,q-1} \tag{17}$$

where i denotes a team and n the number of fielded players.[30] F is positive (negative) when the players underperformed (overperformed) in the last match.

Columns 7 and 8 of Table D.8 show how our instrumental variable estimates react when controlling for the current form of the players. Compared with our baseline regressions, we only observe minor changes. For example, the optimal level of birthplace diversity estimated by our first instrumental variable approach shifts from 0.608 (p-value: 0.073) to 0.593 (p-value: 0.066) if F serves as an additional control variable in our regression models.

4.2.9 Alternative clustering methods

In our main regressions, we cluster the standard errors at the club-by-season-by-round-by-manager level and thus in accordance with our main panel dimension. An objection against this strategy may be that the number of observations per cluster is too low for obtaining reliable estimation results.[31] To allay this concern, we show in Table D.10 how our results change when using alternative clustering methods. In Columns 1 and 2, we cluster at the club-by-season level. The number of observations per cluster is 34 in this case. In Columns 3 and 4, we apply a two-way clustering. Following the guidelines of Cameron et al. (2012) and Cameron and Miller (2015), we cluster at the cross-sectional (club-by-season-by-round-by-manager) and the panel dimension (match day). In Columns 5 and 6, we use standard errors that are clustered at the club-by-season and the match day level to determine how statistically significant our estimates are. For each of the alternative clustering methods and both instrumental variable approaches, we find that the Lind-Mehlum-Test confirms the hump-shaped effect of birthplace diversity on team performance.

---

[30]If a particular player did not attend in the previous match, we set Δ equal to 0 and thus implicitly assume that he has his average form.

[31]The maximal number of observations per cluster is 17 when using our baseline approach. The average number of observations per cluster is 13.74

4.2.10 Piecewise linear regressions

Our baseline estimates suggest that the optimal level of birthplace diversity is around 0.6. This result implies that an increase in birthplace diversity enhances (reduces) team performance if the actual level of birthplace diversity is below (above) this threshold. Put differently, if we estimate a linear regression model and restrict the sample to the observations with $B < B^*$ ($B > B^*$), we should observe that birthplace diversity has a positive (negative) effect on team performance. Using our first instrumental variable approach, we show in Columns 1 and 2 of Table D.11 that this hypothesis indeed holds. Notably, the slope parameters of the two regression lines have similar absolute values. This finding suggests that the quadratic functional form assumption made by the Lind-Mehlum-Test is appropriate.

Combining the piecewise regression approach with our second instrumental variable approach is slightly more difficult because it exploits changes in birthplace diversity to create plausibly exogenous variation. More specifically, assume that the actual level of birthplace diversity exceeds the ideal level ($B > B^*$) and that an unexpected change in birthplace diversity takes place ($Z \neq 0$). If this change is positive (negative), birthplace diversity moves away from (towards) the optimal level of birthplace diversity. We thus expect a negative (positive) effect when estimating a linear model and restricting the sample to those observations with $Z > 0$ ($Z < 0$). Columns 3 and 4 of Table D.11 are consistent with this expectation.[32]

4.2.11 Simonsohn's two-lines test

So far, we have used the Lind-Mehlum-Test to check whether the relationship between birthplace diversity and team performance is hump-shaped. A valid concern regarding this approach is that it imposes a quadratic functional form. To address this issue, we exploit an procedure proposed by Simonsohn (2018). The idea behind this method is to estimate two regression lines (one for low and another for high levels of the variable of interest) and to test whether the slope parameters of these regression lines differ in their sign. To determine the threshold value that splits the sample in two parts, Simonsohn (2018) proposes a method, called Robin-Hood-Algorithm, that consists of five steps. In the first step, this algorithm estimates a cubic spline for the relationship between the explanatory variable (x) and the dependent variable (y). Afterwards, it computes the most extreme internal fitted value ($\hat{y}_{max}$). In the third step, the Robin-Hood-Algorithm identifies the predicted values that differ from $\hat{y}_{max}$ by at most one standard deviation. Predicted values that meet this condition belong to the set of potential threshold values ($\mathcal{Y}_{flat}$). In the next step, an interrupted regression model is estimated, using the median value of x within $\mathcal{Y}_{flat}$. The output of this regression will be two slope parameters (z_1, z_2). In the last step, the Robin-Hood-Algorithm produces the actual threshold value as the $\frac{|z_2|}{|z_1|+|z_2|}$th percentile of the x values that are associated with $\mathcal{Y}_{flat}$ (for details and

---

[32]We cannot conduct this robustness check for the case $B < B^*$ since the sample size is too small.

illustrative examples, see Simonsohn, 2018).

When applying Simonsohn's procedure to our data, we obtain a threshold value of 0.62. Reassuringly, this threshold is virtually the same as the predicted optimal level of birthplace diversity that we find when using the Lind-Mehlum-Test. Column 5 of Table D.11 shows that the results of an interrupted instrumental variable regression which uses 0.62 as the threshold as threshold value confirm the presence of a non-linear effect of birthplace diversity on team performance. Importantly, we also observe that both of the slope parameters are statistically significant at conventionally used levels (p-values: 0.069, 0.060).

4.3 Mechanisms and moderating factors

The regression results presented in the previous sections provide strong evidence for a hump-shaped effect of birthplace diversity on team performance. This section discusses potential explanations for this relationship. In addition, we study whether the optimal level of birthplace diversity depends on contextual factors.

Several existing theories suggest that diversity increases the range of skills and thus enhances team performance (see e.g. Alesina et al., 2016, Lazear, 1999). In our setting, this implies that our results are only plausible if football players have country-specific skills. We are convinced that this key prerequisite holds because various sports science studies suggest that country-specific football skills exist. For example, Basevitch et al. (2013) and Sarmento et al. (2013) suggest that players from Italy have particularly high tactic skills, whereas Spanish players have outstanding passing skills (for more further evidence, see Dellal et al., 2011, Mitrotasios et al., 2019, Sarmento et al., 2014).[33]

A common argument for why an increase in birthplace diversity does not necessarily improve team performance is that highly diverse groups are more likely to suffer from communication problems (Lang, 1986, Lazear, 1999).[34] Unfortunately, we cannot run an empirical analysis that directly confirms this assumption. The main reason is that the language skills of *Bundesliga* players cannot be systematically measured due to limited data availability. However, we know various interviews in which players and managers stress the great importance of language skills and interpersonal communication on the pitch (for some examples, see Table D.14). In addition, linguistic studies suggest that communication barriers influence the performance of football teams. More specifically, Kellerman et al. (2005) and Ringbom (2012) run surveys among managers and players in the Netherlands and Finland and find that language problems are perceived as a key reason for misunderstanding if teams consist of players from multiple countries (see also Lavric et al., 2008).

---

[33]We found some interviews in which players and managers suggest that country-specific skills exist in soccer. For example, Nuri Sahin, a famous soccer player in Germany, mentioned in an interview with the football magazine *Kicker* that Japanese offensive players have special tactical skills (see https://www.kicker.de/732654/artikel).

[34]For evidence on the negative effect of linguistic diversity on team performance, see Dale-Olsen and Finseraas (2020).

Another potential explanation for why teams might benefit from birthplace diversity only up to a certain degree is that social categorization leads to performance-reducing inter-group biases when the team members are highly diverse (O'Reilly et al., 1989, van Knippenberg et al., 2004). Although we lack data that allows us to precisely assess the relevance of this mechanism, we think for two reasons that severe inter-group biases are relatively unlikely in our case. First, Price et al. (2013) provide evidence from the NBA, suggesting that racial differences do not produce inter-group biases in sports teams. Of course, ethnic diversity is not the same as birthplace diversity, but we think that inter-group biases are more prevalent in racially diverse working groups than in nationally diverse working groups. Second, some studies suggest that the risk of inter-group biases is relatively low if a team has a common goal or competes with another team (see e.g. Gaertner et al., 1993, 2000, Lowe, 2021).

To empirically support our explanations, we exploit the facts that the performance of football teams depends on two different aspects (goal scoring, goal preventing) and that interpersonal exchange on the field is more important for the defensive performance than for the offensive performance. Put differently, if our hunch about the channels at work is correct, we should find that the ideal level of birthplace diversity is larger for offensive than for defensive players.[35] Table D.12 suggests that this is indeed the case because we find that the optimal level of birthplace diversity for offensive performance ($B^* = 0.65$) exceeds the optimal level for defensive performance ($B^* = 0.52$). However, for two key reasons, we need to interpret this result with some caution. First, the two estimates of the optimal levels of birthplace diversity are only statistically significant at relatively low levels, according to the Lind-Mehlum-Test (p-val.: 0.171, 0.097). Second, this test does not allow to check whether the two predicted optimal levels of birthplace diversity are statistically different from each other. However, to substantiate that the importance of interpersonal exchange is likely to be a moderating factor for the relationship between birthplace diversity and team performance, we can argue that the difference of the two estimates is larger than one standard deviation.

5 Conclusion

Multinational working groups are becoming more and more common in most advanced economies. For many practitioners, it is thus important to know whether a nationally diverse team performs better than a homogeneous team. We study this question, using rich information on 7,208 matches and 3,266 players from the *Bundesliga*, the highest division of German male football. In our analysis, we apply two instrumental variable strategies to identify the causal effect of birthplace diversity on team performance. Our results suggest that this effect is hump-shaped. In other words, we find that birthplace diversity positively affects team performance only up to a certain level. Beyond this threshold, birthplace

---

[35]We also think that such a pattern could hardly be explained with inter-group biases since such an explanation needs that the defensive players are more likely to discriminate their co-workers than the offensive player. If anything, our anecdotal evidence suggests the opposite.

diversity reduces team performance.

An objection against our work might be that professional soccer is a rather special industry and that the external validity of our findings is thus low. We doubt that this is the case since effective exchange and diversity in skills are factors for success in various sectors. Examples include consultancy, arts and music, research and development, and marketing. Obviously, the optimal level of birthplace diversity differs within as well as between these working fields. In line with the view that this level is moderated by task-specific factors, we present suggestive evidence implying that the optimal structure of a team depends on how important interpersonal exchange is throughout the production process.

We believe that the contribution of our paper goes beyond the result that birthplace diversity has a hump-shaped effect on team performance. This is the case because we illustrate how and why our results considerably differ from the results reported in other studies with comparable settings. More specifically, we illustrate that the two-way fixed effect approach leads to (downward-)biased estimates of the optimal level of birthplace diversity. The key reason for this bias is that team compositions endogenously change during the production process. In our particular setting, we observe that the managers (unintentionally) increase the birthplace diversity of their team during a match if their team does not perform well. We show that neglecting this pattern largely changes the interpretation of the results. An important question in this regard is whether we must expect similar problems when studying other types of diversity or when using data that does not come from the sports industry. From our perspective, this question cannot be negated because endogenous changes in team compositions exist in various fields.[36] We therefore believe that conclusions should always be drawn with a lot of caution when detailed information on changes of team structures during the production process is not available.[37]

Our study suggests two paths for future research. First, we think that it is crucial to examine whether the fixed effect estimates reported in related studies can be confirmed when using a method that addresses the problem of reverse causality. Our paper helps to answer this question because our two instrumental variable approaches can easily be adjusted to different types of diversity and institutional environments. Second, we believe that future studies should provide more evidence on factors that moderate the effect of diversity and team performance. Without such evidence, we can hardly provide specific practical advices on how to structure a working group.

---

[36]For instance, young researchers may be more likely to ask a senior researcher to join their team if they have problems with their project. Similarly, managers may change the age or gender composition of a team depending on how it performs.

[37]Of course, if the problem of reverse causality exists, it depends on the specific research environment whether a fixed effect approach creates a downward-biased or an upward-biased estimate of the optimal level of diversity.

References

Addesa, F., Pazzona, M., and Rossi, G. (2022). Migrant diversity and team performance in a high-skilled labour market. *Kyklos*, 75(3):365–384.

Ager, P. and Brueckner, M. (2013). Cultural diversity and economic growth: Evidence from the US during the age of mass migration. *European Economic Review*, 64(1):76–97.

Alesina, A., Harnoss, J., and Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, 21(2):101–138.

Alesina, A. and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.

Alvarez, J., Forrest, D., Sanz, I., and Tena, J. d. D. (2011). Impact of importing foreign talent on performance levels of local co-workers. *Labour Economics*, 18(3):287–296.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30.

Arcidiacono, P., Kinsler, J., and Price, J. (2017). Productivity spillovers in team production: Evidence from professional basketball. *Journal of Labor Economics*, 35(1):191–225.

Ashraf, Q. and Galor, O. (2013). The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1):1–46.

Balsmeier, B., Frick, B., and Hickfang, M. (2019). The impact of skilled immigrants on their local teammates' performance. *Applied Economics Letters*, 26(2):97–103.

Basevitch, I., Yang, Y., and Tenenbaum, G. (2013). Is the best defense a good offense? Comparing the Brazilian and Italian soccer styles. *International Journal of Fundamental and Applied Kinesiology*, 45(2):213–221.

Bazzi, S. and Clemens, M. A. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics*, 5(2):152–86.

Bettinger, E. P. and Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review*, 95(2):152–157.

Bossert, W., d'Ambrosio, C., and La Ferrara, E. (2011). A generalized index of fractionalization. *Economica*, 78(312):723–750.

Bove, V. and Elia, L. (2017). Migration, diversity, and economic growth. *World Development*, 89(1):227–239.

Bryson, A. and Chevalier, A. (2015). Is there a taste for racial discrimination amongst employers? *Labour Economics*, 34(1):51–63.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2012). Robust inference with multiway clustering. *Journal of Business and Economic Statistics*, 29(2):238–249.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.

Cox, T. H., Lobel, S. A., and McLeod, P. L. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal*, 34(4):827–847.

Dale-Olsen, H. and Finseraas, H. (2020). Linguistic diversity and workplace productivity. *Labour Economics*, 64(1):101813.

Delis, M. D., Gaganis, C., Hasan, I., and Pasiouras, F. (2017). The effect of board directors from countries with different genetic diversity levels on corporate performance. *Management*

Science, 63(1):231–249.

Dellal, A., Chamari, K., Wong, d. P., Ahmaidi, S., Keller, D., Barros, R., Bisciotti, G. N., and Carling, C. (2011). Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European Journal of Sport Science*, 11(1):51–59.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, 132(4):1593–1640.

Dezsö, C. L. and Ross, D. G. (2012). Does female representation in top management improve firm performance? A panel data investigation. *Strategic Management Journal*, 33(9):1072–1089.

Dobson, S. and Goddard, J. (2011). *The Economics of Football.* Cambridge University Press.

Docquier, F., Turati, R., Valette, J., and Vasilakis, C. (2020). Birthplace diversity and economic growth: Evidence from the US states in the Post-World War II period. *Journal of Economic Geography*, 20(2):321–354.

Dohmen, T. and Sauermann, J. (2016). Referee bias. *Journal of Economic Surveys*, 30(4):679–695.

Dohmen, T. J. (2008a). Do professionals choke under pressure? *Journal of Economic Behavior & Organization*, 65(3-4):636–653.

Dohmen, T. J. (2008b). The influence of social forces: Evidence from the behavior of football referees. *Economic inquiry*, 46(3):411–424.

Earley, C. P. and Mosakowski, E. (2000). Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal*, 43(1):26–49.

Freeman, R. B. and Huang, W. (2015). Collaborating with people like me: Ethnic coauthorship within the United States. *Journal of Labor Economics*, 33(S1):S289–S318.

Frick, B. (2011). Gender differences in competitiveness: Empirical evidence from professional distance running. *Labour Economics*, 18(3):389–398.

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., and Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology*, 4(1):1–26.

Gaertner, S. L., Dovidio, J. F., Samuel, G., et al. (2000). *Reducing intergroup bias: The common ingroup identity model.* Psychology Press.

Garicano, L. and Palacios-Huerta, I. (2014). Making the beautiful game a bit less beautiful. In *Beautiful game theory: How soccer can help economics.* Princeton University Press.

Guillaume, Y. R., Dawson, J. F., Otaye-Ebede, L., Woods, S. A., and West, M. A. (2017). Harnessing demographic differences in organizations: What moderates the effects of workplace diversity? *Journal of Organizational Behavior*, 38(2):276–303.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 1(4):34–68.

Haas, H. and Nüesch, S. (2012). Are multinational teams more successful? *International Journal of Human Resource Management*, 23(15):3105–3113.

Herrmann, M. A. and Rockoff, J. E. (2012). Worker absence and productivity: Evidence from teaching. *Journal of Labor Economics*, 30(4):749–782.

Hoogendoorn, S., Oosterbeek, H., and Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, 59(7):1514–1528.

Hornung, E. (2014). Immigration and the diffusion of technology: The huguenot diaspora in prussia. *American Economic Review*, 104(1):84–122.

Horwitz, S. K. and Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, 33(6):987–1015.

30

Jones, B. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Review of Economic Studies*, 76(1):283–317.

Joshi, A. and Roh, H. (2009). The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal*, 52(3):599–627.

Kahane, L., Longley, N., and Simmons, R. (2013). The effects of coworker heterogeneity on firm-level output: Assessing the impacts of cultural and language diversity in the National Hockey League. *Review of Economics and Statistics*, 95(1):302–314.

Kellerman, E., Koonen, H., and van der Haagen, M. (2005). "feet speak louder than the tongue": A preliminary analysis of language provisions for foreign professional footballers in the Netherlands. In *Second Language Needs Analysis*, pages 200–222. Cambridge University Press.

Kesavan, S., Staats, B. R., and Gilland, W. (2014). Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Science*, 60(8):1884–1906.

Kleven, H. J., Landais, C., and Saez, E. (2013). Taxation and international migration of superstars: Evidence from the European football market. *American Economic Review*, 103(5):1892–1924.

Krumer, A. and Lechner, M. (2017). First in first win: Evidence on schedule effects in round-robin tournaments in mega-events. *European Economic Review*, 100:412–427.

Krumer, A. and Lechner, M. (2018). Midweek effect on soccer performance: Evidence from the German Bundesliga. *Economic Inquiry*, 56(1):193–207.

Lang, K. (1986). A language theory of discrimination. *Quarterly Journal of Economics*, 101(2):363–382.

Lavric, E., Pisek, G., Skinner, A., and Stadler, W. (2008). *The linguistics of football.* Narr Francke Attempto Verlag.

Lazear, E. P. (1999). Globalisation and the market for team-mates. *Economic Journal*, 109(454):15–40.

Lichter, A., Pestel, N., and Sommer, E. (2017). Productivity effects of air pollution: Evidence from professional soccer. *Labour Economics*, 48(1):54–66.

Lind, J. T. and Mehlum, H. (2010). With or without U? The appropriate test for a U-shaped relationship. *Oxford Bulletin of Economics and Statistics*, 72(1):109–118.

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review*, 111(6):1807–44.

Lyons, E. (2017). Team production in international labor markets: Experimental evidence from the field. *American Economic Journal: Applied Economics*, 9(3):70–104.

Maderer, D., Holtbrügge, D., and Schuster, T. (2014). Professional football squads as multicultural teams: Cultural diversity, intercultural experience, and team performance. *International Journal of Cross Cultural Management*, 14(2):215–238.

Mitrotasios, M., Gonzalez-Rodenas, J., Armatas, V., and Aranda, R. (2019). The creation of goal scoring opportunities in professional soccer. tactical differences between Spanish La Liga, English Premier League, German Bundesliga and Italian Serie A. *International Journal of Performance Analysis in Sport*, 19(3):452–465.

O'Reilly, C. A., Caldwell, D. F., and Barnett, W. P. (1989). Work group demography, social integration, and turnover. *Administrative Science Quarterly*, 34(1):21–37.

Ottaviano, G. I. and Peri, G. (2005). Cities and cultures. *Journal of Urban Economics*, 58(2):304–337.

Papps, K. L., Bryson, A., and Gomez, R. (2011). Heterogeneous worker ability and team-based production: Evidence from Major League Baseball, 1920–2009. *Labour Economics*, 18(3):310–319.

Prat, A. (2002). Should a team be homogeneous? *European Economic Review*, 46(7):1187–1207.

Price, J., Lefgren, L., and Tappen, H. (2013). Interracial workplace cooperation: Evidence from the NBA. *Economic Inquiry*, 51(1):1026–1034.

Price, J. and Wolfers, J. (2010). Racial discrimination among NBA referees. *Quarterly Journal of Economics*, 125(4):1859–1887.

Prinz, J. and Wicker, P. (2016). Diversity effects on team performance in the tour de france. *Team Performance Management*.

Ringbom, H. (2012). Multilingualism in a football team: the case of IFK Mariehamn. In *Cross-linguistic Influences in Multilingual Language Acquisition*, pages 185–197. Springer.

Sarmento, H., Anguera, M. T., Pereira, A., Marques, A., Campaniço, J., and Leitão, J. (2014). Patterns of play in the counterattack of elite football teams – a mixed method approach. *International Journal of Performance Analysis in Sport*, 14(2):411–427.

Sarmento, H., Pereira, A., Matos, N., Campaniço, J., Anguera, T. M., and Leitão, J. (2013). English Premier League, Spaińs La Liga and Italỳs Seriés A – What's different? *International Journal of Performance Analysis in Sport*, 13(3):773–789.

Schwab, A., Werbel, J. D., Hofmann, H., and Henriques, P. L. (2016). Managerial gender diversity and firm performance: An integration of different theoretical perspectives. *Group & Organization Management*, 41(1):5–31.

Simmons, R. (1997). Implications of the Bosman ruling for football transfer markets. *Economic Affairs*, 17(3):13–18.

Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of u-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science*, 1(4):538–555.

Spolaore, E. and Wacziarg, R. (2016). Ancestry, language and culture. In *Palgrave Handbook of Economics and Language*, pages 174–211. Springer.

Staats, B. R., Milkman, K. L., and Fox, C. R. (2012). The team scaling fallacy: Underestimating the declining efficiency of larger teams. *Organizational Behavior and Human Decision Processes*, 118(2):132–142.

Stahl-Rolf, S., Holtmannspötter, D., Hutapea, L., Mecks, E., Pfaff, D., von Proff, S., and Reuß, K. (2018). *Faktor Vielfalt – Die Rolle kultureller Vielfalt für Innovationen in Deutschland*. Bertelsmann Stiftung.

Suedekum, J., Wolf, K., and Blien, U. (2014). Cultural diversity and local labour markets. *Regional Studies*, 48(1):173–191.

Tan, T. F. and Netessine, S. (2014). When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science*, 60(6):1574–1593.

Tan, T. F. and Netessine, S. (2019). When you work with a superman, will you also fly? An empirical study of the impact of coworkers on performance. *Management Science*, 65(8):3495–3517.

van Knippenberg, D., De Dreu, C. K., and Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, 89(6):1008.

van Knippenberg, D. and Mell, J. N. (2016). Past, present, and potential future of team diversity research: From compositional diversity to emergent diversity. *Organizational Behavior and Human Decision Processes*, 136:135–145.

Williams, K. Y. and O'Reilly, C. A. I. (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior*, 20:77–140.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production

of knowledge. *Science*, 316(5827):1036–1039.

For online publication

A Theoretical framework

In this supplementary section, we develop a simple theoretical framework that illustrates how birthplace diversity might affect team performance. The structure of our model is greatly inspired by Ashraf and Galor (2013) and Lazear (1999).

We consider a working group that includes n people who originate from m countries. Each worker i has two traits: an origin-specific trait ($q_i \in [0,1]$) and a personality trait ($p_i \in [0,1]$) that does not depend on the country of origin. $\mathbf{A} = \{a_{ij}\}_{i,j=1,\ldots,n}$ denotes a similarity matrix in which $a_{ij} = 1 - |q_i - q_j| \in [0,1]$ reflects the extent to which the origin-specific traits of workers i and j resemble each other. As Bossert et al. (2011), we define that the birthplace diversity of a team ($\delta \in [0,1]$) only depends on the similarity matrix and assume that:

$$\delta = \delta(\mathbf{A}) = 0 \quad \Leftrightarrow \quad a_{ij} = 1 \quad \text{for all } i,j = 1, \ldots, n \tag{A1}$$

$$\delta = \delta(\mathbf{A}) = 1 \quad \Leftrightarrow \quad a_{ij} = 0 \quad \text{for all } i,j = 1, \ldots, n \text{ and } i \neq j \tag{A2}$$

$$\delta_{a_{i,j}} = \frac{\partial \, \delta(\mathbf{A})}{\partial \, a_{i,j}} < 0 \quad \text{for all } i,j = 1, \ldots, n \text{ with } a_{ij} \in (0,1). \tag{A3}$$

Our model presumes that the performance of teams depends on two factors. The first factor is the available knowledge:

$$H = H(\delta, \rho, \Sigma_Q, \Sigma_P) \geq 1 \quad \text{with} \quad \Sigma_Q = \sum_{i=1}^{n} q_i \quad \text{and} \quad \Sigma_P = \sum_{i=1}^{n} p_i \tag{A4}$$

where Σ_P and Σ_Q denote the abilities of the workers and $\rho \in [0,1]$ the diversity of the personality traits. We assume that birthplace diversity increases the available knowledge because individuals born in different countries grow up under different educational and cultural systems and are thus likely to have distinct productive skills (see Alesina and La Ferrara, 2005, Alesina et al., 2016, Ashraf and Galor, 2013):[38]

$$H_\delta = \frac{\partial \, H}{\partial \, \delta} > 0 \quad \text{with} \quad \lim_{\delta \to 0} H_\delta = \infty \quad \text{and} \quad \lim_{\delta \to 1} H_\delta = 0 \quad \text{and} \quad H_{\delta\delta} \leq 0. \tag{A5}$$

The second factor of team performance is the efficiency with which team members collaborate with each other:

$$E = E(\delta, \rho) \in (0,1] \quad \text{with} \quad E(0,0) = 1. \tag{A6}$$

The efficiency of the collaboration decreases in birthplace diversity since workers from different places are likely to face cultural and linguistic barriers (Freeman and Huang, 2015, Lazear, 1999). In addition, inter-group biases and conflicts are more likely to exist

---

[38] In line with several related studies, we implicitly assume that different skills complement each other.

in diverse teams (Earley and Mosakowski, 2000).

$$E_\delta \; = \; \frac{\partial E}{\partial \delta} \; < \; 0 \quad \text{with} \quad \lim_{\delta \to 0} E_\delta \; > \; -\infty \quad \text{and} \quad E_{\delta\delta} \; \leq \; 0. \tag{A7}$$

Assuming a Cobb-Douglas production function, the total output of the team is then given by:

$$Y \; = \; E\big(\delta, \, \rho\big)^\alpha \cdot H\big(\delta, \, \rho, \, \Sigma_Q, \, \Sigma_P\big)^{1-\alpha} \quad \text{with} \quad \alpha \in (0, 1), \tag{A8}$$

and the first-order condition with respect to δ can be written as:

$$0 \; = \; \alpha \cdot E_\delta \cdot \left(\frac{H\big(\delta, \rho, \Sigma_Q, \Sigma_P\big)}{E\big(\delta, \rho\big)} \right)^{1-\alpha} + \; (1 - \alpha) \cdot H_\delta \cdot \left(\frac{E\big(\delta, \rho\big)}{H\big(\delta, \rho, \Sigma_Q, \Sigma_P\big)} \right)^\alpha \tag{A9}$$

From (A9), we obtain the following results:

Proposition 1. *Conditional on all other determinants of team performance, our model suggests that:*

(a) Team output Y is a concave function of birthplace diversity δ.

(b) Team output Y is maximized at an intermediate level of birthplace diversity $\delta^* \in (0, 1)$.

(c) The more important efficient collaboration is for the team output, the lower is the optimal level of birthplace diversity:

$$\frac{\partial \delta^*}{\partial \alpha} \; < \; 0 \quad \text{with} \quad \lim_{\alpha \to 0} \delta^* \; = \; 1 \quad \text{and} \quad \lim_{\alpha \to 1} \delta^* \; = \; 0.$$

B The effect of performance-based substitutions on birthplace diversity

In this supplementary section, we illustrate why the conventional fixed effect approach produces downward-biased estimates of the optimal level of birthplace diversity. More specifically, we present regression results, showing that the level of birthplace diversity increases during a match because of performance-based substitutions.

Our analysis consists of three parts The first part shows how team managers react during a match if their team does not perform well. To this end, we divide each match into eighteen non-overlapping five-minute periods and estimate the regression model:

$$\Delta \, O_{isrmdp} \; = \; \alpha \cdot O^b_{isrmdp} \; + \; \beta \cdot P^b_{isrmdp} \; + \; \gamma \cdot \mathbf{X}^b_{isrmdp} \; + \; \xi_{isrmd} \; + \; \theta_p \; + \; \varepsilon_{isrmdp} \tag{B1}$$

with

$$\Delta \, O_{isrmdp} \; = \; O^e_{isrmdp} \; - \; O^b_{isrmdp} \tag{B2}$$

Table B.1 Changes in the team composition during a match.

	(1)	(2)	(3)
Behind (begin of period)	0.074***	0.001***	-0.000
	(0.000)	(0.004)	(0.456)
Forwards (end of period)			0.016***
			(0.000)
Outcome variable	Δ Forwards	Δ Birthplace Div.	Δ Birthplace Div.
Observations	143,208	143,208	143,208
Seasons	13	13	13
Fixed effects	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes
Quality controls	Yes	Yes	Yes

Notes: The table reports results from within-match fixed effect regressions. We cluster standard errors at the club-by-match level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

where i denotes a club, s a season, m a manager, d a match day, r a round, and p a five-minute period. O^b (O^e) is the number of offensive players at the beginning (end) of a period. P is a dummy that is equal to 1 only if a team is behind at the beginning of a period. $\mathbf{X}$ includes team and opponent controls. ξ is a club-by-match fixed effect, θ a period fixed effect, and ε the error term.

Column 1 of Table B.1 shows results from estimating (B1). Consistent with Garicano and Palacios-Huerta (2014), we find that team managers increase the number of fielded forwards if their team is behind.

In the second part of this supplementary analysis, we study whether the birthplace diversity of the fielded players depends on the score of a match. To this end, we use the regression model:

$$\Delta \, \mathrm{B}_{isrmdp} = \alpha \cdot \mathrm{B}^b_{isrmdp} + \beta \cdot \mathrm{P}^b_{isrmdp} + \gamma \cdot \mathbf{X}^b_{isrmdp} + \xi_{isrmd} + \theta_p + \varepsilon_{isrmdp} \quad \text{(B3)}$$

with

$$\Delta \, \mathrm{B}_{isrmdp} = \mathrm{B}^e_{isrmdp} - \mathrm{B}^b_{isrmdp} \quad \text{(B4)}$$

where B^b (B^e) denotes the birthplace diversity at the beginning (end) of a five-minute period. All other components have the same meaning as above.

In Column 2 of Table B.1, we report results from estimating (B1). We observe that the level of birthplace diversity increases if a team is behind. This result implies that reverse causality is a severe problem in our setting.

The final part of our analysis provides an explanation for why birthplace diversity increase if teams do not perform well. Two simple aspects play a role here. First, team managers dislike losses and thus replace defensive with offensive players if their team is behind (see Column 1 of Table B.1). Second, the share of foreigners in the *Bundesliga* is much larger among offensive players than among defensive players.[39] Put differently,

---

[39]Our raw data suggests that the share of foreign defenders and midfielders is around 45 percent, whereas the share of foreign forwards is around 65 percent.

we argue that the increase in the number of forwards is the channel through which team performance affects birthplace diversity. The results shown in Column 3 of Table B.1 support this view since we find that the estimate of the effect of team performance on birthplace diversity is statistically insignificant if we augment regression model (B3) by the number of offensive players at the end of a five-minute period.

C Additional figures

Figure C.1 Market values and performance measures.

(a) Minutes played.   **(b)** Goals scored.   **(c)** Assists made.
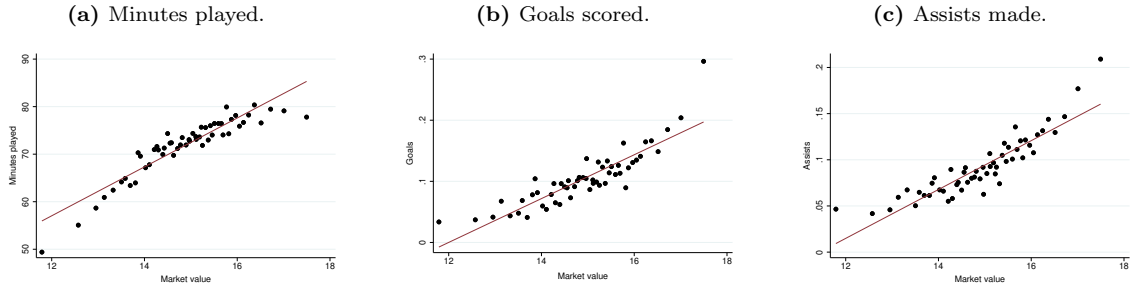
Notes: This figure shows the correlation between the market value and several performance measures at the individual level. All correlations are adjusted for positional fixed effects.

Figure C.2 FIFA rating and performance measures.

(a) Minutes played.   **(b)** Goals scored.   **(c)** Assists made.
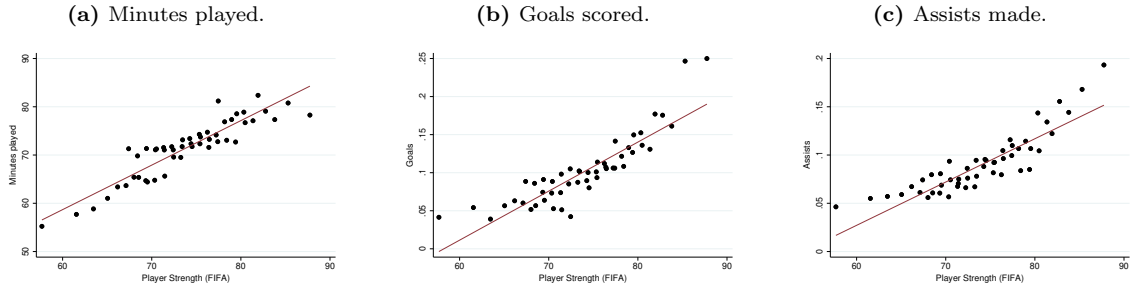
Notes: This figure shows the correlation between the FIFA playing strength and several performance measures at the individual level. All correlations are adjusted for positional fixed effects.

Figure C.3 Classification of players (example).

<div align="center">

Predicted Line-up   **Starting Line-up**

Predicted Line-up	Starting Line-up
Ulreich	Ulreich
Kimmich Süle Hummels Alaba	Kimmich Süle Hummels Rafinha
Martinez Thiago James	Tolisso Thiago James
Müller Lewandowski Ribery	Müller Lewandowski Ribery

</div>

Notes: The left panel shows the expected starting line-up of *Bayern München* for the match on Saturday, 12th May 2018. *Kicker* published this line-up on Friday, 11th May 2018. The right panel shows the actual starting line-up. We use *blue* characters to indicate players (Alaba, Martinez) that belong to group $\mathcal{A}$, *red* characters to indicate players (Rafinha, Tolisso) that belong to group $\mathcal{B}$, and *black* characters to indicate players that belong to group $\mathcal{C}$.

Figure C.4 Classification of players (example).

<div align="center">

| Predicted Line-up	Starting Line-up
Ulreich	Ulreich
Kimmich Süle Hummels Alaba	Kimmich Süle Hummels Rafinha
Martinez Thiago James	Tolisso Thiago James
Müller Lewandowski Ribery	Müller Lewandowski Ribery

</div>

Notes: The left panel shows the expected starting line-up of *Bayern München* for the match on Saturday, 12th May 2018. *Kicker* published this line-up on Friday, 11th May 2018. The right panel shows the actual starting line-up. We use *blue* characters to indicate players (Alaba) that belong to group $\widetilde{\mathcal{A}}$, *red* characters to indicate players (Rafinha) that belong to group $\widetilde{\mathcal{B}}$, and *black* characters to indicate players that belong to group $\widetilde{\mathcal{C}}$. Players (Martinez) colored in *violet* belong to neither of the three groups. David Alaba is a member of group $\widetilde{\mathcal{A}}$ because he unexpectedly missed the match because of back problems.

Figure C.5 Zero-stage relationship (all unexpected changes).
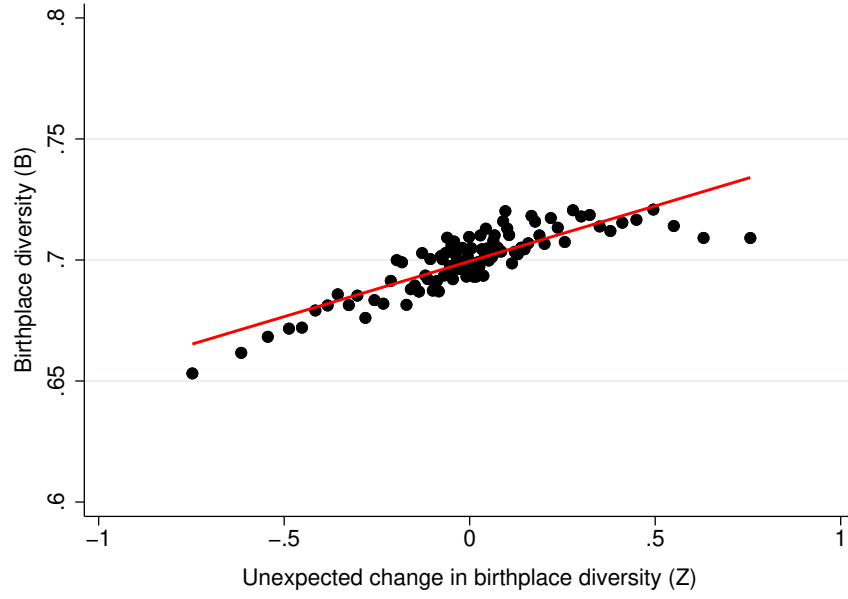
Notes: This figure depicts the positive correlation between the birthplace diversity of the fielded players (B) and our instrumental variable Z as defined in Equation (9). For the sake of vividness, we use a binned scatterplot to present the underlying raw data.

Figure C.6 Zero-stage relationship (injuries).
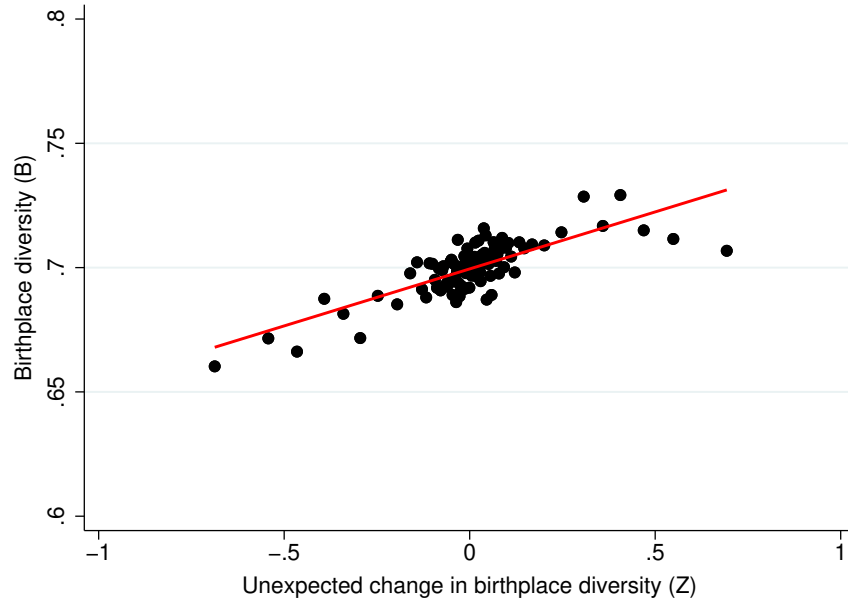
Notes: This figure depicts the positive correlation between the birthplace diversity of the fielded players (B) and our instrumental variable Z as defined in Equation (15). For the sake of vividness, we use a binned scatterplot to present the underlying raw data.

Figure C.7 Leave-one-out analysis (seasons).
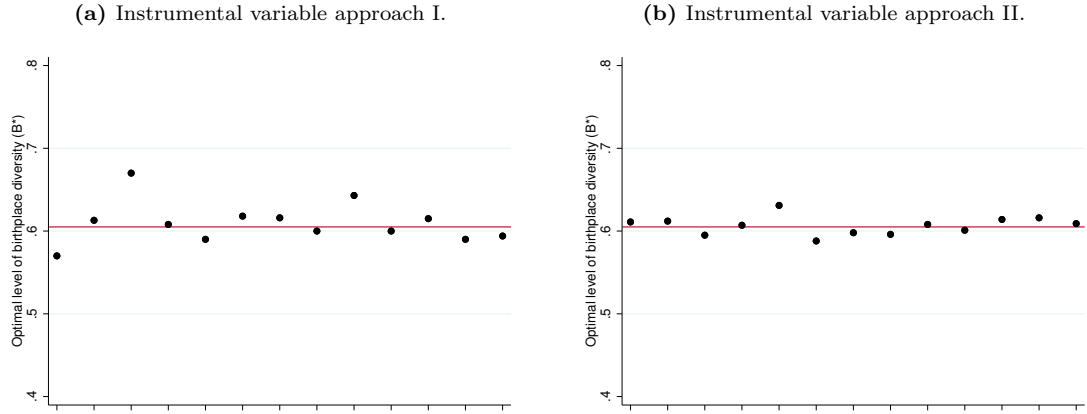
(a) Instrumental variable approach I.   **(b)** Instrumental variable approach II.

Notes: This figure shows the optimal level of birthplace diversity (B^*) estimated with our instrumental variable strategies for different estimation samples. In each sample, we drop one season from our data set. The red line represents the estimate that we obtain when using the full sample.

Figure C.8 Leave-one-out analysis (clubs).
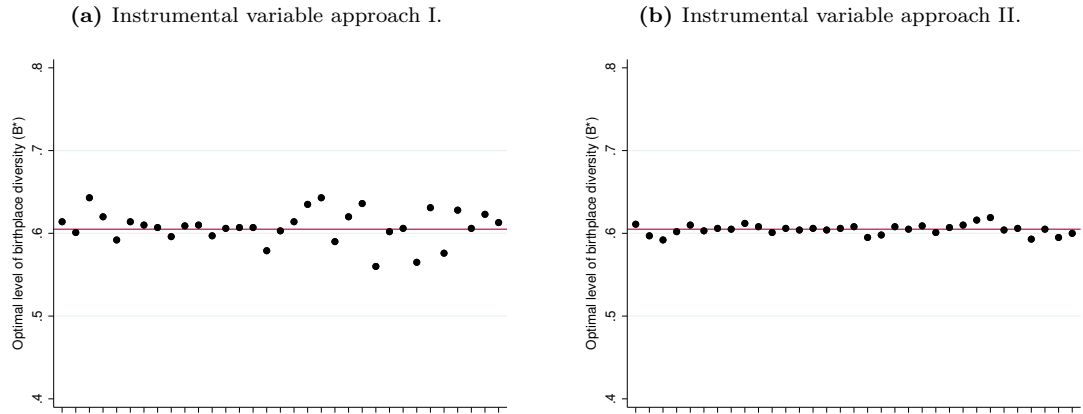
(a) Instrumental variable approach I.   **(b)** Instrumental variable approach II.

Notes: This figure shows the optimal level of birthplace diversity (B^*) estimated with our instrumental variable strategies for different estimation samples. In each sample, we drop one club from our data set. The red line represents the estimate that we obtain when using the full sample.

Figure C.9 Deviation from B* and seasonal performance.
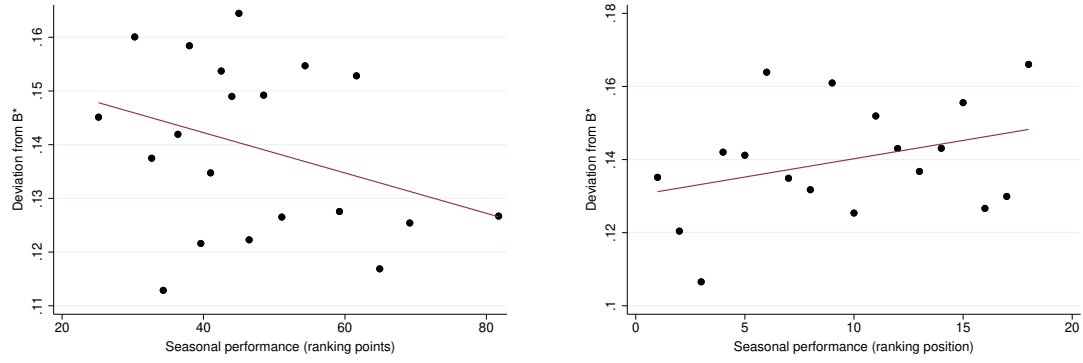
Notes: This figure illustrates the deviation from the optimal level of birthplace diversity (B*) against the final performance of the team in a season measured by (a) the achieved number of points and (b) the final ranking position.

Figure C.10 Deviation from B* by team quality.
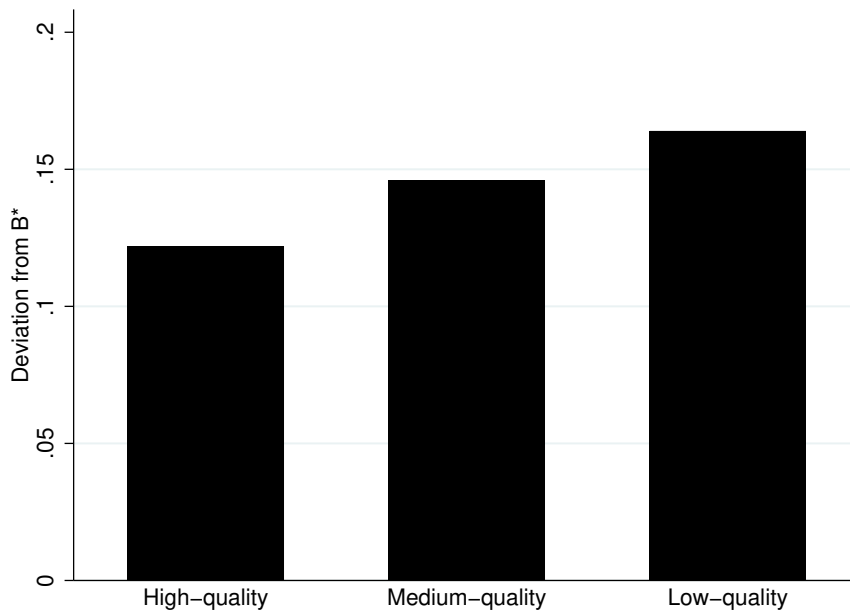
Notes: This figure illustrates the deviation from the optimal level of birthplace diversity (B*) grouped by team quality. Team quality is calculated as the average yearly performance between the season 2005/06 and the season 2017/2018. Table D.13 indicates which clubs belong to which category.

D Additional tables

Table D.1 List of countries.

Country	Country	Country
Albania (13)	Gambia (3)	Norway (36)
Algeria (1)	Georgia (9)	Paraguay (5)
Argentina (36)	Germany (1534)	Peru (6)
Armenia (1)	Ghana (26)	Poland (89)
Australia (16)	Greece (19)	Portugal (16)
Austria (71)	Guinea (3)	Romania (28)
Azerbaijan (1)	Guinea-Bissau (2)	Russia (16)
Belarus (3)	Hungary (27)	Senegal (11)
Belgium (38)	Iceland (6)	Serbia (47)
Benin (2)	Iran (10)	Sierra Leone (3)
Bolivia (1)	Israel (5)	Slovakia (23)
Bosnia and Herzegovina (35)	Italy (13)	Slovenia (14)
Brazil (141)	Ivory Coast (12)	South Africa (11)
Bulgaria (24)	Jamaica (2)	Spain (38)
Burkina Faso (3)	Japan (28)	Suriname (4)
Cameroon (25)	Kazakhstan (5)	Sweden (39)
Canada (9)	Korea (11)	Switzerland (79)
Cape Verde (1)	Kosovo (17)	Syria (1)
Chile (8)	Latvia (2)	Tajikistan (3)
China (3)	Lebanon (2)	Togo (4)
Colombia (8)	Liechtenstein (1)	Trinidad and Tobago (1)
Congo Demo. Rep. (12)	Lithuania (1)	Tunisia (10)
Cong Rep. (4)	Luxembourg (2)	Turkey (17)
Costa Rica (1)	Macedonia (16)	Ukraine (20)
Croatia (71)	Mali (2)	United Kingdom (15)
Cyprus (1)	Malta (1)	United States (32)
Czech Republic (70)	Mexico (8)	Uruguay (7)
Denmark (75)	Moldova (2)	Venezuela (5)
Ecuador (3)	Montenegro (9)	Zambia (3)
Egypt (7)	Morocco (8)	Zimbabwe (1)
Equatorial Guinea (3)	Mozambique (1)	
Estonia (1)	Namibia (2)	
Finland (12)	Netherlands (72)	
France (92)	Nigeria (26)	

Notes: This table lists the countries from which the players in our data set originate. In parentheses, we report the number of players born in a particular country.

Table D.2 List of control variables.

Variable	Description
Home	Dummy variable that is equal to 1 if a club had the home field advantage.
Cup[a]	Dummy variable that is equal to 1 if a team played in the national cup in the week before (or the week after) a *Bundesliga* match.
ECT[a]	Dummy variable that is equal to 1 if a team played in an European club tournament in the week before (or the week after) a *Bundesliga* match.
Age[a,b]	Average age (in years).
Tenure[a,b]	Average duration of club membership (in years).
Bundesliga matches[a]	Average number of *Bundesliga* matches.
European matches[a]	Average number of matches in the European club tournaments.
Championship matches[a]	Average number of matches in European and World Championship matches.
Top-League matches[a]	Average number of matches in highest soccer divisions in England, France, Italy, and Spain.
Market value[a,c]	Average of logged market values.
FIFA rating[a,c]	Average playing strength in video game *FIFA*.
Age (manager)[a]	Age of the manager.
Tenure (manager)[a]	Number of years in which the manager was in charge of the team.
Bundesliga matches (manager)[a]	Total number of *Bundesliga* matches that a manager was in charge of a *Bundesliga* team.
Opponent-Season fixed effects	Set of season-specific dummy variables that indicate the opponent of the team.
Ranking fixed effects	Set of dummy variables that capture the current ranking positions of the opposing teams.

Notes: This table lists all variables that are included in the vector $\mathbf{X}$. (a) indicates that $\mathbf{X}$ includes this variable for both the team and its opponent. (b) indicates that $\mathbf{X}$ includes the squared term of the variable. (c) indicates that $\mathbf{X}$ does not include this variable if we use the full sample.

Table D.3 Summary statistics (full sample).

Variable	Mean	Std. Dev.	Min	Max
Goal difference	0	1.872	-8	8
Ranking points	1.372	1.313	0	3
Main explanatory variable				
Birthplace diversity	0.669	0.153	0	0.918
Control variables				
Home	0.500	0.500	0	1
Cup	0.101	0.301	0	1
ECT	0.168	0.374	0	1
Age	26.824	1.332	22.321	31.029
Tenure	2.247	0.779	0.321	6.25
Bundesliga matches	86.826	34.924	0.5	223.857
European matches	12.710	10.117	0	72.429
Championship matches	1.228	1.629	0	15.286
Top-league matches	8.492	11.073	0	90.923
Age (manager)	47.330	6.684	28.6	73.7
Tenure (manager)	2.007	2.348	0	14
Bundesliga matches (manager)	131.599	140.297	0	829

Notes: This table presents summary statistics for our reduced sample. This sample consists of 14,076 observations. In Table D.15, we indicate how strongly correlated the team characteristics are.

Table D.4 Summary statistics (reduced sample).

Variable	Mean	Std. Dev.	Min	Max
Dependent variables				
Goal difference	0	1.872	-8	8
Ranking points	1.373	1.315	0	3
Main explanatory variable				
Birthplace diversity	0.699	0.1354	0	0.918
Control variables				
Home	0.5	0.5	0	1
Cup	0.101	0.301	0	1
ECT	0.186	0.389	0	1
Age	26.261	1.237	22.321	30.6
Tenure	2.117	0.705	0.321	5.357
Bundesliga matches	81.529	30.862	6.071	193
European matches	14.010	10.873	0	72.429
Championship matches	1.276	1.840	0	15.286
Top-league matches	9.397	11.849	0	80.923
Market value (log)	14.869	0.721	13.007	17.288
FIFA rating	73.956	3.501	63.071	86.714
Age (manager)	47.660	6.964	28.6	73.7
Tenure (manager)	1.719	2.099	0	14
Bundesliga matches (manager)	129.759	140.863	0	829

Notes: This table presents summary statistics for our reduced sample. This sample consists of 7,956 observations. In Table D.16, we indicate how strongly correlated the team characteristics are.

Table D.5 First-stage estimates (non-linear model).

	(1)	(2)
	First inst. var. approach	
Birthplace divers. in Kicker line-up	0.364***	0.035
	(0.000)	(0.517)
Birthplace divers. in Kicker line-up sq.	-0.032	0.308***
	(0.528)	(0.000)
	Second inst. var. approach	
Predicted level of birthplace divers.	0.996***	1.020***
	(0.000)	(0.000)
Predicted level of birthplace divers. sq.	0.008	0.137
	(0.930)	(0.207)
Outcome variable	B	B^2
Observations	7,956	7,956
Seasons	13	13
Fixed effects	Yes	Yes
Basic controls	Yes	Yes
Quality controls	Yes	Yes

Notes: This table presents first-stage estimates. For the related second-stage estimates, see Columns 3 and 4 of Table 1. Standard errors are clustered at the club-by-season-by-round-by-manager level and p-values are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table D.6 Reduced-form estimates (non-linear model).

	(1)	(2)
	Lind-Mehlum-Test	
Optimal level of birthplace diversity	0.588*	0.723**
	(0.068)	(0.052)
	Regression coefficients	
Birthplace divers. in Kicker line-up	1.803	
	(0.103)	
Birthplace divers. in Kicker line-up sq.	-1.533*	
	(0.092)	
Predicted level of birthplace divers.		6.375**
		(0.046)
Predicted level of birthplace divers. sq.		-4.408**
		(0.026)
Outcome variable	Goal Diff.	Goal Diff.
Observations	7,956	7,956
Seasons	13	13
Fixed effects	Yes	Yes
Basic controls	Yes	Yes
Quality controls	Yes	Yes
Sample mean (B)	0.699	0.699

Notes: This table reports reduced-form estimates. For the related second-stage estimates, see Columns 3 and 4 of Table 1. Standard errors are clustered at the club-by-season-by-round-by-manager level and p-values are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table D.7 Robustness checks (instrumental variable estimates, non-linear model, part 1).

	(1)	(2)	(3)	(4)	(5)	(6)
	Lind-Mehlum-Test					
Optimal level of birthplace	0.631**	0.549**	0.566**	0.595*	0.575**	0.593**
diversity ($\hat{B}^*$)	(0.049)	(0.044)	(0.031)	(0.063)	(0.044)	(0.020)
	Regression coefficients					
Birthplace divers. ($\hat{\beta}_1$)	41.991*	4.434*	33.855*	4.233	41.922*	45.081**
	(0.098)	(0.089)	(0.061)	(0.105)	(0.088)	(0.040)
Birthplace divers. sq. ($\hat{\beta}_2$)	-33.289*	-4.038*	-29.875*	-3.557*	-36.465*	-37.979**
	(0.093)	(0.056)	(0.057)	(0.093)	(0.086)	(0.035)
Estimation technique	IV	IV	IV	IV	IV	IV
IV Approach	2nd	1st	2nd	1st	2nd	2nd
Outcome variable	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Observations	7,956	7,956	7,956	7,956	7,956	7,956
Seasons	13	13	13	13	13	13
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes	Yes	Yes
Quality controls	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic (B)	9.8	165.8	9.4	140.90	5.76	12.9
First-stage F-statistic (B^2)	10.2	337.9	9.8	312.33	5.96	13.3
Sample mean (B):	0.699	0.684	0.684	0.680	0.680	0.699

Notes: The table reports results from six instrumental variable regressions. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. In Column 1, we use injuries as source of plausible exogenous variation. In Columns 2 and 3, we only take into account those players who participated in a match for at least 30 minutes. In column 4 and 5 we weight players according to the number of minutes they played. In Column 6, we use an alternative procedure for measuring the extent to which birthplace diversity changes due to an unexpected replacement in the starting line-up.

Table D.8 Robustness checks (instrumental variable estimates, non-linear model, part 2).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					Lind-Mehlum-Test			
Optimal level of birthplace diversity ($\hat{B}^*$)	0.576**	0.602**	0.657*	0.537***	0.566**	0.569**	0.593*	0.609**
	(0.018)	(0.029)	(0.085)	(0.038)	(0.028)	(0.040)	(0.066)	(0.022)
					Regression coefficients			
Birthplace divers. ($\hat{\beta}_1$)	8.163**	40.010*	7.092**	24.838*	7.041**	52.442*	5.265	43.108**
	(0.035)	(0.058)	(0.013)	(0.076)	(0.039)	(0.073)	(0.116)	(0.039)
Birthplace divers. sq. ($\hat{\beta}_2$)	-7.082**	-33.244*	-5.397**	-23.112*	-6.216**	-46.104*	-4.437*	-35.419***
	(0.022)	(0.055)	(0.031)	(0.063)	(0.034)	(0.075)	(0.098)	(0.039)
Estimation technique	IV	IV	IV	IV	IV	IV	IV	IV
IV Approach	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Outcome variable	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Observations	7,488	7,488	7,956	7,956	7,956	7,956	7,956	7,956
Seasons	13	13	13	13	13	13	13	13
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quality controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic (B)	89.6	11.7	132.4	14.9	153.7	7.4	126.6	13.7
First-stage F-statistic (B²)	215.7	12.6	265.3	15.8	303.4	7.60	286.7	14.4
Sample mean (B):	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699

Notes: The table reports results from eight instrumental variable regressions. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively. In Column 1 and 2, we control for the performance of the club in the last two matches. In Columns 3 and 4, we take into account that some players migrated during their childhood. In Columns 5 and 6, we use an alternative measure of birthplace diversity. In Columns 7 and 8, we control for form fluctuations.

A14

Table D.9 Birthplace diversity and team performance (non-linear model, alternative outcome).

	(1)	(2)	(3)	(4)
Panel A: Linear regression model				
Regression coefficients				
Birthplace divers. $(\hat{\beta}_1)$	-0.815***	-0.913***	-0.164	-0.575
	(0.000)	(0.001)	(0.798)	(0.603)
First-stage F-statistic	-	-	640.5	272.80
Panel B: Non-linear regression model				
Lind-Mehlum-Test				
Optimal level of birthplace	0.236	0.269	0.578*	0.586*
diversity $(\hat{B}^*)$	(0.222)	(0.252)	(0.062)	(0.072)
Regression coefficients				
Birthplace divers. $(\hat{\beta}_1)$	0.558	0.725	3.799	19.162
	(0.444)	(0.504)	(0.124)	(0.143)
Birthplace divers. sq. $(\hat{\beta}_2)$	-1.183**	-1.347	-3.284*	-16.351
	(0.049)	(0.118)	(0.100)	(0.132)
First-stage F-statistic (B)	-	-	126.3	13.7
First-stage F-statistic (B^2)	-	-	286.5	14.3
Estimation technique	OLS	OLS	IV	IV
IV approach	-	-	1st	2nd
Outcome variable	Points	Points	Points	Points
Observations	14,076	7,956	7,956	7,956
Seasons	23	13	13	13
Fixed effects	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes
Quality controls	No	Yes	Yes	Yes
Sample mean (B):	0.669	0.699	0.699	0.699

Notes: The table reports results from fixed effect and instrumental variable regressions. We cluster standard errors at the club-by-season-by-round-by-manager level and show p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table D.10 Alternative clustering (instrumental variable estimates, non-linear model).

	(1)	(2)	(3)	(4)	(5)	(6)
	Lind-Mehlum-Test					
Optimal level of birthplace diversity ($\hat{B}^*$)	0.608*	0.604**	0.608*	0.604**	0.608*	0.604*
	(0.071)	(0.032)	(0.083)	(0.042)	(0.081)	(0.052)
	Regression coefficients					
Birthplace divers. ($\hat{\beta}_1$)	5.373*	41.924*	5.373*	41.924*	5.373**	41.924
	(0.068)	(0.065)	(0.069)	(0.083)	(0.036)	(0.105)
Birthplace divers. sq. ($\hat{\beta}_2$)	-4.420*	-34.670*	-4.420*	-34.670*	-4.420*	-34.670*
	(0.070)	(0.062)	(0.080)	(0.077)	(0.056)	(0.098)
Estimation technique	IV	IV	IV	IV	IV	IV
IV approach	1st	2nd	1st	2nd	1st	2nd
Outcome variable	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Observations	7,956	7,956	7,956	7,956	7,956	7,956
Seasons	13	13	13	13	13	13
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes	Yes	Yes
Quality controls	Yes	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic (B)	136.9	9.91	127.8	13.6	137.2	9.9
First-stage F-statistic (B^2)	307.5	10.61	222.5	14.4	241.3	10.7
Sample mean (B):	0.669	0.699	0.699	0.699	0.699	0.699

Notes: The table reports results from instrumental variable regressions. In Columns 1 and 2, we cluster standard errors at the club-by-season level. In Columns 3 and 4, we apply a two-way clustering procedure (club-by-season-by-round-by-manager and match day level). In Columns 3 and 4, we cluster at the club-by-season and the match day level. We report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table D.11 Piecewise and interrupted linear regression (instrumental variable estimates).

	(1)	(2)	(3)	(4)	(5)
	Piecewise linear regression				
Birthplace diversity	3.735**	-3.073*	-17.755*	18.519***	
	(0.042)	(0.072)	(0.093)	(0.003)	
	Interrupted regression (Simonsohn, 2018)				
Birthplace diversity (low)					3.328*
					(0.069)
Birthplace diversity (high)					-3.006*
					(0.060)
Estimation technique	IV	IV	IV	IV	IV
IV approach	1st	1st	2nd	2nd	1st
Observations	2,083	5,784	1,633	1,217	7,956
Seasons	13	13	13	13	13
Outcome	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.	Goal Diff.
Fixed effects	Yes	Yes	Yes	Yes	Yes
Basic controls	Yes	Yes	Yes	Yes	Yes
Quality controls	Yes	Yes	Yes	Yes	Yes
First-stage F-statistic	89.8	370.1	10.2	30.1	345.41
Diversity level ($BDiv$)	< 0.6	> 0.6	> 0.6	> 0.6	-
Change in Diversity (Z)	-	-	> 0.0	< 0.0	-

Notes: The table reports results from instrumental variable regressions. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table D.12 Heterogeneity in optimal level of birthplace diversity.

	Performance in offense	Performance in defense
	(1)	(2)
	Lind-Mehlum-Test	
Optimal level of birthplace diversity ($\hat{B}^*$)	0.649	0.516*
	(0.171)	(0.097)
	Regression coefficients	
Birthplace divers. ($\hat{\beta}_1$)	2.778*	-3.037
	(0.062)	(0.194)
Birthplace divers. sq. ($\hat{\beta}_2$)	-2.139	2.943
	(0.114)	(0.130)
Estimation technique	IV	IV
IV approach	1st	1st
Observations	7,956	7,956
Seasons	13	13
Outcome variable	Goals scored	Goals allowed
Fixed effects	Yes	Yes
Basic controls	Yes	Yes
Quality controls	Yes	Yes
First-stage F-statistic (B)	148.98	185.85
First-stage F-statistic (B^2)	284.78	359.96
Team	Midfielder + Forwards	Defender + Midfielder

Notes: The table reports results from instrumental variable regressions. We cluster standard errors at the club-by-season-by-round-by-manager level and report p-values in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table D.13 Classification of clubs.

High-quality	Medium-quality	Low-quality
Bayern München	1. FC Kaiserlautern	Alemannia Aachen
Bayer 04 Leverkusen	1. FC Köln	Eintracht Braunschweig
Borussia Dortmund	1. FSV Mainz 05	FC St. Pauli
Schalke 04	Arminia Bielefeld	Fortuna Düsseldorf
RB Leipzig	Borussia Mönchengladbach	SC Paderborn
Werder Bremen	Eintracht Frankfurt	SV Darmstadt 98
	Energie Cottbus	SpVgg Greuther Fürth
	FC Augsburg	
	FC Ingolstadt	
	Hamburger SV	
	Hannover 96	
	Hansa Rostock	
	Hertha BSC Berlin	
	Karlsruher SC	
	MSV Duisburg	
	SC Freiburg	
	TSG 1899 Hoffenheim	
	VfB Stuttgart	
	VfL Bochum	
	VfL Wolfsburg	

Notes: This table lists how clubs are classified in high-, medium, and low-quality clubs. The classification is based on the overall performance of the clubs in the sample period.

Table D.14 Newspaper articles about the importance of communication in soccer.

Interviewed person	Interviewer	Sources
Alassane Pléa (player)	Express (newspaper)	https://www.express.de/sport/fussball/borussia-moenchengladbach/gladbach-stuermer-p1%C3% A9a-spricht-nach-rekord-transfer-ueber-favre--dante-und-seine-nummer-30957612
Max Meyer (player)	Sport1 (TV-channel)	https://www.sport1.de/internationaler-fussball/premier-league/2020/02/ max-meyer-exklusiv-so-lief-das-mit-schalke-04-sein-leben-in-london
Uli Hoeneß (manager)	Sport-Bild (newspaper)	https://sportbild.bild.de/bundesliga/vereine/bundesliga/ hoeness-fordert-deutsch-pflicht-in-bayern-kabine-49570212.sport.html
Rene Paasch (psychologist)	Blog entry	https://www.die-sportpsychologen.de/2017/01/dr-rene-paasch-sprachbarriere-im-fussball/
Franck Ribery (player)	FC Bayern (online-interview)	https://fcbayern.com/de/news/2017/07/interview-mit-franck-ribery-teil-1/
Ciro Immobile (player)	Focus (newspaper)	https://www.focus.de/sport/fussball/bundesliga1/ verlorenes-jahr-in-dortmund-groesster-heil-profi-der-bundesliga-immobile-klagt-ueber-bvb-klopp-und-deutschlnad_ id_4873285.html/
Giovane Elber (player)	Hamburger Morgenpost (newspaper)	https://www.mopo.de/sport/hsv/ santos-und-walace-aussen-vor-giovane-elber-erklaert-die-brasilianer-krise-beim-hsv-26873000/
Arne Friedrich (player)	Zeit (newspaper)	/https://www.zeit.de/sport/2011-10/arne-friedrich-sprachen-lernen-deutsch-loew/
Birgit Kirschke (teacher)	Zeit (newspaper)	https://www.faz.net/aktuell/sport/fussball/bundesliga/ im-interview-die-deutschlehrerin-der-auslaendischen-cottbus-spieler-1464254.html/
Pep Guardiola (manager)	Zeit (newspaper)	https://www.welt.de/sport/fussball/article182738950/ Sprachprobleme-Guardiola-wollte-bei-Bayern-nach-zwei-Monaten-hinschmeissen.html/
Robert Peters (reporter)	Rheinische Post (newspaper)	https: //rp-online.de/sport/fussball/international/spanien/sprache-ist-der-schluessel_aid-12594963/
Joachim Born (linguist)	scientific work	http://geb.uni-giessen.de/geb/volltexte/2015/11823/pdf/LU_8_Born_Gloning.pdf/
Ellis Palmer (reporter)	online interview	https://www.bbc.com/news/world-europe-44624066/
Anonymous (reporter)	tweet	https://translation-blog.trustedtranslations.com/ language-sports-communication-european-football-2017-06-21.html/
J. Steiner und E. Lavric (linguist)	scientific work	https://www.uibk.ac.at/romanistik/personal/lavric/sprache_fussball/publications/b115_merspr_i_ fussball_fallstudie_span_legionaer.pdf/
RB Leipzig	homestory	https://rblive.de/ueber-rb-leipzig/ dolmetscherin-bei-rb-leipzig-hat-hausaufgaben-fuer-keita-burke-und-bernardo-1523/

Table D.15 Correlations between team characteristics (full sample).

	Birthplace diversity	Home	Cup	ECT	Age	Tenure	Bundesliga Matches	European Matches	Championship matches	Top-league matches
Birthplace diversity	1.000									
Home	0.010	1.000								
Cup	0.014	0.001	1.000							
ECT	0.026	0.003	0.107	1.000						
Age	-0.019	0.011	-0.002	-0.072	1.000					
Tenure	-0.147	0.008	-0.002	0.140	0.247	1.000				
Bundesliga matches	-0.058	0.005	0.042	0.282	0.315	0.591	1.000			
European matches	0.185	0.006	0.074	0.369	0.069	0.389	0.682	1.000		
Championship matches	0.102	0.005	0.071	0.329	0.135	0.323	0.584	0.807	1.00	
Top-league matches	0.175	0.005	0.063	0.233	0.064	0.094	0.318	0.572	0.605	1.000

Table D.16 Correlations between team characteristics (reduced sample).

	Birthplace diversity	Home	Cup	ECT	Age	Tenure	Bundesliga Matches	European Matches	Championship matches	Top-league matches	Market value (log)	FIFA rating
Birthplace diversity	1.000											
Home	0.008	1.000										
Cup	0.018	-0.001	1.000									
ECT	0.017	0.003	0.096	1.000								
Age	0.207	0.008	-0.009	-0.063	1.000							
Tenure	-0.050	0.011	0.020	0.172	0.225	1.000						
Bundesliga matches	0.090	0.005	0.065	0.306	0.252	0.628	1.000					
European matches	0.190	0.006	0.085	0.383	0.166	0.480	0.748	1.000				
Championship matches	0.144	0.003	0.082	0.347	0.180	0.402	0.581	0.830	1.00			
Top-league matches	0.240	0.002	0.062	0.236	0.156	0.134	0.310	0.570	0.600	1.000		
Market value (log)	0.124	0.005	0.103	0.460	-0.167	0.427	0.705	0.783	0.668	0.456	1.000	
FIFA rating	0.148	0.007	0.099	0.443	-0.024	0.457	0.763	0.790	0.697	0.472	0.938	1.000