NATIONALIST 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 represent a major threat 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 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 over time, which suggests that increasing professionalization of sports also may limit the influence of subjective and nationalist impulses.

Keywords: nationalist bias, skijumping, judges, fixed effects

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 biases in several ways. First, we assess to what extent nationalistic bias can be explained by unobserved characteristics of the performance. Second, we consider national 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.

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 concludes that 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 concieve of bias resulting from at least three different mechanisms. 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 princple 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 influences by the works of Boltanski & Thevenon (Boltanski and Thévenot 1991). 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 evalution. 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 would 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 stricted 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 (Zitzewitz 2006), 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 reexamine the existence of nationalist biases, study how these biases vary across the countries from which evaluators are drawn, and study whether nationalist biased have decreased in magnitude over time, due to, for example, 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.

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, historical and professional ideosyncracies at the national level. The strength of the bias is assumed to be lower in countries without strong ski jumping cultures and higher in countries where the sport is more peripheral in the larger sports field.

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

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 for results according to our pre-specified instructions. In accordance with norms of open social science, the software used to download and scrape the PDF files are documented in the GitHub repository associated with this article.[^1: See https://github.com/torkildl/skijumping for data and software.]

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
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 hypoteses. We estimate the models separately for the two periods we have data for, as there may be 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 bias using fixed effects models

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 test our different hypotheses.

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

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

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. Hypothesis #1 is tested by assessing the direction and magnitude of B.

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

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

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. The B estimate recovered from this model will be compared with the corresponding parameter from model 1, and the result of this comparison can be applied to test hypothesis #2.

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 an adjusted model #2b:

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

This model is an improvement over the previous ones, but limitations still remain. For example, there could be so-called compensating biases, where other judges on thethe same panel as a suspected biased judge compensate by lowering their scores when they know that one of the others will up his score. Finally, the model also assumes that bias is constant across all judges from all nations, a crucial assumption that we subsequently question in hypothesis #3.

Variations across nations and time periods

In order to test our third hypothesis, we want to allow for variations in bias between groups of judges from the same nation by interacting the judge fixed-effect in model #2 with the indicator function for judge and competitor nation correspondence. This implies the followung model:

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

where L_j is an added judge country fixed-effect. The nation-specific estimates for B provide an

assessment of the average bias in scores by judges from different countries.[1] We also rank countries by our selected instrument of ski jumping history, namely the historical record of Olympic ski jumping gold medals in the decades before the seasons we study. The hypothesis is tested by estimating the Spearman rank correlation ρ across the countries. A negative correlation indicates support for the hypothesis.

Our fourth hypothesis regards potential changes in the variation of nationalist bias over time. In order to assess changes in the variation in nationalist bias, we estimate a model akin to model #2, but with an added interaction with a dummy representing the second season (with the first season as a reference group). The parameter estimate for this interaction effect, Γ provides a direct test of hypothesis with regard to changes in nationalist bias. The model is written as

$$s_{ijp} = B{\cdot}\phi(I = J) + \Gamma{\cdot}Y{\cdot}\phi(I = J) + q_p + \epsilon_{ijp} \tag{5}$$

RESULTS AND DISCUSSION

Our analysis of nationalistic bias proceeds in four steps, with each step corresponding to a test of one of our four hypotheses.

Nationalistic biases in judge scores

Table 2 presents the results from the first three estimations. 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, for example, cultural and historical differences between countries that lead to an over- or underappreciation of certain styles.

[1] In principle, we could obtain individual bias estimates for each judge and then analyze these as samples of judge biases scores grouped by judges' nationalities. This would be very demanding of our data and experiments indicate that estimates would be very imprecise. We thus refrain from doing this, and only allow the bias to vary between the different judge nationalities.

Table 2: Results from regression models of national bias

	Model 1: Naïve	Model 2a: Within-jump	Model 3a: Within-judge
National bias	0.136*** (0.088 0.185)	0.065*** (0.049 0.082)	0.075*** (0.059 0.091)
Constant	17.430*** (17.416 17.445)		
Observations	23,945	23,945	23,945
R^2	0.001	0.919	0.926
Adjusted R^2	0.001	0.898	0.907
Residual Std. Error	1.100 (df = 23943)	0.351 (df = 19155)	0.336 (df = 18985)
F Statistic	30.710*** (df = 1; 23943)		

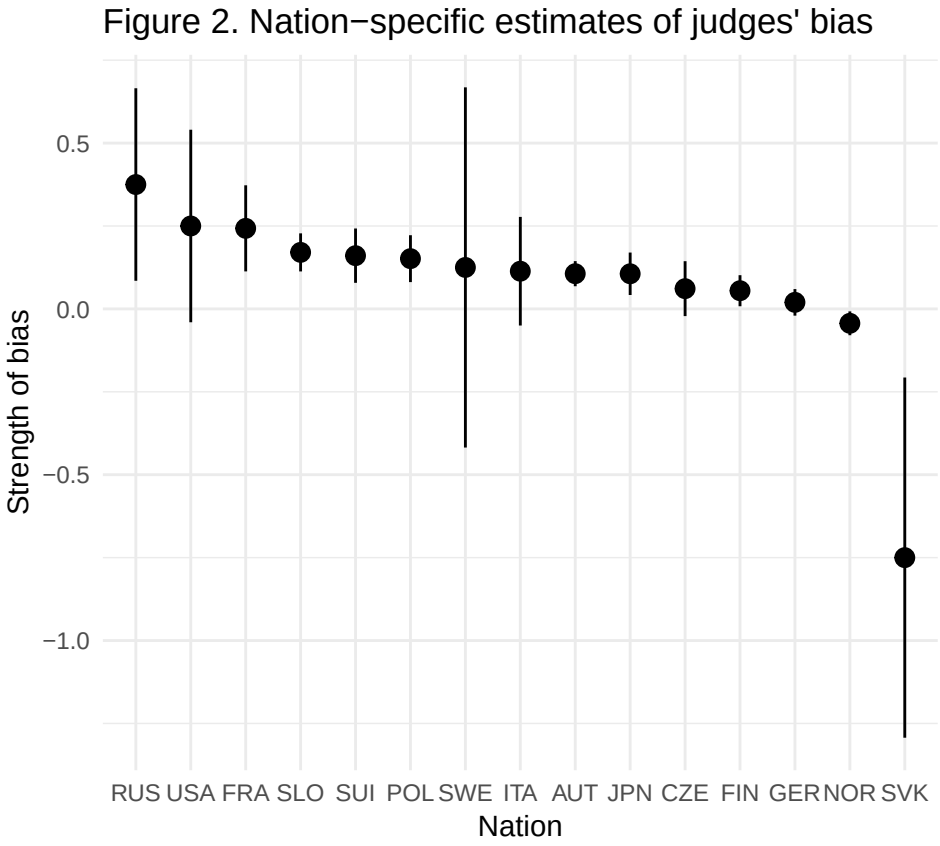
Note: *p<0.05; **p<0.01; ***p<0.001
In models 2 and 2b, the intercept is suppressed by the fixed effects estimation procedure.

This was the concern of the second hypothesis, were we assumed that the bias level would decrease once we controlled for other, unobserved characteristics of the jump. The second model takes up exactly this challenge, 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. In our adjusted second model #2b, where we also include measures of individual judges' leniency, we obtain a comparable estimate of nationalistic bias.

Variations in bias across nations and time periods

Our third hypothesis concerned variation in nationalistic biases for judges from different countries, and whether such variation is correlated with an indicator of the historical importance of ski jumping in that country.

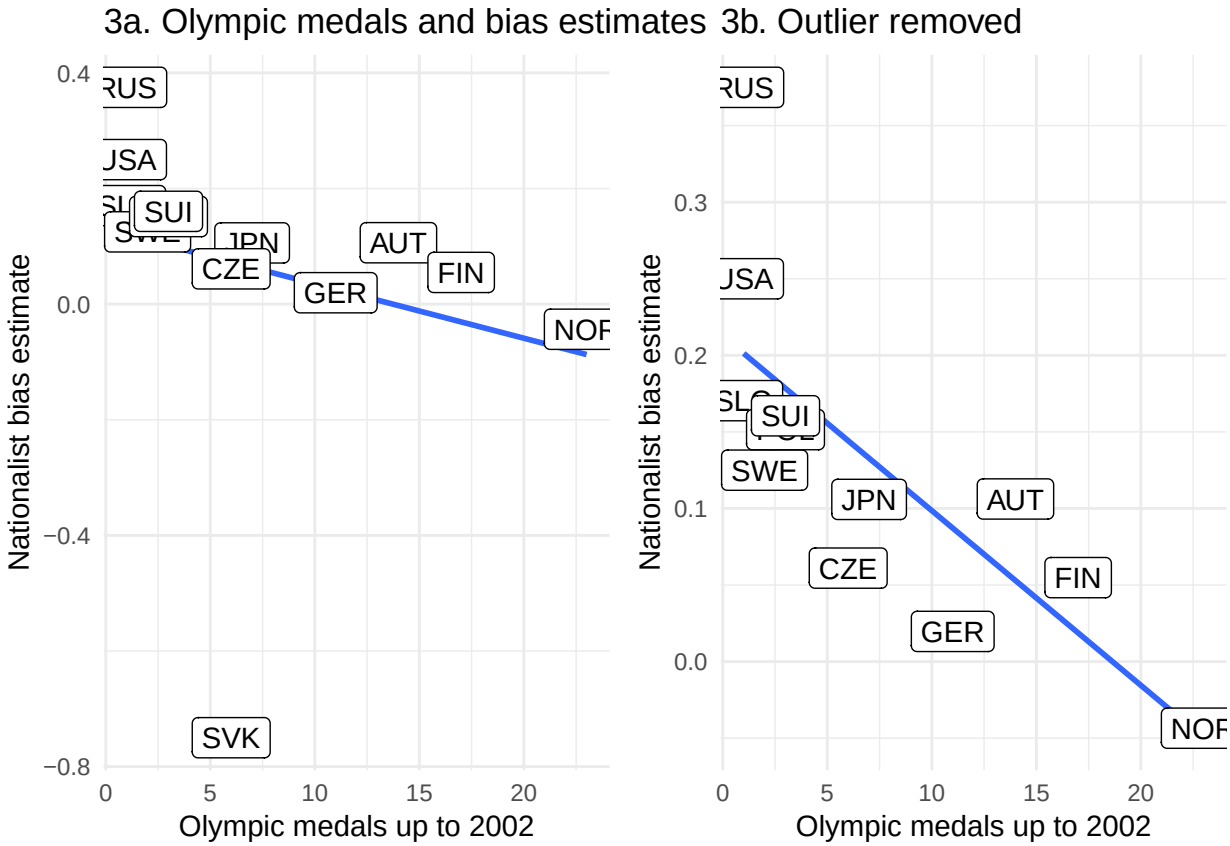
Our third model allowed estimating such variation, and the results are shown in Figure 2. Judges from Russia, the United States and France appear most biased. Norwegian and Slovakian judges seems to hold a negative nationalistic bias, where same-nation jumpers are punished more severely than others. Overall, the estimated biases seem 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 magnitude. The caseof Slovakia is curious, as this country show particularly strong negative bias. The estimate lie well outside of the rest of the estimates, and we will treat it as an outlier in the following analysis.



To properly test the third hypothesis, we then estimate the rank correlation of the nation-specific bias coefficient and the nation's Olympic medal harvest. This exercise was done by scraping existing public statistics on Olympic medals, and calculating how many medals each country has collected.

The Spearman Rho rank correlation coefficient of nationalist bias and Olympic medals was estimated at `spearman$estimate` which thus supports the hypothesis that the magnitude of bias is negatively related to the historical importance of ski jumping in the country (as measured by our index of Olympic medals).

Figure 3 shows the scatterplot between Olympic medals and natonalistic bias estimates, with and without the outlier Slovakia included.



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.

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

In this paper, we report an analysis of data from the entire FIS World Cup 2006/2007 and 2007/2008 seasons in Nordic ski jumping. Our overarching research question is whether there is nationalistic biases in ski jumping, and whether there are systematic variations in such bias when influences of the actual performance are ruled out. From our results, we confirm a tenet of the extant literature, and document nationalistic biases in judges' evaluations. The biases were generally small, and do not seem to pose a major threat to the validity of results from skijumping 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 ski jumping that is already in place most likely provide adequate mitigation of judges' nationalist impulses so that the objectivity and fairness of the results are not threatened.

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

Boltanski, Luc, and Laurent Thévenot. 1991. "De La Justification." *Les économies de La Grandeur*.

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.

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.

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.

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.

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.