

NATIONALIST AND HOME FIELD BIAS IN SPORT PERFORMANCE EVALUATIONS

An example from the ski jumping World Cup

High stakes and prizes in international sports competitions necessitate efficiently identifying legitimate winners. Illegitimate biases of non-independent judges and home field advantages represent two major threats to the legitimacy of international competitions. Judges' nationalities represent one source of bias, as they may prefer athletes of their own nation. Judges may also favour athletes competing within their home countries, creating another potential bias. Such biases, conscious or unconscious, may affect the outcomes of competitions. The literature offers no complete consensus on the magnitude of such biases, hence not whether they affect competition outcomes or not. In this paper, we shed light on these problems, using international ski jumping as an example. We draw on data from the FIS World Cup competitions in the 2006-2008 and 2015-2016 seasons, and employ several fixed-effects based regression designs. Our results reaffirm suspicions of national bias and home field advantage in major ski jumping competitions, but also show their magnitude is too little to be of major relevance to competition outcomes. There is a trend towards smaller nationalist biases

Keywords: nationalist bias, home field bias, skijumping, judges

INTRODUCTION

Many sports use qualified experts to rank and evaluate performances. For example, ski jumping competitions have a panel of five judges who independently evaluate each performance. Similar systems are used in diving, figure skating, and numerous other sports. The experts may completely or partially decide the outcome of competitions.

A prerogative of the system is that experts judge performances in an unbiased and neutral manner. However, there is a large potential for biases in such judgements (Meyer and Booker 1991). Experts may, for example, overestimate performances by competitors from their own country, overvalue certain stylistic elements, or unconsciously adjust evaluations in response to spectator expectations. Such biases may then be consequential in determining the outcome of competitions. Expert judgements are used well beyond the sports domain, and the relevance of studying expert judgements may inform several important societal domains. Expert judgements are widely used, for example, when evaluating applicants to universities, evaluating research grant proposals (Marsh et al. 2008), and in assessing job applicants (cf. Fasang (2006); Bursell (2012)). Understanding the extent of and processes underlying such biases is important for any attempts to remove their effects.

In the present paper, we set out to study two possible types of bias in evaluations of ski jumping performances: nationalistic bias and home-field advantage bias. Nationalistic bias refers to biases in which the judges (positively) evaluate contenders from their own country, whereas home-field advantage bias arises if judges favour contenders from the country in which the contest is held. To answer our questions, we use data from two periods, the 2006-2008 seasons and the 2015-2016 season, of the Federation International de Ski (FIS) World Cup series of ski jumping contests.

We contribute to the literature on nationalistic and home field advantage biases in several ways. First, we analyse nationalistic bias and assess to what extent such bias can be explained by unobserved characteristics of the performance. Second, we consider variations in such bias, and discuss this variation in light of sociological theory of evaluation and valuation. Third, by using our data from two time periods and comparing our results to those obtained with similar data but from the

markedly different context of the Olympic games, a highest-stakes-contest (Zitzewitz 2006), we can shed light on the role of different institutional and incentive contexts and changes over time in the magnitude of bias. Fourth, we study the social and physical ramifications of home advantages.

THEORY AND RELEVANT LITERATURE

There is an extant scholarly literature that explores the existence and magnitude of subjective biases in sports evaluations. These biases may affect results or other outcomes in favor of what is in principle irrelevant aspects of competitors or other actors in the sport. Research have documented biases that promotes profits for the league owner or organization, racial and ethnic biases, nationalistic biases, as well as home advantage biases.

Research on biases in basketball illustrates the breadth of the field. In their study of American professional basketball, Price, Remer and Stone (J. Price, Remer, and Stone 2012) found that the referees were biased towards calls that would increase profits for the league. More specifically, in turnovers, the referees tend to favour the home team, the losing team in a particular game, and the losing teams of the play-off series. In turn these biases can lead to increased ticket sales and ad-revenue due to more exciting, closer games; and ultimately even extend the playoff series itself, garnering more revenue. Even though these biases may operate towards profit-maximizing goals, the authors note that they need not result from any conscious behaviour on behalf of the referees.

Two related papers explore a potential racial bias among NBA referees, and its consequences for betting markets (J. Price and Wolfers 2010, Larsen, Price, and Wolfers (2008)). The authors note that there is a significant difference between fouls called when the athlete and referees are of the same race versus when they are not. The difference is large enough to be an important factor in determining game outcomes, making it a viable tactic to pursue in betting markets. Analysing data from two seasons, they find that when the majority of judges are white, betting on the team that has more minutes (expected) played by white players will beat the point spread over 50% of the time.

Nationalistic or cultural bias, where judges favour athletes from their home country or culturally similar ones in a systematic way would be most evident in certain nationalistic competitions such as

world cups or Olympics. Evidence of these kinds of biases have been found in many different sports, including rugby (L. Page and Page 2010) , muay thai (Myers et al. 2006), figure skating (Yang 2006), gymnastics (Callahan, Mulholland, and Rotthoff 2016), and surfing (Sampaio 2012).

In their study of the 2000 Olympic diving competition, Emerson, Seltzer and Lin (Emerson, Seltzer, and Lin 2009) found evidence of strong biases in some of the judges; both nationalistic ones, and against competitors from certain other nations. Based on their model, they re-estimated the results, removing the biases to obtain the objective quality of the dive, and isolated a case where the medals might have changed hands – had the judges been completely unbiased. Zitzewitz (2006) made a similar analysis of datasets from the 2002 Salt Lake City Winter Olympics and found evidence of nationalistic biases in both ski jumping and figure skating. Evidence of bloc voting and vote trading was also found in figure skating where Germany, USA, Canada and Italy are more likely to be negatively biased towards Russia, Ukraine, France and Poland, and vice versa (Zitzewitz 2006). Bloc voting in figure skating is nothing new, however. Seltzer and Glass (1991), for example, documented consistent bloc voting along cold-war lines in addition to nationalistic biases in the Winter Olympic games held from 1968 till 1988.

Vote trading in the 2002 Olympics led to a revamp of the scoring system for figure skating after a French judge admitted to being pressured to vote in favour of the Russian pair over the Canadian pair in the pairs' competition final. This effectively handed the Russian pair the gold medal. Allegedly this was part of a deal with the Russian team, where they would reciprocate in a later competition in favour of the French. This scandal was examined quantitatively by Lock and Lock (2003) using a bootstrap technique to look for inconsistencies in the scoring. Looking at the correlation between judges' rank (of the athlete) and the final actual rank (a high correlation indicating an accurate judge), they isolated one inconsistent judge, but it was not the French one. The French judge had the best correlation of all the judges in that event; in other words, she was the most spot-on in her scoring, and no evidence of wrongdoing was found.

Theorizing nationalistic biases in ski jumping

Our reading of the empirical literature on nationalistic and other evaluative biases suggests there is evidence for such biases in a variety of sports. Given its existence, what are the root causes of such bias and how do they emerge? Theoretically, we can conceive of bias in at least three different ways. We denote these three mechanisms the social psychological mechanism, the cultural legitimacy mechanism and the differential professionalization mechanism, and discuss each of them in turn below.

The social psychological mechanism: Implicit or explicit nationalist bias

The first social psychological mechanism is the simple folk psychology theory of judges' nationalistic impulses at the subconscious or even conscious level. They prefer their own to win, and give higher scores to those athletes. Importantly, this theoretical mechanism does not by itself predict variation in the magnitude of biases across countries. It could in principle be combined with other theoretical ideas to arrive at predictions about any such variation.

Theory of evaluation work, as applied to nationalist biases in ski jumping

The second way of theorizing bias is to consider the role of judges and their efforts as evaluation work. Evaluative work and the justification for evaluations as a sociological study has been profoundly influenced by the works of Boltanski & Thevenon (???). A central tenet in their sociology of evaluations is the room for evaluation.

The room for evaluation decides what is relevant and good in a field. The evaluative room may be wide, and allow for many different evaluative statements, or narrow, and restrict evaluative statements. The main point is that the evaluative room allows for criticisms by field actors that in turn may affect the structure of the room for evaluation. Criticisms within the room for evaluation are recognized as forces of change in organizations and society at large. Actors cannot expect to have any criticism heard, and it must be recognized by multiple actors before it is heard. If enough actors repeat and (re)formulate a criticism, it will potentially lead to change and a new room for evaluation.

Applied to ski jumping as a field, judges' practices for scoring represents the room for evaluation. Within this room there are likely coexisting practices that may be correlated with judges' nationalities, giving rise to national variations in judge scores. Some countries, for example, Norway and Finland, enjoy long ski jumping traditions, whereas neighbouring Sweden, an otherwise similar country, does not. Most likely the long traditions also imply that judges from these countries to a larger degree than Swedish judges will define the room for evaluation. This became salient when Jan Boklöv, a Swede, pioneered the V-style in ski jumping. His style was not *comme-il-faut* among judges and competitors in ski jumping at the time, and thus heavily penalized by judges. Eventually he won competitions, due to the technical advantages of the V-style, and other actors started to mimic the style (Müller 2008). These changes then led to a change in the room for evaluations, where V-style was accepted and it is currently completely dominant in the field.

Mere national variation in judge scoring practices does not yield nationalistic biases in individual scores. However, when we also consider the fact that competitors are trained and socialized in their respective national ski jumping associations and competitions. This implies that young competitors share the more experienced judges' evaluative standards, and in turn this leads to nationalistic biases in judge scoring.

Professionalisation of sports and the evolution of evaluation

The third source of national variation in ski jumping evaluations is related to the process of professionalization in sports. An essential ingredient in evaluation are standards. In sports, standards are continuously evolving. The continued professionalization of sports could suggest a higher degree of standardization and a narrowing of the room for evaluations. Several aspects of contemporary ski jumping suggests that professionalization will limit the magnitude of nationalistic biases.

First, FIS licences judges that takes international, high-level judging assignments. Professionalization of ski jumping entails strict quality control of judge performances and training. Currently FIS requires national skiing associations to hold at least one national-level judge seminar every year to contribute to educating new judges, and all judges are required to participate in such seminars at

least once every two years.

Second, There is also a system in place for auditing judges. Chapter 10 of the FIS Rules documents explains that judges are subject to an evaluation by a FIS Sub-committee working together with a data team, and that their work may result in allotting individual judges with “points” that indicate non-fair scoring practices. If judges accumulate many points, they may not be nominated to higher-level competitions or wholly excluded from judging. Such sanctioned judges may re-qualify for judging assignments after scoring performances from television broadcasts.

Third, the competition rules also limit the influence of such biases by removing the best and worst of the five scores from the total score. This certainly limits the magnitude of any bias, but as previous research have found significant biases (???), it does not seem to root it completely out.

The relevance of professionalism for the *national variation* in biases emerges when we combine the idea of evaluative work with the likely different levels of professionalism among the national associations that are responsible for judge training. In countries where ski jumping is a relatively important sport, such as Finland and Norway, the bias should be less due to an assumed stronger professionalization.

Summing up, it seems meaningful to probe the existence of nationalist biases in different countries, but also study whether nationalist biased have decreased in magnitude over time, due to the general tendency towards increased professionalization in sports.

In addition to the social psychological, cultural legitimacy, and differential professionalization mechanisms discussed above, there may well be yet other mechanisms that also contribute to nationalistic biases of this kind, for example desires on part of individual judges to grow the sport's influence in their countries and that try to boost their own competitors placements.

The extant literature on home field advantages

Home field advantage is another type of potential bias in ski jumping. Such advantages, where some competitors benefit from advantages from competing on the home field, may arise in many different ways. Several of the mechanisms that may produce home field advantages are psychological sources

of superior performances, while others involve judges overrating home field competitors. This point illustrates that the mere presence of a biases is not necessarily illegitimate. In order to improve scoring and fairness in sports, one must understand the mechanisms behind them.

Some research into the home-field advantage has attempted to move our understanding of the causal mechanisms behind the advantages. Carmichael and Thomas (2005) outline the four most common explanations of this phenomena using soccer as an example. The first is a familiarity explanation – the home team is more familiar with the field, the stadium, the locker room, and so forth, which puts the players more at ease in the situation, allowing them to concentrate on performing at their best. Second, traveling to another field could result in fatigue, invoke various stressors, and cause disruptions in pre-competition routines for the away team. All these factors may lead to inferior performances by the away team, relative to their likely performance at the home field. Third, specific rules may favour the home team, giving them the upper hand. Last, there could be a crowd effect. The size of the crowd at the home field might improve individual or team performance. It might also make the judge to be more lenient for the home team, through psychological mechanisms.

Page and Page (2010) attributes some of the crowd effect to crowd noise, noting that the judge is under both time pressure (making a correct decision in a short time) and social pressure (from the crowd). A study by Nevill et al. (2002) showed that crowd noise was related to the home-field advantage. Setting up an experiment with one group of judges watching a football match with sound, and another group without, they found that the group watching with sound called 15.5% fewer fouls against the home team than the group who watched in silence. This could also indicate that judges use crowd noise as a guideline in their judging work. The crowd effect may also take another direction, as Anders and Rotthoff (2014) find that football clubs with notoriously violent fans have a larger home-field advantage than peaceful ones. Here, the theory is that the prospect of fans-to-judge violence causes a more lenient judge.

Research on the home-field advantage has also been done in various winter sports. Balmer et al. (2001) analysed home-field advantages the winter Olympics from 1908 to 1998, and found some

evidence for it in figure skating, freestyle skiing, ski jumping, alpine skiing and short track speed skating. The authors examined the relevance of three of the explanations listed above (familiarity with local conditions, time zones and crowd effect), and found evidence for home condition familiarity. Conditional on the time zones travelled, travel time as such did not matter. Most evidence, however, was found for the crowd effect theory, as sports that were judged only subjectively had a bigger home-field advantage than sports in which assessments were based on both subjective and objective criteria.

Hypotheses

We propose several hypotheses on biases in international ski jumping competitions, based on the literature and theory reviewed above. On the basis of earlier findings, we propose a first hypothesis that there are nationalistic biases in ski jumping. Specifically, we expect there to be a positive bias in scores when judges score performances by athletes who are from the same nation as themselves.

Our second hypothesis states that the gross bias is partly explained by controlling for characteristics of the jump and athlete. The empirical implication of is that the magnitude of the bias will attenuate once all unobserved factors related to the jump are controlled.

A third hypothesis concerns variation in bias across countries. Nationalistic biases may vary across countries due to cultural-professional variations, where the strength of the bias is lower in countries with strong jumping cultures and higher in countries where the sport is more peripheral in the larger sports field.

Our fourth hypothesis states that nationalistic biases in ski jumping evaluations decrease over time, possibly as a results of continued professionalization of the sport.

For home field advantages, we propose an additional hypothesis that home field advantages are measurable and impacts athletes performances. Finally, we distinguish between the social aspects of competing on the home field and the physical aspects of competing on the home field, and hypothesize that the physical home field advantage is less salient than the social home field advantage, and thus less important for performances.

DATA AND METHODS

To test the hypotheses proposed above, we use data from the FIS World Cup competitions in ski jumping from the 2006-2007 and 2007-2008 seasons as well as from the 2015-2016 seasons. The original data for 2006-2007 and 2007-2008 seasons were downloaded from the FIS web site, coded, edited for errors, and collated into complete data set for analysis. For the season 2015-2016, we wrote automated web scraping software that identified relevant PDF files on the FIS servers, downloaded these files, and scraped them according to our pre-specified instructions. The software used to download and scrape the PDF files are documented in the GitHub repository associated with this article.

Both data sets include detailed information on each jump in each competition, including the five individual judge scores from the jump along with the identity and nationality of each judge in each competition. Characteristics of the jump such as speed and length are included, as well as the points allotted. The total points for the jump is the sum of the length score and the stylistic score. We also know where and when the competition was held. For the later period, additional details on wind, starting gate and point adjustments for such conditions are also available.

We exclude team events from both seasons, as team events follow a different structure than individual competitions and also because extracting data from team events are markedly more difficult than it is for individual events. Table 1 outlines the dimensionality of the two data sets.

Table 1: Descriptive statistics

Statistic	2005-2007	2015-2016
	Mean / Count	Mean / Count
No. of competitor nations	26	15
No. of unique skijumping hills	22	18
No. of unique competitors	154	123
No. of unique judges	95	104

Statistic	2005-2007	2015-2016
No. of jumps	3772	1017
No. of competitions	51	43
Avg. judge score	17.39	17.63
Avg. jump length	125.48	125.71
Avg. points	115.51	126.07

Our analysis proceeds in several steps. First, we estimate nationalistic biases using a sequence of regression models where more complexity is incorporated for each model, each model speaking to one of the first three hypotheses. We estimate the models separately for the two periods we have data for, as there may have been change over time in the nature of nationalistic biases, the subject of our fourth hypothesis. In a second step, we estimate home field advantages in both jump performance and stylistic evaluation.

Estimation of nationalistic biases

Following the general approach used in most studies of nationalistic biases (in particular, Emerson, Seltzer, and Lin (2009); Zitzewitz (2006)), we analyse systematic tendencies in scores by individual judges and by the judges' nationality. We begin with a simple model and add complexity in subsequent steps to approach a stricter test of nationalistic bias. In the first, simple, model, a dummy variable $\phi(I = J)$ indicates whether the judge comes from the same country (J) as the athlete (I) or not. The model can be written as

$$s_{ijp} = B \cdot \phi(I = J) + \epsilon_{ijp} \quad (1)$$

where s_{ijp} is the score given by judge j , on performance p , by a certain athlete i , and ϵ_{ijp} is the error term. The B coefficient in this model estimates the difference in expected score if the athlete and the judge come from the same country.

However, such a model does not take into account relevant differences between nations. A Norwegian judge may give higher scores to a fellow countryman's performance due to a correspondence of their definitions of valuable stylistic elements, in accordance with national variations in evaluation practices.

Our next step is an attempt to remove such confounding factors, by exploiting the other judges' evaluations of the same jump in estimation of nationalistic bias. Assuming that the other four judges give an unbiased score, a more precise way to estimate the bias is to add a fixed effects-term for each jump to the model (in effect including a dummy variable for each jump). In such a model, our estimator is based on individual judges' score deviances from the average score for each jump. If the average score represents a more objective assessment of the performance, deviations from it can be interpreted as signs of bias. Our second model is therefore

$$s_{ijp} = B \cdot \phi(I = J) + q_p + \epsilon_{ijp} \quad (2)$$

where q_p is the jump fixed effect. This model will not confuse differences in stylistic culture with nationalistic biases, comparing the potentially biased judge against the four other judges for each jump. Following Zitzewitz (2006), we also estimated a model with so-called compensating biases, relaxing the assumption of objectivity with the four remaining judges. This model allows for the remaining judges to compensate a potentially biased judge, by lowering their scores when they know that one of the others will up his score. We did not find any significant compensating bias in our data, but this particular model did not converge and we thus abstain from reporting it in our set of results.

A limitation of the second model is that it does not capture any inherent leniency the judges may display. If a judge consistently scores performances higher or lower than other judges, then we cannot be sure that the observed bias is not just a particular judge always scoring lower or higher than the average judge. To control for this, we add a judge fixed effect, l_j , which leads to our third model:

$$s_{ijp} = B \cdot \phi(I = J) + q_p + l_j + \epsilon_{ijp} \quad (3)$$

This model is an improvement over the previous ones, but limitations still remain. The model assumes that bias is constant for each judge across all judges from all nations. Ideally, we would want to allow for variations in bias between individual judges or groups of judges by interacting the judge fixed-effect with the indicator function for judge and competitor nation correspondence. This would in principle let us obtain individual bias estimates for each judge, but our data unfortunately do not allow for such a specific estimation. They do however allow for estimating an interaction between nationalistic bias and the home country of the judge. This leads to the model:

$$s_{ijp} = B \cdot (L_j \cdot \phi(I = J)) + q_p + \epsilon_{ijp} \quad (4)$$

where L_j is an added judge country fixed-effect. In this fourth model, the estimates for B provide an assessment of the average bias in scores by judges from different countries.

Estimation of home field advantage

To analyse home field advantage we need to separate the problem of bias into the physical and the social domains. That is, whether an athlete actually performs better on the home field or whether he gets better scores by the judges due to social pressure or other factors. Thanks to the natural separation of style scores and length scores in ski jumping, this can be uncovered using similar models.

In order to observe a home field effect in points, be it length points or the judges' style points, we have to compare scores on a home field with scores on away fields – leading to the assumption that a jumper is somewhat consistent in performance throughout a season. Previously, we used the four other judges as a contrast measure of performance quality. For home field advantage we have to assume that an athlete maintains a relatively constant quality of jumping throughout the season, using his away jumps as a guide to his general performance level. This can be a bold assumption,

given that local weather conditions and individual athlete’s “daily form” also may affect these scores, as well as the tiny margins of timing required to “push off” at the right moment. Another problem is that countries with a strong tradition for ski jumping will have both stronger jumpers, host competitions more often due to existing infrastructure, and have more judges in the judge panel.

Again, we construct a dummy-variable using a function, $(I = H)$, which will take the value 1 if the jump takes place in country I, the athlete’s home country, and H being the competition country. We then wish to estimate the following model:

$$s'_{irs} = B' \cdot (I = H) + a_i + r + \epsilon_{ir} \quad (5)$$

where s_{irs} are the length points of the athlete, in a specific season, in a specific competition. B' is the estimated home field effect in points and a_i is a fixed effect for the athlete-season combinations. The fixed effect also removes any effect of individual jumpers that “strong jumpers” from traditionally strong ski jumping nations by out estimating each athlete, in each season, separately. Lastly, taking into account the different sizes of the hills, which may result in a skewness that the K-point system fails to adjust for, we add a fixed effect for the hill represented by r .

For the social part of the home field advantage we wish to estimate a similar model where style points awarded by the judges serves as the dependent variable.

$$s_{ijprs} = B \cdot (I = H) + a_{is} + r + l_j + x_p + \epsilon_{ijprs} \quad (6)$$

In this model, we add fixed effects for judges as in (3) to capture leniency of individual judges. We cannot use jump fixed-effects to capture the inherent objective quality of the jump, because home advantage is invariant within a jump, but we can include two central jump characteristics in our data, namely length and speed. This is because “all else being equal, a ski jumper with the same take-off speed but better form while airborne will travel further” (Zitzewitz 2006, 78). We therefore add these two variables to the model, where they and their coefficients are represented by the x_p .

In accordance with norms of open social science, we offer our data and analysis log in an open

repository at the time of publication.

RESULTS AND DISCUSSION

Nationalistic biases in judge scores

Our analysis of nationalistic bias proceeds in four steps. Table 2 presents the results from the first three estimations.

Table 2: Results from regression models of national bias

	Model 1a: Naïve	Model 2a: Within-jump	Model 3a: Within-judge
National bias	0.146*** (0.088 0.203)	0.068*** (0.050 0.087)	0.081*** (0.062 0.099)
Constant	17.378*** (17.361 17.395)		
Observations	18,860	18,860	18,860
R ²	0.001	0.925	0.931
Adjusted R ²	0.001	0.906	0.914
Residual Std. Error	1.153 (df = 18858)	0.354 (df = 15087)	0.339 (df = 14994)
F Statistic	24.981*** (df = 1; 18858)		

Note:

*p<0.05; **p<0.01; ***p<0.001

In models 2a and 3a, the intercept is suppressed by the fixed effects estimation procedure.

Table 3: Results from regression models of national bias for season 2015-2016

	Model 1b: Naïve	Model 2b: Within-jump	Model 3b: Within-judge
National bias	0.101* (0.021 0.181)	0.053** (0.018 0.087)	0.056** (0.023 0.089)
Constant	17.624*** (17.600 17.649)		
Observations	5,085	5,085	5,085
R ²	0.001	0.870	0.891
Adjusted R ²	0.001	0.837	0.860
Residual Std. Error	0.848 (df = 5083)	0.342 (df = 4067)	0.317 (df = 3967)
F Statistic	6.062* (df = 1; 5083)		

Note:

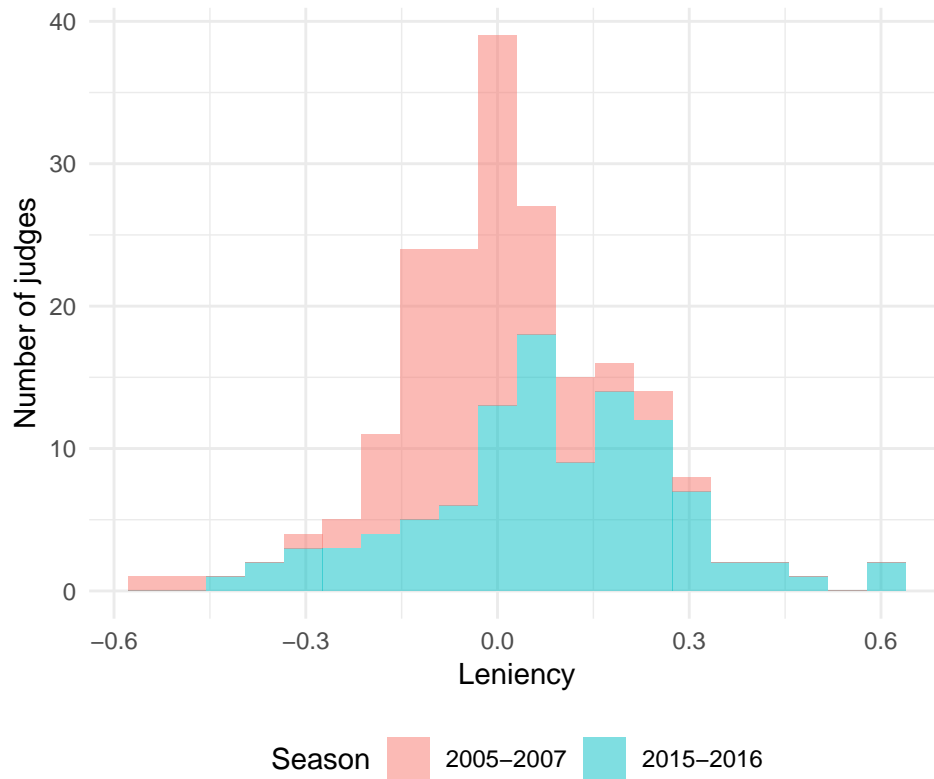
*p<0.05; **p<0.01; ***p<0.001

In models 2b and 3b, the intercept is suppressed by the fixed effects estimation procedure.

Our first model yields a positive and statistically significant estimate: judges award an average of 0.146 extra points to jumpers from their home country. This is approximately one-third of the minimum increment in judge scores in ski jumping (0.5). Compared with within-jump variance it only amounts to a quarter of a standard deviation. This result already suggests that the magnitude of nationalistic bias in judges' scores is limited. The estimate may be biased by cultural and historical differences between countries that lead to an over- or underappreciation of certain styles. The second model takes up exactly this problem, and provides a within-jump estimate of the nationalistic bias. In this model, the coefficient is estimated at 0.068, markedly lower than in the first model. It is statistically significant at the 0.1% level. The magnitude is low, but there is clearly a tendency towards nationalistic bias.

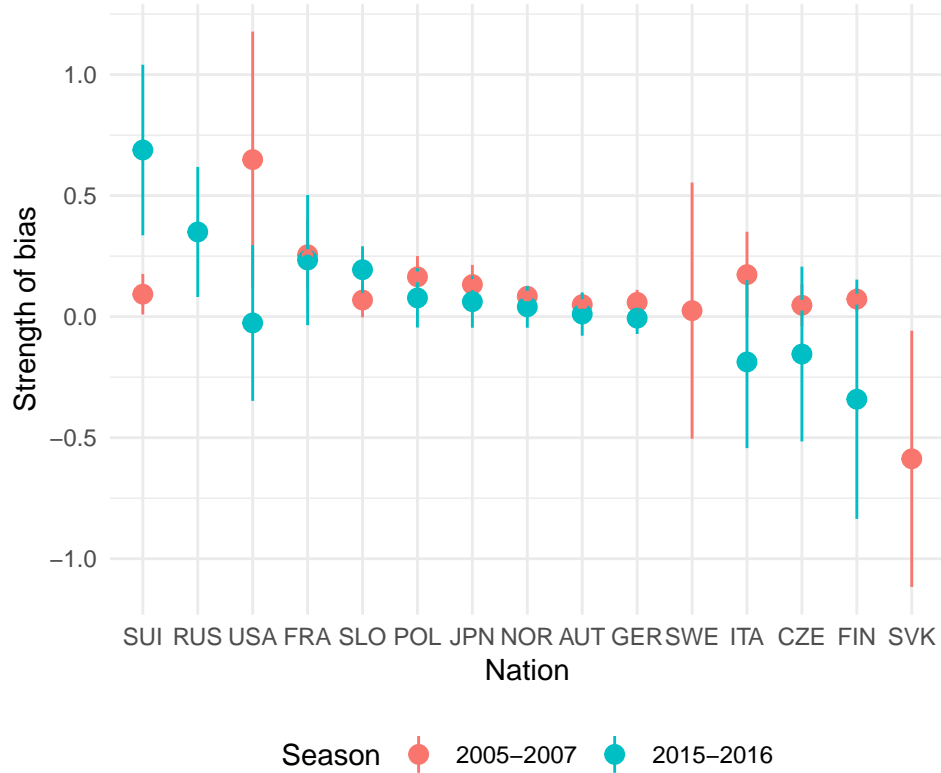
The third model let us also include judge leniency, a measure of how strict a judge is in his or her evaluations. From this model, we obtain a similar estimate of nationalistic bias of 0.081. All estimates are well below 0.5, the minimum increment in judge scores used in ski jumping. Judges' leniency display important variation, however, as the range of judges leniency coefficients is larger than 0.8 points. Figure 1 show this variation in a histogram.

Figure 1. Histogram of judges' leniency



The fourth and final model allowed estimating the variation in nationalistic biases for judges from different countries. The results are shown in Figure 2. Countries with fewer than 25 observations are not shown, as their estimates have large standard errors and thus are imprecisely estimated. French judges appear most biased, followed by Polish, and Japanese judges. For the other countries, the estimated biases are fairly small and in several cases not statistically significant. It is difficult to provide a substantive interpretation of this variation in national biases, but it is nevertheless clear that the biases are relatively small in most cases.

Figure 2. Nation-specific estimates of judges' bias



Home field advantages

Table 3 presents the results for the two models, where we included effects of the hypothesized home field advantage. Following the distinction between the physical and social dimensions of such advantages, we specified one model for the actual length of the jump and another for style scores. The first column of Table 3 show the parameter estimates for the length model (Model 5). This model includes competitor-season-fixed effects, which implies that variations among different athletes and their form is taken into account and the estimate for differential jump lengths for home field competitors are purged of those influences. Our main interest lies with the dummy variable for the athlete performing in his home country. The parameter estimate for this dummy variable is not significantly different from zero. Athletes thus do not seem to perform better in their home countries, than when competing in other countries. The physical dimension of home field advantage is thus undetectable in our data.

Table 4: Results from regression models of home field advantage

	Model 5: Length points	Model 6: Style points
Home field speed length	0.772 (−0.675 2.218)	0.074*** (0.036 0.112) −0.050*** (−0.064 −0.036) 0.067*** (0.065 0.068)
Observations	3,772	18,650
R ²	0.789	0.539
Adjusted R ²	0.773	0.530
Residual Std. Error	13.555 (df = 3504)	0.785 (df = 18286)

Note: *p<0.05; **p<0.01; ***p<0.001
Both models include athlete and jumping hill fixed effects, and model 6 also include judge fixed-effects.

We examine the social dimension of home field advantage in Model 6, where we include a parameter for bias in judge's scores if the athlete is competing in his or her home country. If there is home field advantage, this parameter should be measurably larger than zero. The estimate of this parameter is statistically significant, and larger than zero, suggesting that there is a social dimension of home field advantage in ski jumping. The effect is small and roughly of the same magnitude (0.074) as for nationalistic bias. The bias only amounts to a quarter of a standard deviation within each jump and it is well below the minimum increment for judge scores of 0.5 point. It seems unlikely that home field advantages is a serious threat to the validity of results of any competition.

Our results in light of the extant literature

Our bias estimates were smaller, however, than for example those reported from the Salt Lake City 2002 Winter Olympics (Zitzewitz 2006). The finding of generally smaller effects than those previously reported is intriguing and point to a potentially important element affecting subjective evaluation of performances: when stakes are higher, as in the prestigious Olympic games, the potential influence of biased judgments is likewise increased. The reason our results show less biases than previous research may thus be that the stakes in each of the World Cup series events are lower than the stakes in e.g. Olympic events (cf. Zitzewitz 2006). The mechanisms that lead to biases are somehow more often invoked or invoked in a more influential version when stakes are higher. Another potential explanation points to individual judges' career considerations, as objective performance in regular contests can later be rewarded by participation in higher prestige events, such as the Olympics. Our results thus complement Zitzewitz's (2006) findings by pointing out situations in which biases are more (less) likely. Future research could compare competitions that differ in status and prizes, for example competitions at the regional and the national levels, and further test this idea of bias as emergent in high-stakes competitions.

Our estimates of nationalistic biases were also small enough that they would not alter the rankings in any of the contests. A comparison with the within-jump variance indicates that the bias only amounts to around a quarter of a standard deviation (cf. Model 3 in Table 2). In other words, the

variation between the judges' scores that is not due to nationalistic bias is much larger. If this variation is considered to be random deviations from a "true" score, it is clear that competitors should worry more about chance variations in judges' evaluations (and possible idiosyncratic opinions) than about nationalist biases affecting the outcome of the competition.

We conducted separate analyses for home-field bias in performance (as indicated by the length of the jumps) and in judges' evaluations of the stylistic elements in each jump. The results showed no effects on the jumpers' actual performance, implying that the jumpers did not benefit from familiarity with the hill or from home-audience support in terms of jump length. Instead, home-field bias showed up in the judges' evaluations of the jumps, even though the effect was here again of very small magnitude. Whether the bias stemmed from "crowd-pleasing"—which functions as a conflicting motivation to objectivity in the judges' performance—or some other bias-generating mechanism is beyond the scope of this study.

As has been pointed out in previous research, biases in evaluations need not result from conscious actions by the judges. In related research on hiring practices in symphony orchestras (cf. Fasang 2006), researchers have claimed that evaluators have non-conscious schemata based on external factors such as gender, which affect the perception of performances and which can be removed by blinded evaluations. In a similar manner, judge evaluations in sport contests can be affected by non-conscious biases. A judge may over- or underrate an athlete's performance based on her or his nationality, or the judge may use the behaviour of the audience as information on this performance. In such cases, ways to minimize the effects of such factors should be promoted to reduce the effects of biases on evaluations.

CONCLUSION

We estimated nationalistic and home-field advantages in ski jumping using data from the entire FIS World Cup 2006/2007 and 2007/2008 seasons. From our results, we confirm parts of the extant literature, and document both national and home field biases in judges' evaluations. The biases were generally small, and do not seem to pose a major threat to the validity of results from ski jumping.

competitions. However, in league or series-type competition tournaments, the cumulative effects of nationalistic bias may be important. An important takeaway from this study is that the system of regulations in

Biases in subjective evaluations of judges and referees may carry important implications for sports results, as exemplified by findings of potential effects on final rankings of athletes or on betting markets (Larsen, Price, and Wolfers 2008, J. Price and Wolfers (2010)).

Knowledge of the sources of evaluative biases in competitive sports can potentially help us better understand biases and discriminatory practices in other fields, such as student and employee recruitment and the allocation of research grants. Our findings of nationalistic and home-field biases have provided evidence that, in particular when assessed together with other related findings, add to our understanding on bias-generating practices. This knowledge may aid sport organizations, as well as other types of organizations, in their control of similar biases, and produce fairer, performance evaluations in the future.

REFERENCES

Anders, Anne, and Kurt William Rotthoff. 2014. "Is Home-Field Advantage Driven by the Fans? Evidence from Across the Ocean." *Applied Economics Letters* 21 (16): 1165–8.

Balmer, Nigel J., Alan M. Nevill, and A. Mark Williams. 2001. "Home Advantage in the Winter Olympics (1908-1998)." *Journal of Sports Sciences* 19 (2): 129–39. doi:[10.1080/026404101300036334](https://doi.org/10.1080/026404101300036334).

Bursell, Moa. 2012. "Name Change and Destigmatization Among Middle Eastern Immigrants in Sweden." *Ethnic and Racial Studies* 35 (3): 471–87. doi:[10.1080/01419870.2011.589522](https://doi.org/10.1080/01419870.2011.589522).

Callahan, Brian P, Sean E Mulholland, and Kurt W Rotthoff. 2016. "Cultural Bias: Gymnasts, Judges, and Bilateral Trade Agreements." *The Journal of SPORT* 5 (1): 3.

Carmichael, Fiona, and Dennis Thomas. 2005. "Home-Field Effect and Team Performance: Evidence from English Premiership Football." *Journal of Sports Economics* 6 (3): 264–81.

Emerson, John W., Miki Seltzer, and David Lin. 2009. "Assessing Judging Bias: An Example

from the 2000 Olympic Games.” *The American Statistician* 63 (2): 124–31. doi:[10.1198/tast.2009.0026](https://doi.org/10.1198/tast.2009.0026).

Fasang, Anette. 2006. “Recruitment in Symphony Orchestras: Testing a Gender Neutral Recruitment Process.” *Work, Employment and Society* 20 (4): 801–9.

Larsen, Tim, Joe Price, and Justin Wolfers. 2008. “Racial Bias in the NBA: Implications in Betting Markets.” *Journal of Quantitative Analysis in Sports* 4 (2).

Lock, R, and KF Lock. 2003. “The Statistical Sports Fan: Judging Figure Skating Judges.” *STATS* 36: 20–24.

Meyer, MA, and JM Booker. 1991. “Selecting and Motivating the Experts.” *Eliciting and Analyzing Expert Judgment: A Practical Guide*, 85–98.

Müller, Wolfram. 2008. “Performance Factors in Ski Jumping.” In *Sport Aerodynamics*, 139–60. Springer.

Myers, Tony D, Nigel J Balmer, Alan M Nevill, and Yahya Al Nakeeb. 2006. “Evidence of Nationalistic Bias in Muaythai.” *Journal of Sports Science & Medicine* 5 (CSSI): 21.

Nevill, Alan M, Nigel J Balmer, and A Mark Williams. 2002. “The Influence of Crowd Noise and Experience Upon Refereeing Decisions in Football.” *Psychology of Sport and Exercise* 3 (4): 261–72.

Page, Lionel, and Katie Page. 2010. “Evidence of Referees’ National Favouritism in Rugby.” *Paper Provided by National Centre for Econometric Research in Its Series NCER Working Paper Series with*, no. 62.

Price, Joseph, and Justin Wolfers. 2010. “Racial Discrimination Among NBA Referees.” *The Quarterly Journal of Economics* 125 (4): 1859–87.

Price, Joseph, Marc Remer, and Daniel F Stone. 2012. “Subperfect Game: Profitable Biases of NBA Referees.” *Journal of Economics & Management Strategy* 21 (1): 271–300.

Sampaio, Breno. 2012. “Three Essays on Applied Microeconomics.” Doctoral dissertation, University of Illinois, Urbana-Champaign. <https://core.ac.uk/download/pdf/4837557.pdf>.

Seltzer, Richard, and Wayne Glass. 1991. “International Politics and Judging in Olympic Skating Events: 1968-1988.” *Journal of Sport Behavior* 14 (3): 189.

Yang, Mary. 2006. “Cultural Biases in Olympic Figure Skating Judgments.” Senior Thesis,

Evanston, IL: Northwestern University. <http://mmss.wcas.northwestern.edu/thesis/articles/get/591/Yang%202006.pdf>.

Zitzewitz, Eric. 2006. "Nationalism in Winter Sports Judging and Its Lessons for Organizational Decision Making." *Journal of Economics & Management Strategy* 15 (1): 67–99. <http://onlinelibrary.wiley.com/doi/10.1111/j.1530-9134.2006.00092.x/full>.