Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Quality, Subjectivity, and Sustained Superior Performance at the Olympic Games

David M. Waguespack, Robert Salomon

To cite this article:
David M. Waguespack, Robert Salomon (2016) Quality, Subjectivity, and Sustained Superior Performance at the Olympic Games. Management Science 62(1):286-300. http://dx.doi.org/10.1287/mnsc.2014.2144

MANAGEMENT SCIENCE
Vol. 62, No. 1, January 2016, pp. 286–300
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.2144

Quality, Subjectivity, and Sustained Superior Performance at the Olympic Games

David M. Waguespack
Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20742, dwaguesp@rhsmith.umd.edu

Robert Salomon
Leonard N. Stern School of Business, New York University, New York, New York 10002, rsalomon@stern.nyu.edu

In some competitions, performance evaluation includes a substantial subjective component. We argue that the inherent uncertainty and ambiguity in subjective evaluation can lead to favorable ex post treatment for reputationally privileged competitors. Post consumption, judges may infer quality that is not directly observed and/or make conservative choices to assuage accountability concerns. We examine these issues in the context of the Olympic Games, comparing country-level performance outcomes across Olympic sports. We find that past performance is predictive of current performance in all sports, but the effect is stronger in subjective outcome sports versus objective outcome sports. That is, past performance is a better predictor of future performance in sports where external judges and referees can influence the outcomes. We find the same pattern in individual boxing matches, with past country-level performance having a stronger effect on subjective boxing outcomes (judges' decisions) than objective boxing outcomes (knockouts).

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mnsc.2014.2144.

Keywords: reputation; status; contests
History: Received July 10, 2013; accepted December 3, 2014, by Jesper Sørensen, organizations. Published online in *Articles in Advance* June 29, 2015.

1. Introduction

Why are past winners likely to win again? One conventional explanation highlights differences in durable skills, such as the various heterogeneous resources and capabilities of interest to strategy scholars. In this case, past and present performance are linked by inalienable endowments or to hard-to-replicate investments in quality. A complementary explanation suggests that persistent performance differences can arise from favorable ex ante treatment provided to competitors that facilitates superior performance (Merton 1968, Podolny and Phillips 1996, Correll et al. 2012). That is, certain competitors, perhaps based on past performance, are given access to better resources and/or opportunities, which leads to better performance in subsequent rounds of competition (Sorenson and Waguespack 2006).

In this paper, we highlight and provide evidence for an additional mechanism likely to be driving persistent superior performance, one that we refer to as *subjective cumulative advantage*. In contrast with the mechanisms described above, subjective cumulative advantage entails favorable ex post treatment of reputable actors—those for whom future performance expectations are based on past performance outcomes (see Jensen et al. 2012). We suggest that favorable ex post treatment is likely to manifest in cases where the evaluation of performance includes a subjective component and when performance outcomes are determined by third-party judges. In such situations, we posit that the uncertainty and ambiguity associated with quality prompts evaluators to look for social clues believed to correlate with quality (Podolny 1993). Moreover, subjectivity is likely to evoke accountability concerns among judges. Reputations can therefore serve as a powerful decision-making guide (Correll et al. 2012) and/or a crutch that helps evaluators justify decisions (Correll et al. 2012, Jensen 2006, Waguespack and Sorenson 2011). The net result is that subjective performance judgments anchor on past winners, leading to systematic patterns of performance persistence.

We examine these issues in the context of the Olympic Games, a setting with a mix of objective and subjective contests. We conduct two distinct tests to see whether country-level reputation drives performance more in subjective contests. First, we examine the relationship between past and present country-level medal performance across objective, refereed, and subjective sports. In regression models that parallel Bernard and Busse (2004) but are disaggregated by sport, we find that past performance is predictive of current performance in all sports, but the effect is more persistent in refereed and subjective sports than

objective sports. Second, we investigate the relationship between past performance and current performance in the outcomes of individual boxing matches, which have the special feature of mixing objective and subjective outcomes. Although boxers from countries with more prior medals are more likely to win, reputation has a stronger affect when outcomes come down to judges' decisions. Taken together, the results suggest that reputation influences performance more in subjective contests, as judges exhibit a preference for participants from countries that have performed well in the past. The findings therefore provide support for a reputational mechanism long thought to be driving persistent performance but that has received little empirical substantiation.

The remainder of this paper proceeds as follows. In §2, we briefly review the literature on persistent performance and sketch out the theoretical mechanisms behind *subjective cumulative advantage*. In §3, we provide an overview of the data, describe the modern Olympic Games, and discuss the institutional features that make it a good setting for detecting persistently superior performance resulting from subjective quality standards. In §4, we present tests for the effects of prior medal performance on current medal performance by nation across objective, refereed, and subjective sports. In §5, we present tests for the effects of prior country-level medal performance on the outcomes in individual boxing matches. In §6, we addresse sensitivity and interpretation issues. In §7, we conclude by discussing implications and possible extensions to this work.

2. Persistence in Performance and Subjective Cumulative Advantage

There are several broad sets of explanations in the economics and sociology literatures that explain persistent superior performance. One conventional explanation revolves around the possession of superior capabilities. For example, unobservable differences in innate ability might lead some individuals, or organizations, to persistently outperform others. In an Olympic context, consistent superior performance by Michael Phelps (who won six gold medals in 2004 and eight gold medals in 2008) might be a result of inalienable advantages in wing span, foot size, and/or lung capacity. At the national level, countries might similarly possess immutable comparative advantages in certain sports as a consequence of natural endowments—e.g., Kenya's hills for distance running and Switzerland's mountains for skiing (Bernard and Busse 2004).

In addition to inalienable or immutable comparative advantages, persistence in performance can result from targeted investments in skills and capabilities. Individuals can dedicate themselves to skill development through training, and nations can invest in specific sports and/or specialized training facilities in an effort to generate sustainable advantage in a particular domain. Consistent with that conjecture, Bernard and Busse (2004) suggest that one reason why some countries systematically outperform others at the Olympic Games is because they have developed hard-to-imitate capabilities born out of specialization, and repeated investment, in certain sports, e.g., Cuba in boxing or Japan in baseball.

Another explanation for persistent performance has to do with reputational effects (Merton 1968, Podolny 1993, Gould 2002). That is, individuals or organizations that establish a reputation for quality based on past performance might be rewarded with resources that help perpetuate their advantage in subsequent contests. In a sporting context, one form of ex ante preferential treatment can be observed in contests where competitors are seeded based on past performance (see Weiss 1986). Seeding favors reputable competitors by affording them an "easier" route to the championship. Another set of studies documents that elite athletes are more likely to be born in the months immediately following the youth age cutoff in their sport. Because children born in the months after the age cutoff are older than those they compete against, they have more developed gross and fine motor skills, and they therefore exhibit greater skill. As a result, they are more likely to be singled out by their coaches to receive better training, beginning a snowball effect that influences their future career trajectories. This is known as the "relative age effect" (see Musch and Grondin 2001).

Although skill possession, skill acquisition, and reputational advantages are all viable explanations for observed patterns of persistence in performance outcomes, in this study we highlight an alternative mechanism that has received relatively little empirical attention. We refer to this mechanism as *subjective cumulative advantage*. In contrast with the ex ante reputational advantages described above, it hinges on the preferential ex post evaluation for reputationally privileged actors. The result is performance gaps that are more persistent than what underlying ability variance or favorable ex ante treatment alone will produce. In the following sections, we discuss the nature of subjectivity in competition and how it relates to reputation. We then return to the issue of detecting subjective cumulative advantage in applied settings.

In many competitive environments, quality is objective in that market participants broadly agree on its underlying dimensions, and outputs can be observed and measured. In such situations, it is easier for observers and evaluators to discriminate among competitors ex post because winners are ordinarily those

whose outputs display greater levels of those underlying objective quality dimensions. In some settings, however, quality is uncertain and ambiguous, comprised of subjectively determined attributes that are more difficult to observe and measure. In such cases, it is more difficult to discriminate amongst competitors, and evaluators play a much larger role in determining outcomes.

The process of evaluation and the influence of evaluators on performance outcomes in subjective contests has received attention in several fields. Perhaps most directly related to the context at hand is the literature on personnel performance appraisals, where the distinction between objective and subjective indicators, and the potential for bias on the part of evaluators, is a subject of extensive attention (Arvey and Murphy 1998). In a review and meta-analysis, Bommer et al. (1995) note that outcomes of objective and subjective personnel performance appraisals are only weakly correlated ($r = 0.389$). They also point out that subjectivity has two distinct archetypal interpretations.

At one extreme, subjectivity reflects uncertainty with respect to well-defined, but costly to measure, objective constructs. For example, in a sporting event, a well-defined outcome might have taken place but the evaluator has difficulty properly observing it, as with judgment calls among sporting referees. This notion of subjective error underpins calls for the use of instant replay in sports (Kim and King 2014).

At the other extreme, subjectivity is a matter of post-transaction uncertainty and ambiguity. This fundamental ambiguity is inherent to situations where different actors can interpret the same multidimensional construct differently. For example, consider the difference between one athletic contest where the winner is the "fastest" competitor and another where the winner is the "best" competitor. In the former case, "fast" is straightforward and measurable, although it might be hard to observe this with the naked eye, as in the case of a photo finish. In the latter case, "best" is an ambiguous construct open to interpretation. Moreover, there can be various subjective dimensions of "best" that include relative power, grace, artistry, etc. When subjectivity of this sort exists, evaluation is more complicated.

The notion that evaluation is complicated by post-transaction uncertainty and ambiguity is resonant with descriptions in other fields. For example, Darby and Karni (1973) proposed that the quality of certain goods and services, such as the provision of repair services, is costly to ascertain, even after experience with them. Such "credence goods" complicate market mechanisms by potentially creating a stubborn gap between intrinsic quality and perceived quality. Karpik (2010) takes things a step further by proposing that "singular" goods and services, such as artistic performances, are marked by incommensurability, multidimensionality, and uncertainty. This creates ambiguity, he argues, and implies an inability to compare objective quality attributes for such goods. In such situations, market mechanisms break down, and relative value becomes more a function of hedonics (see also Neelamegham and Jain 1999, Yogev 2010). In the case of subjective performance appraisal, therefore, multidimensionality and scaling choices can result in the inconsistent use of metrics (Campbell 1990). There is also scope for manipulation. White (2002) even goes so far as to suggest that when quality is multidimensional and subjective, the decision about superior performance can come first, with the factors supporting it second.

Uncertainty and ambiguity, however, do not necessarily imply that evaluation is systematically biased as opposed to simply noisy. For the second leg of our argument that subjectivity allows for advantages to persist—i.e., to be cumulative ex post—we draw on the status dynamics literature, which has two important insights with respect to the maintenance of hierarchies. First, that literature suggests that reputational advantages derived from past success can serve as a signaling device, prompting evaluators to infer quality that they might not have observed in outputs (Merton 1968, Sauder et al. 2012, Zuckerman 2012). Advantages might therefore persist because reputationally disadvantaged competitors are, effectively, held to higher standards (Foschi 1996). Second, reputational advantages derived from past success may serve as a coordination device, enabling evaluators to gauge the likely responses of other evaluators to the same information (Correll et al. 2012). In this scenario, evaluators face accountability concerns as to how their own opinion is regarded and anticipate that conservative choices will be subsequently easier to justify and defend (Jensen 2006; see also Garicano et al. 2005).[1]

Recent work has applied these insights to ex post evaluation. For example, Kim and King (2014) demonstrate that high status baseball players are more likely to benefit from judgment calls made by umpires, nicely illustrating how quality can be inferred even when not precisely observed. Waguespack and Sorenson (2011) find that, holding film content constant, rating boards are more likely to deem movies from high status studios appropriate for younger audiences. The bias was even more pronounced in countries with industry-funded review boards, indicative of a situation in which accountability to the industry drives ratings. Correll et al. (2012) drill more precisely into both subjective evaluation mechanisms.

[1] Interestingly, in the case of gender bias, Foschi (1996) finds that accountability concerns reduce the propensity to favor the advantaged group.

They asked subjects in an experimental setting to evaluate two identical chocolates distinguished only by wrapper color. When subjects were later told that the chocolate they had not selected was the higher status brand, 15% switched their preference in a second round of testing, possibly inducing consumers to think that there was some subtle quality element that they initially missed. The switching rate was even greater (50%) among subjects told subsequently that they must recommend one of the chocolates to others, apparently evoking accountability concerns about the recommendation.

For all the reasons described above, we believe that preferential ex post treatment is likely to have an effect on persistent superior performance beyond skill heterogeneity and preferential ex ante treatment. Although scholars recognize the possibility for such an ex post reputational effect, testing this supposition has proved challenging because durable skill heterogeneity and preferential ex ante treatment are also likely to be operative in settings where we observe persistent performance gaps. The challenge is in empirically disentangling these factors.

Our goal, therefore, is to construct parallel worlds where we can compare performance persistence with and without subjective quality evaluations, similar to the Salganik et al. (2006) study that explores social influence in music markets. Toward that end, we will compare results across objective and subjective competitions over time at the Olympic Games, while seeking, to the extent possible, to hold constant skill differential and favorable ex ante treatment. If our priors are correct, we expect superior performance to persist for longer in subjective competitions than in objective competitions. This is because evaluators—i.e., judges—are likely to anchor on reputation after the fact. The size of the subjective reputational advantage, if present, should manifest as a greater marginal effect of past performance on present performance in subjective contests and can be viewed as a form of returns to reputation.

The next section of the paper briefly outlines the history of the modern Olympic Games and discusses institutional features that allow us to isolate subjective reputational effects.

3. The Olympic Games

Given that an estimated 70% of the world's population watched some portion of the 2008 Summer Olympics in Beijing (Nielsen 2008), we assume that most readers are somewhat familiar with the Olympic Games. The basic format in the modern Olympic Games is that athletes and teams from each participating country compete in a series of events. In each contest, the top three athletes or teams in order are awarded gold, silver, and bronze medals.[2] The scope of, and participation in, the Olympics has grown steadily over time. At the Athens Olympiad in 1896, 241 athletes from 14 countries participated in 43 discrete medal-awarding events. For the 2006 Winter Games, 2,508 athletes from 80 countries were represented, competing in 84 events. In the 2008 Summer Games, 10,942 athletes from 204 countries participated in 302 events (IOC 2013).

Organizationally, the modern Olympic Games began with the founding of the International Olympic Committee (IOC) in 1890 by Pierre de Coubertin. Between 1896 and 1924, the Olympics were run in conjunction with World's Fairs (Guttman 1994). In 1924 the Olympics split into separate Summer and Winter Games, and the Olympic movement assumed the same basic organizational structure it holds today. At the top of the organizational chart is the IOC, which manages the selection of the Olympic host city and serves a coordinating and sanctioning role for constituent organizations (IOC 2010). Three constituents are particularly important. National Olympic Committees (NOCs) manage the participation of athletes from each country, including procedures used to select national representatives. International Sports Federations (ISFs) are nongovernmental organizations that establish the rules and procedures for each of the various sports competitions. Finally, Organizing Committees (OCs) are responsible for running the games in the host city.

Beyond this broad sketch, several institutional features of the games are important for understanding our analysis of persistent performance. First, the Olympics embrace a broad participatory and egalitarian ethic, while also limiting the number of athletes from any given country. For instance, NOCs with no swimmers that meet the minimum qualifying standards may still enter athletes, while each nation is limited to two qualified racers per event (Fédération Internationale de Natation 2010). For tournament sports with limited spots, such as basketball, qualification is determined by global and regional tournaments (FIBA 2012). In practice, however, the Olympic athletes are not necessarily the best in the world, largely because de Coubertin hoped to advance internationalism over the forces of nationalism. Although there is certainly great latitude at the NOC level in the selection and training of athletes, an important institutional consideration for this paper is that countries cannot simply be excluded from competition (Guttman 1994).

Moreover, all competitors are given equal treatment at the games. For instance, all athletes, regardless of expected performance, are given equal time to train in the competition facilities. Furthermore, for sports

[2] Boxing, an exception, awards two bronze medals.

with head-to-head competition such as basketball or boxing, the pairing of competitors is random. These stipulations mean that there are limited ex ante positional advantages afforded to elite competitors, such as a less competitive route to the medal rounds. The country organizing the games may gain some advantages, however, because it automatically qualifies to enter competitors in all events and because it benefits from greater familiarity with local conditions (Bernard and Busse 2004).

Second, Olympic officiating is, in essence, a highly scrutinized and professionalized career in itself. The IOC and ISFs report devoting considerable attention and resources to attracting qualified experts to judge the events and to detecting cheating (IOC 2008). Thus, we have some confidence that reputational effects do not simply reflect novices' inability to detect subtle objective quality attributes. More importantly, and much like the career pressures faced by other market intermediaries (for example, see Hong and Kubik 2003 on financial analysts), Olympic officials are ostensibly concerned with performance evaluation. One manifestation of this accountability concern is a tendency for judging to reflect nationalistic biases (Emerson et al. 2009) and for those biases to increase when officiating appointments are determined by participant nations rather than an unaffiliated international organization (Zitzewitz 2006).

Although each sporting federation has its own processes for selecting judges, there is anecdotal evidence to suggest that officials compete for the distinction of working at medal events. The ISFs have historically struggled to balance perceived ability and perceived impartiality in these selections. For instance, in the 2014 Olympics, the International Hockey Federation controversially appointed a crew of experienced Canadian officials to referee the gold medal match between the historically dominant Canadians and Sweden (Longley 2014). Likewise, a boxing official at the Beijing Olympics noted that even while boxing encourages judges to behave independently, those with deviant scores do not advance to judge medal rounds (McNally 2012). The precise constraints and incentives for Olympic officials are clearly germane to decision-making bias. Unfortunately, detailed longitudinal information on officiating procedures and individual officials is rarely available (Emerson et al. 2009).

Ambiguity about historical officiating procedures notwithstanding, and as we explain in more detail in the following section, the Olympics includes a variety of contests readily classifiable as objective or subjective in nature. Their objective and subjective characteristics, along with the stability of those classifications across sports over time, allows us to isolate evidence of systematic reputational bias, as manifest in performance persistence across event types.

4. Country-Sport Analysis

To examine the influence of past reputation on current performance outcomes, we test whether persistent medal performance differences exist over time across countries in sports of varying objectivity/subjectivity at the modern Olympic Games. Our data include results from both the modern Summer and Winter Olympics. The Summer Games were first held in 1896 and have taken place every four years since, with the exception of 1916, 1940, and 1944. The Winter Games were first held in 1924 and took place in years corresponding with the Summer Games until 1992.[3] Since then, the Summer and Winter Games have alternated every two years.

To generate the sample, we collected data from the IOC for every Olympiad from 1896 through to 2006 (IOC 2008). From this list we excluded participants from countries that were part of an "international" team (i.e., teams made up of participants from several countries) and countries that participated in only one Olympic Games. Table A1 in the online appendix (available as supplemental material at http://dx.doi.org/10.1287/mnsc.2014.2144) lists the participating nations included in our sample.

In this study, we focus on country performance at the sport level. This represents a slight departure from previous analyses of Olympic performance. For example, Bernard and Busse (2004) conducted analyses of performance at the country-Olympiad level. They were interested in explaining medal counts aggregated up to the national level. Although we build upon their work, we account for the exact sports in which individual countries participate. We do so because we are interested in explaining country-specific persistence in performance in particular domains over time. That is, we are more interested in understanding how a certain country (e.g., China) performs in a given sport (e.g., table tennis) than the number of total medals it has garnered across all sports at a particular Olympic Games.

At the same time, we focus on the sport level (e.g., swimming) rather than the disaggregated event level (e.g., 200 m freestyle) because there often is a substantial skill overlap among individual events in a sports category. This overlap can lead to dominance of a number of similar events by one person (e.g., Michael Phelps in swimming events at the 2008 Beijing Games) or a group of persons (e.g., China in table tennis singles and team events). Looking at the sport level gives us the advantage of better accounting for unobservable individual- and nation-specific

[3] Winter sports such as figure skating and ice hockey were played in the Summer Games before the Winter Games were introduced. Refer to Table A2 in the online appendix for the years of introduction.

skill heterogeneity, and averages out random components in event-level competition. It can also help us better account for preferential ex ante treatment provided athletes in a given sport because the sport-specific component is isolated. We note that we do not include sports in our sample that are categorized as demonstration sports (e.g., roque) or sports that had a brief history and are no longer played (e.g., tug of war). Table A2 in the online appendix lists the sports included in this study with their respective years of introduction and years played.

Because we are interested in examining sports based on whether outcomes are determined using more objective or subjective criteria, we classify each sport into one of three categories—objective, refereed, and subjective—based on our reading of the rules and regulations of each sport as published by each Olympic Sporting Federation recognized by the IOC.

The objective category refers to those sports, such as track and swimming, in which outcomes are determined objectively. In objective sports, winners are determined based on finite outcomes (e.g., fastest time or first to reach a specific point), and officials mostly serve to document outcomes. Refereed sports include those in which on-field officials make frequent in-game judgments, as in soccer or basketball, about rules violations that can alter the course of play. Although the officials in refereed sports do not necessarily directly determine the outcomes, they can certainly influence the outcomes. Finally, in sports we refer to as subjective, such as gymnastics or diving, official judges play a central role in determining outcomes. Table 1 details the sports assigned to each category.

Using the various sources of country and sport data collected from the IOC, we created a matrix of participating nations at every Olympiad in every sport. This creates a sample of 208 countries, 42 sports, 45 Olympiads (25 summer and 20 winter games), yielding 20,277 country-sport-Olympiad observations. As we limit our analysis to those observations for which we have medal totals for three games prior to the focal game, our final usable sample is reduced to 18,046 country-sport-Olympiad observations.

Table 1 Classification of Olympic Sports

	Objective	Refereed	Subjective
Summer	Archery	Badminton	Diving
	Canoe/kayak	Baseball	Gymnastics
	Cycling	Basketball	Judo
	Equestrian	Fencing	Synchro swimming
	Field	Handball	Taekwondo
	Modern pentathlon	Hockey	
	Rowing	Soccer	
	Sailing	Softball	
	Shooting	Table tennis	
	Swimming	Tennis	
	Track	Volleyball	
	Triathlon	Water polo	
	Weightlifting	Wrestling	
Winter	Alpine skiing	Ice hockey	Figure skating
	Biathlon		
	Bobsleigh		
	Curling		
	Lugeing		
	Nordic skiing		
	Speed skating		

Note. Boxing, which includes elements of all three types, is excluded from the cross-sport analysis.

4.1. Data

Because we are interested in performance—i.e., country performance in each sports category—we include a measure of medal performance as our dependent variable. *Medal_total* is a count variable that captures the total number of medals (gold, silver, and bronze) that a country won in a given sport in a given Olympiad. The data on medal performance come from the IOC.

Turning to the independent variables, to determine whether there is persistence within sports across countries in Olympic medal performance, we include dynamics—i.e., measures of past medal performance in a given sport—to gauge the impact of past performance on current performance. Following prior literature, we include performance lags of *Medal_total* for the three prior Olympiads. For example, $Medal\_total_{(t-3)}$ is the number of medals that a country has won in that sport three Olympiads prior to the current one.[4]

A positive coefficient on the past performance variables would indicate that countries that have won medals in a given sport in prior games are more likely to win medals in future games. We expect some persistence of this sort due to country/individual characteristics conducive to certain sports, country/individual investment and specialization in given sports, and/or preferential treatment received by individuals/teams that allow them to dominate for long periods of time. That said however, we generally expect the effects to erode with increasing lags. An eroding dynamic effect suggests that one country (or set of countries) does not uniformly dominate the medal podium in a given sport.

In addition to the variables described above, we include a number of control variables that have

[4] When a nation boycotts an Olympiad, we interpolate past medal performance in a given year as the average medal total over the prior three Olympiads in which the country had participated. As an alternative, we replaced interpolated values with zero. Results did not vary. Results also did not vary when using different lag durations.

the potential to impact medal performance. We control for overall population (*Ln_population*) and gross domestic product (*Ln_GDP*, expressed in nominal U.S. dollars). The size of the population determines the pool from which athletic talent can be drawn, and the wealth of a country influences the resources that countries have to invest in sport (e.g., Bernard and Busse 2004, Johnson and Ali 2004). Not surprisingly, GDP and population have been found to be the best predictors of success at the Olympic Games (e.g., Bernard and Busse 2004, Johnson and Ali 2004, Moosa and Smith 2004, Novikov and Maksimenko 1973). Given the length of the period under study, we had to consult a number of sources for GDP and population data. For population data from 1950 (inclusive), we used the United States Census Bureau (2008) International Database. For any missing data in that period, we referred to the United Nations Department of Economic and Social Affairs (UN-DESA) (2008) database. Population data from before 1950 was collected primarily from the International Historical Statistics (Mitchell 1992, 1995, 2003) and the Global Financial Database (2008). For GDP we were able to collect data for the 1972–2006 period from the United Nations Statistics Division (2008). Data for the period 1960–1968 came from the World Bank (2008) World Development Indicators (WDI) Database. For GDP data prior to 1960, we consulted the International Historical Statistics database.

Along with GDP and population, we control for whether the country acts as the host nation. We define *Host_nation* as a dummy variable that captures whether the country hosted the games. Studies demonstrate that host countries are more likely to win medals than nonhost countries (see Bernard and Busse 2004 for a review). Turning to more game-specific variables, we control for whether a particular Olympiad was subject to a significant boycott. We define *Boycott* as one if the Olympiad in question was boycotted by a significant number of countries, and zero otherwise. The Summer Games of 1976 in Montreal, 1980 in Moscow, and 1984 in Los Angeles were major boycott games. We expect that the absence of competitors due to boycott will make it more likely for participating countries to medal in any given sport.

We also control for the number of nations participating in a given sport at a given Olympic Games (*Participating_nations*) and the number of events per sport (*Events_per_sport*). On average, we expect that the greater the number of participants, the smaller the statistical probability of any single country winning a medal in a given sport. At the same time, a larger number of events per sport increases the chances of any single country winning a medal. Finally, we control for season. *Season* is a dummy variable that takes the value of one for the Summer Games and zero otherwise. The Summer Games and Winter Games are differentiated because they attract a different set of participating countries (not just in composition, but also in total numbers) and involve sports that require different skills and resources.

4.2. Methods

The dependent variable in this study is a count measure (*Medal_total*) that takes only nonnegative, integer values. For dependent variables of this sort, scholars generally recommend a Poisson regression. However, as Cameron and Trivedi (1998) point out, the Poisson regression is sensitive to its underlying assumptions, and we found evidence of overdispersion in our sample. We therefore present results using a negative binomial regression.

Given the panel nature of our data, it is possible for a systematic unobserved component to be embedded in the error term leading to serial correlation. Either fixed or random effects can be used to account for unobserved heterogeneity of this sort (see Greene 2007). We opt for country random effects in our specification because some of the countries in our sample exhibit little variance in the dependent variable. Some countries have never won a medal in any sport, and many have competed, but not won, in certain sports.

As mentioned previously, we are interested in persistence in performance at the Olympic Games. We therefore include dynamics in our regression specifications (see Al-Osh and Alzaid 1987, Alzaid and Al-Osh 1990, Brännäs and Hellström 2001). We add lags of the dependent variable as explanatory variables to examine whether and, if so, how much, persistence exists in performance across Olympiads. The regression takes the following functional form:

$$Medal\_total_{ijt} = \exp(\gamma\, Medal\_total_{ij(t-k)} + \beta X_{ijt} + u_j + \varepsilon_{ijt}), \qquad (1)$$

where $Medal\_total_{ijt}$ reflects the observed number of medals won by country i in sport j at Olympiad t; γ is a vector of persistence parameters for *Medal_total*, where k varies from 1 to 3 (i.e., we include *Medal_total* lags of up to three Olympiads); X_{ijt} represents a matrix of control variables for country i in sport j at Olympiad t; β is a vector of parameter estimates; u_j captures the country-specific disturbance; and ε_{ijt} is an independent error term with a gamma distribution.

Because we wish to examine whether persistence is more prevalent in objective, refereed, or subjective sports, we split the sample by sport category to examine the structural stability of the persistence parameters. Greater persistence in subjective sports would lend credence to the claim that reputation influences

Table 2 Country-Sport Descriptive Statistics and Correlations ($N = 18{,}046$)

	1	2	3	4	5	6	7	8	9	10	11
1. *Medal_total*$_{(t)}$	1.00										
2. *Medal_total*$_{(t-1)}$	0.77	1.00									
3. *Medal_total*$_{(t-2)}$	0.68	0.75	1.00								
4. *Medal_total*$_{(t-3)}$	0.60	0.65	0.73	1.00							
5. *Boycott*	0.00	0.01	0.00	−0.01	1.00						
6. *Events_per_sport*	0.08	0.08	0.07	0.07	−0.01	1.00					
7. *Participating_nations*	−0.04	−0.02	−0.02	−0.01	−0.06	0.79	1.00				
8. *Season*	−0.01	0.00	0.00	0.01	−0.01	0.21	0.27	1.00			
9. *Host_nation*	0.13	0.06	0.06	0.06	0.00	−0.08	−0.09	−0.03	1.00		
10. *Ln_population*	0.21	0.20	0.18	0.17	0.00	−0.17	−0.20	−0.06	0.11	1.00	
11. *Ln_GDP*	0.14	0.15	0.15	0.15	0.16	−0.05	0.00	−0.16	0.10	0.60	1.00
Mean	0.72	0.63	0.56	0.50	0.22	11.29	51.40	0.84	0.03	16.35	23.78
SD	2.08	1.99	1.92	1.84	0.42	8.23	39.22	0.37	0.17	1.82	2.62
Min	0.00	0.00	0.00	0.00	0.00	1.00	3.00	0.00	0.00	9.29	11.18
Max	52.00	52.00	52.00	43.00	1.00	38.00	193.00	1.00	1.00	22.14	30.21

the results. We therefore split the pooled sample from Equation (1) into three subgroups: objective sports, refereed sports, subjective sports:

$$Medal\_total_{ijt} = \exp(\gamma_1 Medal\_total_{ij(t-k)} + \beta X_{ijt} + u_j + \varepsilon_{ijt}), \quad \forall Objective\_sports, \quad (2a)$$

$$Medal\_total_{ijt} = \exp(\gamma_2 Medal\_total_{ij(t-k)} + \beta X_{ijt} + u_j + \varepsilon_{ijt}), \quad \forall Refereed\_sports, \quad (2b)$$

$$Medal\_total_{ijt} = \exp(\gamma_3 Medal\_total_{ij(t-k)} + \beta X_{ijt} + u_j + \varepsilon_{ijt}), \quad \forall Subjective\_sports, \quad (2c)$$

where we are interested in exploring the relative impact of the persistence parameters γ_1, γ_2, and γ_3 across Equations (2a), (2b), and (2c), respectively. A finding whereby $\gamma_3 > \gamma_1$ and/or $\gamma_2 > \gamma_1$ would indicate that there is greater persistence in performance in more subjective sports (i.e., countries are more dominant in more subjective than objective sports). This would be consistent with our priors.[5]

4.3. Results

Table 2 presents descriptive statistics and product moment correlations for the full sample. Correlations are generally as expected, and there is evidence that prior winners are more likely to win again. Winning medals in a given sport in previous Olympiads is strongly associated with winning medals in the same sport at the current games. The relationship weakens over time, with medal totals in the most recent games linked (0.77) to a greater likelihood of winning a medal in the current games, whereas medal totals three games previously is more weakly linked (0.60) with medal totals at the current games. This suggest that, although persistence exist, countries do not uniformly dominate over time.

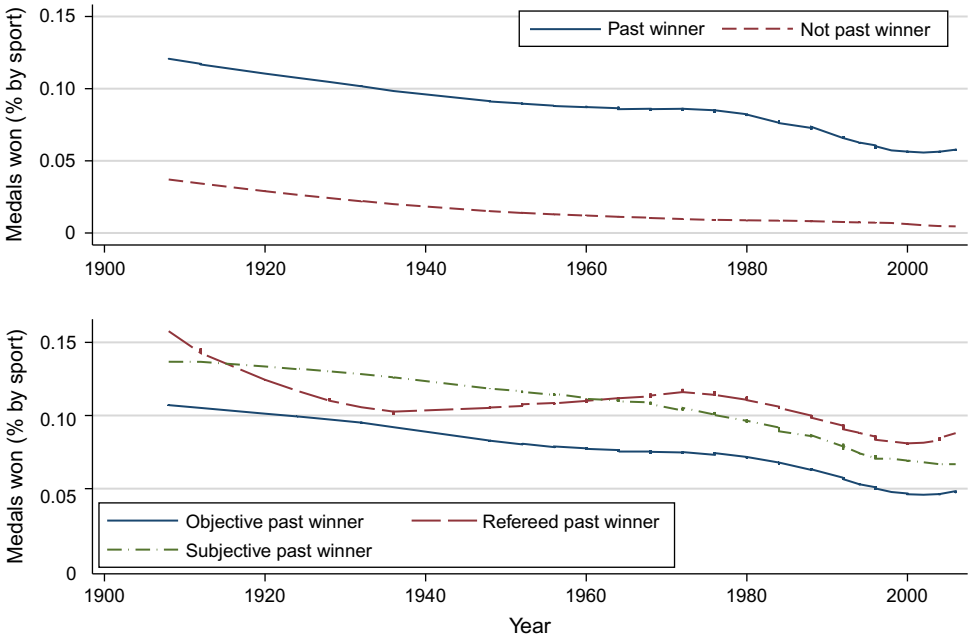
Digging a bit deeper, Figure 1 presents descriptive evidence consistent with persistence and with differences in persistence across sports categories. The top panel demonstrates that prior winners are more likely to win medals in subsequent games. Although there is a general secular trend favoring less dominance in any given sport over time, the performance gap between prior winners and nonwinners narrows only slightly. The bottom panel of Figure 1 disaggregates prior winners into objective, subjective, and refereed sports. The evidence suggests that there is greater persistence in subjective/refereed Olympic contests than objective Olympic contests.

Although the correlations and descriptives are interesting in their own right, it would be premature to conclude that persistence is prevalent or that significant persistence differences exist across sporting types because the results presented in Table 2 and Figure 1 do not control for other intervening factors that have the potential to simultaneously influence the independent and dependent variables. We therefore turn to the multivariate regression in Table 3 to better understand the nature of persistence across objective, refereed, and subjective sports.

The random effects negative binomial regression results appear in Table 3. Column 1 presents results

[5] Although we could have introduced a multiplicative interaction term to estimate the moderating impact of sporting category (i.e., *Medal_total* ∗ *Sport_category*), we prefer subsample splits. This is because multiple category interactions complicate inference, especially in nonlinear regression (Berrington de Gonzalez and Cox 2007, Jaccard et al. 1990). In such situations, testing for moderation through subsample splits is the preferred approach (Jaccard et al. 1990).

Figure 1 **(Color online) Smoothed Scatterplot of Country-Sport-Level Persistence Over Time**



Notes. The vertical axis records the average proportion of medals in a sport won by all countries. In the top panel, countries are split into prior medal winners (past three Olympiads) and nonprior-medal winners. The bottom panel divides prior medals winners into objective, refereed, and subjective categories.

for the full sample. Columns 2, 3, and 4 present results for the objective, refereed, and subjective sport subsamples, as described in Equations (2a), (2b), and (2c), respectively.[6] Beginning with the control variables and consistent with the extant literature (e.g., Ball 1972, Bernard and Busse 2004), we find that *Ln_population* and *Host_nation* are positive and statistically significant ($p < 0.001$) across all conditions. Similarly, we find results in line with expectations for *Events_per_sports* and *Participating_nations*. Results for the other independent variables are less consistent. For example, we find results that match with priors across the refereed and subjective conditions for *Boycott*, but not in the objective condition. Finally, we find that *Ln_GDP* has a positive and statistically significant effect for objective sports, but it does not for refereed and subjective sports. Although somewhat unexpected, this could be due to elevated correlations between *GDP* and *Ln_population* ($r = 0.59$), the fact that country-specific effects are picking up some of the variance due to *GDP*, and/or that once accounting for previous performance, *GDP* has little impact on medal performance.[7]

Turning to the coefficients and marginal effects for the lagged performance variables, we find that past performance positively influences current performance in the full sample and across all subsamples. Not surprisingly, the effects are strongest in the near term. We note that we even find a negative and significant relationship between *Medal_total* in year $(t-3)$ and *Medal_total* at time (t). This suggests that certain countries enjoy a temporal performance advantage, but that advantage is not all encompassing. There is some evidence that sport-specific advantage comes with an expiration date; countries that won medals in a given sport 12 years prior are less likely to win medals in the current games. This is indicative of mean reversion and/or general convergence between weaker and stronger competitors over time.

With respect to the results across sport type, we find significant differences in persistence, with the greatest persistence in more subjective sports. Comparing betas across subsamples using standardized z-scores indicates that the coefficients for the one-year lag $(t-1)$ in the subjective and refereed conditions are each statistically different ($p < 0.01$) from the objective sport condition. The coefficients across subjective and refereed sports do not significantly differ. The same pattern holds for the two $(t-2)$ and three $(t-3)$ year lags—i.e., the coefficients for the subjective and refereed sport subsamples do not statistically differ from

[6] We also ran similar analyses for two other subsamples: objective plus refereed sports and subjective plus refereed sports. The findings were consistent with the patterns presented in Table 3.

[7] In results not reported, we explored several variants of the models presented herein. Indeed, we find that, in models in which we omit *Ln_population* and those in which we do not include a country-specific random effect, *GDP* is positive and significant, consistent with prior findings.

Table 3 Negative Binomial Regression of $Medal\_total_{(t)}$ by Country, Sport, and Olympiad with Country-Sport Random Effects

		Subsamples by sport type		
	Full sample	Objective sports	Refereed sports	Subjective sports
Constant	−6.018**	−6.258**	−4.653**	−6.405**
	(−17.77)	(−15.07)	(−5.95)	(−6.51)
Boycott	−0.164**	−0.217**	−0.090	−0.113
	(−5.21)	(−5.60)	(−1.29)	(−1.15)
	[−0.058]	[−0.079]	[−0.026]	[−0.050]
Events_per_sport	0.042**	0.044**	0.077**	0.014
	(11.28)	(10.30)	(7.29)	(0.87)
	[0.015]	[0.017]	[0.023]	[0.006]
Participating_nations	−0.006**	−0.007**	−0.020**	−0.001
	(−9.05)	(−8.84)	(−7.33)	(−0.26)
	[−0.002]	[−0.003]	[−0.006]	[−0.000]
Season	−0.161+	−0.067	0.006	−0.617*
	(−1.77)	(−0.62)	(0.02)	(−2.42)
	[−0.062]	[−0.026]	[0.002]	[−0.345]
Ln_population	0.379**	0.355**	0.387**	0.433**
	(17.85)	(13.48)	(8.55)	(6.98)
	[0.139]	[0.137]	[0.117]	[0.198]
Ln_GDP	0.007	0.030**	−0.023+	−0.022
	(1.13)	(3.54)	(−1.92)	(−1.22)
	[0.003]	[0.012]	[−0.007]	[−0.010]
Host_nation	0.571**	0.541**	0.419**	0.751**
	(13.63)	(10.49)	(4.60)	(6.26)
	[0.277]	[0.272]	[0.154]	[0.497]
$Medal\_total_{(t-1)}$	0.048**	0.044**	0.090**	0.063**
	(14.51)	(11.68)	(9.35)	(5.14)
	[0.018]	[0.017]	[0.027]	[0.029]
$Medal\_total_{(t-2)}$	0.009*	0.005	0.033**	0.027*
	(2.19)	(1.02)	(3.07)	(2.27)
	[0.003]	[0.002]	[0.010]	[0.012]
$Medal\_total_{(t-3)}$	−0.008*	−0.008+	0.016	−0.011
	(−2.18)	(−1.96)	(1.45)	(−1.11)
	[−0.003]	[−0.003]	[0.005]	[−0.005]
Country-sport random effects	Y	Y	Y	Y
N	18,046	12,419	3,576	2,051
Log-likelihood	−14,482.212	−9,915.74	−2,734.18	−1,738.32
LR chi square	1,568.74**	1,135.62**	379.27**	219.97**

Note. The *t*-statistics appear in parentheses and the marginal effects appear in brackets.
$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$ (two-tailed test).

each other, but each is statistically different from the coefficient in the objective sport subsample ($p < 0.01$).[8]

Differences across each sporting subsample can best be seen in the marginal effects. Specifically, each additional medal won in a given sport at the prior Olympic Games increases the number of medals won in that sport in the current Olympic Games by about 0.017 medals in the objective case, 0.027 medals in the refereed case, and 0.029 medals in the subjective case. Although a 0.012 medal bump from objective to subjective sports seems small, it represents an average return to reputation of roughly 71%. Not only are the marginal effects stronger in the case of subjective and refereed sports for the $(t-1)$ Olympiad, but the number of medals won in a given sport at Olympiad $(t-2)$ is statistically significant only in the refereed and subjective conditions. The relationship between winning past medals in an Olympiad 12 years prior $(t-3)$ even turns negative and significant in the objective sport subsample. The cumulative effect across objective and subjective sports is therefore larger than the single Olympiad $(t-1)$ effect.[9]

Taken together, the results indicate that there is greater persistence in performance in subjective/refereed sports than objective sports. This points to a reputational bias whereby previous winners are likely to receive favorable treatment from officiates, even greater than might otherwise be warranted. Otherwise stated, because results in more subjective sports are to some extent ambiguous, countries that have won in the past are treated differently. This represents returns to reputation that can, on the margin, influence Olympic outcomes.

We are aware of one serious potential caveat with respect to the interpretation of our findings. Specifically, if systematic heterogeneity exists across sports such that there is greater durability in skill differential in subjective and refereed sports than objective sports, then we would expect naturally greater persistence in subjective and refereed sports. However, we believe that the findings for the *Host_nation* variable help corroborate our inference over such an alternative interpretation. The results for the *Host_nation* independent variable indicate that the host country wins, on average, 82% more medals (based on marginal effects) in subjective sports than in objective sports. Again, this points to a bias in subjective sports, whereby judges give preferential treatment to athletes from the host nation.[10]

Given that referees, in contrast with judges, are not empowered with determining results, the finding that

[8] We note that testing for statistical significance across coefficients in nonlinear models is complicated by the underlying distribution and the range of observations over which the comparisons are made. We therefore turn to marginal effects, calculated at the mean, to draw stronger inferences.

[9] In an unreported pooled model with interactions between sport type and lagged medals won, we confirm that the coefficients for refereed and subjective lagged medals are jointly significantly different from the coefficients for objective lagged medals. In further unreported models, we tried a number of alternative specifications: one period lags $(t-1$ only), combined two period lags $(t-1+t-2)$, combined three period lags $(t-1+t-2+t-3)$, and dropping nations without prior appearances in the sport. In all cases, the results were similar to those presented in Table 5.

[10] We return to the issue of skill durability in the sensitivity and robustness section of the online appendix.

persistence is roughly equivalent in refereed and subjective sports is somewhat surprising. The findings hint at a potential bias in referee decisions because referees might similarly be concerned with accountability and/or allow unobserved quality to influence their rulings. We note, however, that whether this equivalence is driven by unobserved heterogeneity factors, such as variance in durable skills or the ability of referees to act unilaterally, is not fully addressable in the cross-sport analysis.

In the following section we, therefore, present an additional set of results that provide further corroboratory evidence to support our claims. Specifically, we examine Olympic outcomes in one sport: boxing. Boxing is an interesting case because the results of individual matches can reflect either objective or subjective outcomes. As a result, analyzing outcomes of boxing matches can help address various concerns about our inferences.

5. Olympic Boxing Match Analysis

Boxing first appeared in the modern games at the 1904 St. Louis Olympics, and by 1920 the Fédération Internationale de Boxe Amateur (FIBA) was formed to govern international competition in the sport. In 1946 the International Boxing Association (AIBA) replaced FIBA as a result of controversial behavior by FIBA officials during World War II. Since that time, Olympic boxing has been governed by the AIBA (AIBA 2010).

An Olympic boxing tournament under the AIBA works as follows. Single elimination tournaments pair off competitors in each weight classification. The number of weight categories between 1948 and 2004 has ranged from eight to 12, with no more than one category added or dropped between successive games. Each country can enter one boxer per weight class, and boxers qualify via regional and global amateur competitions. Matches are seeded randomly, and only the winner advances. The winner of the final match takes the gold medal, the loser the silver medal, and since 1950, the loser in each semifinal contest is awarded a bronze medal.

Each match consists of four rounds of two minutes apiece, and there are four ways to win. First, the opponent can be disqualified (DQ) for things such as failing to appear, failure to continue boxing, or failure to make weight. Second, the referee can stop the fight if one contestant is injured or dangerously outclassed. This result is commonly referred to as a technical knockout (TKO).[11] Third, a contestant wins by a knockout (KO) if his opponent is knocked to the mat and does not resume boxing within 10 seconds. Fourth, a boxer wins on points if none of the other outcomes occurs and a majority of the five judges deem him the winner based on the total number of punches landed. Prior to 1992, each judge tallied his own scorecard, but since that time the AIBA has used a computerized system where judges award points for each blow and a majority of judges must score the same hit within one second of each other for it to count.

5.1. Data

We collected data on the outcomes for individual Olympic boxing matches from 1948, the year the AIBA began running the contests, through 2004. These data were collected, recorded, and published by Polish boxing enthusiast Janusz Majcher (see Majcher 2010). From those data, we computed country medal counts for the prior three Olympiads (as in the country-sport sample), reducing the usable sample range to 1960 through 2004. The analysis parallels the country- and sport-level analysis presented above, with a few exceptions due to the nature of the data. For example, because we have data for each match, and not just the medal outcomes, the unit of analysis in the boxing regressions is the observed outcome (win/lose by KO/TKO or win/lose on points) for each individual competitor from a given country in each match. The analysis from Table 3, by contrast, is at the country-sport level, where the dependent variable reflects the total number of medals won by a given country in a given sport. Because the boxing dependent variable does not reflect the number of medals won in a given sport but is a multioutcome indicator capturing whether the focal country's boxing match ended in a loss, a win by KO/TKO, or a win on points, we match the independent variable to the dependent variable for consistency. We use lagged indicators for whether the focal boxer's home country won a boxing medal in the prior three Olympiads rather than a count of total medals as a predictor of match outcome.[12]

Table A3 in the online appendix presents details about the boxing matches included in our sample. There were 3,786 boxing matches scheduled between 1960 and 2004. From the initial set of matches, we discarded the 199 matches where the win was the result of a disqualification.[13] The remaining 3,587 matches yielded 7,174 observations (3,578 matches $\times$ 2 competitors each). For eight boxers, we were unable to locate

[11] In modern parlance, a TKO is known as a referee stops contest (RSC). In our data, the RSC designation first appears in 1976. For the sake of consistency, we use the TKO label throughout the manuscript.

[12] Results using previous medal counts instead of a medal indicator variable generated similar results.

[13] Most of the DQs occurred in 1976 when a contingent of African nations boycotted the games after the first round matches had been scheduled.

Table 4 Boxing Descriptive Statistics and Correlation ($N = 7{,}166$)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. *Winner*	1.00													
2. *Win: KO/TKO*	0.37	1.00												
3. *Win: points*	0.78	−0.29	1.00											
4. *Medals t − 1* (1 = Yes)	0.23	0.10	0.17	1.00										
5. *Medals t − 2* (1 = Yes)	0.18	0.05	0.15	0.53	1.00									
6. *Medals t − 3* (1 = Yes)	0.16	0.04	0.14	0.43	0.57	1.00								
7. *Boycott*	0.00	0.07	−0.04	−0.04	−0.03	−0.06	1.00							
8. *Participating_nations*	0.00	0.00	0.00	−0.05	−0.06	−0.04	0.10	1.00						
9. *Host_nation*	0.08	0.02	0.07	0.10	0.12	0.19	0.04	−0.02	1.00					
10. *Ln_population*	0.15	0.04	0.13	0.34	0.35	0.32	−0.10	−0.03	0.19	1.00				
11. *Ln_GDP*	0.13	0.02	0.12	0.24	0.28	0.30	0.04	0.17	0.22	0.66	1.00			
12. *Cuba/USA/USSR*	0.22	0.14	0.13	0.71	0.69	0.63	−0.04	−0.05	0.15	0.37	0.26	1.00		
13. *Headgear*	0.00	−0.02	0.01	−0.04	−0.02	−0.01	0.08	0.72	−0.02	0.02	0.31	−0.03	1.00	
14. *Electronic_scoring*	0.00	−0.05	0.03	0.03	0.00	0.04	−0.42	0.25	−0.03	0.10	0.23	0.04	0.53	1.00
Mean	0.5	0.12	0.38	0.52	0.47	0.42	0.35	74.5	0.04	16.94	24.3	0.15	0.55	0.25
SD	0.5	0.32	0.49	0.50	0.50	0.49	0.48	17.21	0.2	1.49	2.37	0.36	0.5	0.43
Min	0	0	0	0	0	0	0	49	0	9.79	16.49	0	0	0
Max	1	1	1	1	1	1	1	105	1	20.98	30.09	1	1	1

country-level economic data, thereby reducing the analysis to a sample of 7,166 individual boxer/match observations. In all subsequent analyses, we treat KOs and TKOs as functionally equivalent because boxing experts informed us that both are objective indicators of injury and/or drastic inferiority.

Table 4 presents descriptive statistics and product moment correlations for the boxing subsample. Relative rates for each of the decisions—objective wins via KO or TKO (*Win*: *KO/TKO*) and subjective decisions awarded by the judges (*Win*: *points*)—can be inferred from the data. The omitted reference category is that the competitor lost the match.

Most of the control variables are as in the prior analysis, but we add several new controls specific to this setting. Specifically, we include *Weight_class* fixed effects and *Round* fixed effects.[14] *Headgear* is a dummy variable capturing the introduction of mandatory headgear in 1984, while *Electronic_scoring* captures the introduction of a computerized scoring system in 1996. Finally, and after speaking with various boxing experts, we decided to include a control for the traditional power countries in boxing (Cuba, USA, and USSR/Russia). These three nations have historically dominated Olympic boxing, and aside from boycott years, have never failed to win boxing medals. We therefore define *Cuba/USA/USSR* as a dummy capturing boxers from any of these respective nations, and zero otherwise. In addition to the nation-specific controls described above, we cluster standard errors by country to account for nation-specific heterogeneity in boxing outcomes.

[14] The round fixed effects account for the specific round in which the match concluded.

5.2. Methods

As in the results presented in Table 3, we are interested in measuring the influence of performance in previous Olympiads on current boxing match outcomes. However, unlike in Table 3, the dependent variable has multiple categories. Our dependent variable of interest in the boxing analysis is whether the focal boxer won or lost the match (in contrast to the number of medals the nation won in a given sport). We are not only interested in whether the boxer in question has won the individual match, but in how the boxer won the match—whether through KO/TKO or by decision (i.e., points). The dependent variable therefore has multiple possible outcomes (lose, win by KO/TKO, or win on points).

Given that the dependent variable of interest has multiple outcomes that are bound by an additive constraint (the probabilities of observing a loss, a win by KO/TKO, or a win by points must equal one), we specify a multinomial logit (Wooldridge 2002). Equation (3) illustrates the functional form of the multinomial specification:

$$\ln \frac{P(Y_{ijt} = m)}{P(Y_{ijt} = 1)} = \alpha_{mijt} + \sum_{m=1}^{M} (\gamma Medal_{ij(t-k)} + \beta X_{ijt}), \quad (3)$$

where m reflects the probability of observing a particular outcome for country i in boxing match j at Olympiad t, γ is a vector of parameters for each of the lagged boxing $Medal_{ij(t-k)}$ indicators, and X_{ijt} is a matrix of independent variables for country i in boxing match j at Olympiad t. When $m = 2$, the multinomial logit is equivalent to a logistic regression with a dichotomous dependent variable. When there are more than two outcomes as in our case (i.e., $m > 2$), the multinomial logit from Equation (3) calculates

$m-1$ log odds ratios interpretable relative to the base outcome. The base outcome in our model is the likelihood of observing a loss in a particular boxing match.

As before, we wish to examine whether persistence is more prevalent for objective or subjective boxing outcomes. We therefore split the sample by outcome to examine the structural stability of the parameters:

$$\ln\frac{P(Y_{ijt}=2)}{P(Y_{ijt}=1)} = \alpha_{2mijt} + \sum_{m=1}^{M}(\gamma_2 Medal_{ij(t-k)} + \beta_2 X_{ijt}), \quad \forall Objective\_outcomes\ (KO/TKO); \quad (4a)$$

$$\ln\frac{P(Y_{ijt}=3)}{P(Y_{ijt}=1)} = \alpha_{3mijt} + \sum_{m=1}^{M}(\gamma_3 Medal_{ij(t-k)} + \beta_3 X_{ijt}), \quad \forall Subjective\_outcomes\ (points\ win). \quad (4b)$$

This setup allows us to both reexamine our cross-sport findings in an individual sport (boxing) and examine outcomes in a context where there are various possible outcomes for each observation—the individual boxing match.

5.3. Results

Table 5 reports results for our sample of Olympic boxing matches. Again, all standard errors are clustered by country to account for country-specific heterogeneity.

In the first column, we present outcomes from a simple logit, where the dependent variable simply captures whether the match ended in a win or a loss. The results indicate that there is a substantial amount of persistence in boxing outcomes. That is, boxers from countries that have won medals in the past are more likely to win the focal match. As with the cross-sport results, the persistence erodes over time, disappearing by Olympiad ($t-3$).

In the second and third columns, we compare results across victory type. The results from columns 2 and 3 indicate that the persistence observed in column 1 is largely driven by subjective outcomes. There is stronger persistence for the subjective outcome, win on points (*subjective win*: *points*), than the objective outcome (*objective win*: *KO/TKO*). Likewise, the *Host_nation* country effect is stronger for subjectively determined outcomes. As before, our interpretation of these results is that past performance is, in part, a proxy for greater ability, development, and/or preferential ex ante treatment. However, there is a greater persistence due to reputational advantages for the subjective outcomes. Consistent with our underlying theory, quality standards are insulated from market discipline when quality is subjective and are thus reflective of a reputational bias when determining outcomes.[15]

Table 5 Logit and Multinomial Logit Estimates of Boxing Match Victory with Country Clustered Standards Errors

	Logit	Multinomial logit	
	Winner	Objective win (KO/TKO)	Subjective win (points)
Boycott	0.024	0.419**	−0.099
	(0.27)	(3.88)	(−1.10)
Participating_nations	0.002	0.004	0.002
	(0.85)	(1.14)	(0.57)
Ln_population	0.047	0.042	0.051
	(1.21)	(0.85)	(1.29)
Ln_GDP	0.015	−0.025	0.028
	(0.52)	(−0.76)	(0.96)
Cuba/USA/USSR	0.965**	1.581**	0.772**
	(4.29)	(6.10)	(3.61)
Headgear	0.040	−0.055	0.064
	(0.38)	(−0.38)	(0.57)
Electronic_scoring	−0.077	−0.217	−0.047
	(−0.64)	(−1.28)	(−0.39)
Host_nation	0.450**	0.239	0.503**
	(2.78)	(1.60)	(2.91)
$Medals_{(t-1)}$ (1 = Yes)	0.634**	0.844**	0.573**
	(7.35)	(7.22)	(6.67)
$Medals_{(t-2)}$ (1 = Yes)	0.160*	−0.022	0.208*
	(2.05)	(−0.21)	(2.47)
$Medals_{(t-3)}$ (1 = Yes)	0.128	−0.021	0.168*
	(1.62)	(−0.19)	(1.97)
Weight class dummies	Y	Y	Y
Round dummies	Y	Y	Y
N	7,166	7,166	
Psuedo R^2	0.067	0.064	
Log-likelihood	−4,631.734	−6,492.561	

Note. *t*-Statistics appear in parentheses.
*$p < 0.05$; **$p < 0.01$ (two-tailed test).

6. Robustness and Sensitivity

To assess the sensitivity and robustness of the findings presented herein, we explored variants of the models and considered the plausibility of alternative explanations. Details can be found in the online appendix.

[15] Coefficient tests confirm that, in the multinomial logit model reported in Table 5, medals ($t-1$) is significantly larger (1%) for the objective outcome. However, for the subjective outcome, medals ($t-2$), medals ($t-3$), and host are significantly larger (5%, 10%, and 10% significance levels, respectively). In unreported models, we tested a number of alternative specifications: one period lags ($t-1$ only), combined three period lags ($t-1+t-2+t-3$), using lagged counts instead of dummy variables, using lagged counts summed over three periods, dropping nations without recent prior appearances in the sport, replacing interpolated values for boycott years with zero, and dropping boxers from Cuba/USA/Russia. In all cases, the results were similar to those presented in Table 5.

7. Discussion and Conclusion

In this study, we examine whether performance gaps are more persistent in subjective competitions. Toward that end, we examined whether there was greater persistence in medal performance in Olympic sports where outcomes are determined more subjectively. We then isolated boxing as a special case to determine whether there was greater persistence in observed performance for subjectively, versus objectively, determined boxing outcomes.

We found that superior country-level performance in Olympic sports is persistent even in objective events, but the persistence is greater when performance includes a subjective component. This suggests that judges pay attention to reputational cues when determining outcomes. Indeed, interviews the authors conducted with Olympic officials corroborate these findings. For example, in an interview conducted with a boxing official from the Beijing Olympics, the official noted how he is aware that there is a perception that certain countries receive unfairly favorable treatment from boxing judges. The official mentioned that it was refreshing to see results that empirically substantiate priors regarding unfair Olympic judging practices. The official even suggested that the results might be used to educate judges about the dangers of overreliance on reputation. Additionally, an Olympic ice skating coach with whom we spoke suggested that it is a widely shared belief among participants and coaches that judges unfairly favor athletes from certain countries. In practical terms, the results in this paper suggest that rooting out reputational bias will remain a challenge.

We believe that there are three promising areas for future inquiry related to the idea of subjective reputational advantage. First, and reflective of the robustness issues discussed in the online appendix, the scope conditions and effect sizes associated with ex post advantages could be more precise. This study utilized aggregate data over an expansive time frame to demonstrate that reputational bias is present. It seems reasonable that environmental changes such as professionalization, standardization, and geopolitical discontinuities could influence the results. Moreover, they are likely better observed with more granular data. We would therefore welcome additional research using more disaggregated data to help refine and extend this work.

Second, we believe that our results support the inference that reputations influence evaluators after the fact (i.e., after the product is consumed). However, it is not entirely clear whether subjective reputational advantage derives primarily from an evaluator who infers quality that they did not, in fact, observe or simply from the justifiability that is afforded officiates when they favor previous winners. Studies that can tease apart the precise mechanisms in a competitive context would certainly further our understanding of reputational bias dynamics.

Third, although we demonstrate the effect of ex post reputational bias on persistent performance in an Olympic context, it would be interesting to see how these findings extend to competitive business environments. That is, it would be interesting to see whether and how reputations influence competitive advantage, and the sustainability of competitive advantages, in industries where performance criteria are determined more subjectively ex post versus industries in which performance criteria are better defined ex ante.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2014.2144.

Acknowledgments

The authors gratefully acknowledge research support from Jin-Hyun Bae. Seminar participants at the University of Chicago, Aalborg University, Universidad Carlos III de Madrid, the Wharton School of the University of Pennsylvania, University of Michigan, National University of Singapore, Drexel University, London Business School, New York University, and University of Maryland provided valuable input. The authors also thank Jóhanna K. Birnir, Brent Goldfarb, Adam Grant, and Olav Sorenson.

References

AIBA (2010) Last accessed October 2010, http://www.aiba.org/.

Al-Osh M, Alzaid AA (1987) First-order integer-valued autoregressive (INAR(1)) processes. *J. Time Ser. Anal.* 8(3):261–275.

Alzaid AA, Al-Osh M (1990) An integer-valued *p*th-order autoregressive structure (INAR(p)) process. *J. Appl. Probab.* 27(2): 314–324.

Arvey RD, Murphy KR (1998) Performance evaluation in work settings. *Annual Rev. Psych.* 49(1):141–168.

Ball DW (1972) Olympic games competition: Structural correlates of national success. *Internat. J. Comparative Sociology* 12(3–4): 186–200.

Bernard AB, Busse MR (2004) Who wins the Olympic Games: Economic resources and medal totals. *Rev. Econom. Statist.* 86(1):413–417.

Berrington de Gonzalez A, Cox DR (2007) Interpretation of interaction: A review. *Ann. Appl. Statist.* 1(2):371–385.

Bommer WH, Johnson JL, Rich GA, Podsakoff PM, Mackenzie SB (1995) On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psych.* 48(3):587–605.

Brännäs K, Hellström J (2001) Generalized integer-valued autoregression. *Econometric Rev.* 20(4):425–443.

Cameron AC, Trivedi PK (1998) *Regression Analysis of Count Data* (Cambridge University Press, New York).

Campbell JP (1990) Modeling the performance prediction problem in industrial and organizational psychology. Dunnette MD, Hough LM, eds. *Handbook of Industrial and Organizational Psychology*, 2nd ed., Vol. 1 (Consulting Psychologists Press, Palo Alto, CA), 687–732.

Correll SJ, Ridgeway CL, Zuckerman EW, Bloch S, Jank S (2012) It's the conventional thought that counts: How third-order inference produces status advantage. Working paper, Stanford University, Stanford, CA.

Darby MR, Karni E (1973) Free competition and the optimal amount of fraud. *J. Law Econom.* 16(1):67–88.

Emerson JW, Seltzer M, Lin D (2009) Assessing judging bias: An example from the 2000 Olympic Games. *Amer. Statistician* 63(2):124–131.

Fédération Internationale de Natation (2010) Beijing 2008—Swimming qualifying procedures. Last accessed October 2010, http://web.archive.org/web/20070304105903/http://www.fina.org/events/OG/Beijing_2008/pdf/qualifyingprocedures_sw.pdf.

FIBA (2012) 2012 Olympic Games basketball tournament—Qualifying process. Last accessed June 2014, http://www.fiba.com/pages/eng/fe/12/olym/who_is_in/p/men.html.

Foschi M (1996) Double standards in the evaluation of men and women. *Social Psych. Quart.* 59(3):237–254.

Garicano L, Palacios-Huerta I, Prendergast C (2005) Favoritism under social pressure. *Rev. Econom. Statist.* 87(2):208–216.

Global Financial Database (2008) Last accessed July 2008, https://www.globalfinancialdata.com/.

Gould RV (2002) The origins of status hierarchies: A formal theory and empirical test. *Amer. J. Sociology* 107(5):1143–1178.

Greene WH (2007) *Econometric Analysis* (Prentice-Hall, Upper Saddle River, NJ).

Guttman, A (1994) *The Olympics: A History of the Modern Games* (University of Illinois Press, Chicago).

Hong H, Kubik JD (2003) Analyzing the analysts: Career concerns and biased earnings forecasts. *J. Finance* 58(1):313–351.

International Olympic Committee (IOC) (2008) Website medal database. Last accessed May 2008, http://www.olympic.org/uk/athletes/results/search_r_uk.asp.

International Olympic Committee (IOC) (2010) The organization. Last accessed October 2010, http://www.olympic.org/en/content/The-IOC/The-IOC-Institution1/.

International Olympic Committee (IOC) (2013) London 2012 Olympic Games: The Official Report. Last accessed June 11, 2015, http://doc.rero.ch/record/32414?ln=en.

Jaccard J, Turrisi R, Wan CK (1990) Interaction effects in multiple regression. *University Paper Series on Quantitative Applications in the Social Sciences*, Vol. 72 (Sage, Newbury Park, CA).

Jensen M (2006) Should we stay or should we go? Status accountability anxiety and client defections. *Admin. Sci. Quart.* 51(1):97–128.

Jensen M, Kim H, Kim BK (2012) Meeting expectations: A role-theoretic perspective on reputation. *The Oxford Handbook of Corporate Reputation* (Oxford University Press, Rochester, NY), 140–159.

Johnson DKN, Ali A (2004) A tale of two seasons: Participation and medal counts at the summer and winter Olympic games. *Social Sci. Quart.* 85(4):974–993.

Karpik L (2010) *Valuing the Unique the Economics of Singularities* (Princeton University Press, Princeton, NJ).

Kim JW, King BG (2014) Seeing stars: Matthew effects and status bias in Major League Baseball umpiring. *Management Sci.* 60(11):2619–2644.

Longley R (2014) Forsberg livid Canadians will ref Olympic gold-medal game. *Toronto Sun* (February 22), http://www.torontosun.com/2014/02/22/forsberg-livid-canadians-will-ref-olympic-gold-medal-game.

Majcher J (2010) Amateur boxing results. Retrieved January 2010, http://amateur-boxing.strefa.pl.

McNally J (2012) Phone interview with author, March 12.

Merton RK (1968) The Matthew effect in science. *Science* 159(3801):56–63.

Mitchell BR (1992) *International Historical Statistics: Europe* (Palgrave-Macmillan, New York).

Mitchell BR (1995) *International Historical Statistics: Africa/Asia/Oceania* (Palgrave-Macmillan, New York).

Mitchell BR (2003) *International Historical Statistics*: *Americas* (Palgrave-Macmillan, New York).

Moosa IA, Smith L (2004) Economic development indicators as determinants of medal winning at the Sydney Olympics: An extreme bounds analysis. *Australian Econom. Papers* 43(3):288–301.

Musch J, Grondin S (2001) Uneqaul competition as an impediment to personal development: A review of the relative age effect in sport. *Developmental Rev.* 21(2):147–167.

Neelamegham R, Jain D (1999) Consumer choice process for experience goods: An econometric model and analysis. *J. Marketing Res.* 36(3):373–386.

Nielsen (2008) Beijing Olympics draw largest ever global TV audience. Last accessed October 2010, http://blog.nielsen.com/nielsenwire/media_entertainment/beijing-olympics-draw-largest-ever-global-tv-audience/.

Novikov AD, Maksimenko M (1973) Soziale und okonomische faktoren und das niveau sportlicher leistungen in verschiedenen landern. *Sportwissenschaft* 2:156–167.

Podolny JM (1993) A status-based model of market competition. *Amer. J. Sociology* 98(4):829–872.

Podolny JM, Phillips DJ (1996) The dynamics of organizational status. *Indust. Corporate Change* 5(2):453–472.

Salganik MJ, Dodds P, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.

Sauder M, Lynn F, Podolny JM (2012) Status: Insights from organizational sociology. Cook KS, Massey DS, eds. *Annual Review of Sociology*, Vol. 38 (Annual Reviews, Palo Alto, CA), 267–283.

Sorenson O, Waguespack DM (2006) Social structure and exchange: Self-confirming dynamics in Hollywood. *Admin. Sci. Quart.* 51(4):560–589.

United Nations Department of Economic and Social Affairs (UN-DESA) (2008) Population Division, World Population Prospects 2006. Last accessed July 2008, http://www.un.org/esa/population/unpop.htm.

United Nations Statistics Division (2008) National Accounts Analysis of Main Aggregates Database. Last accessed July 2008, http://unstats.un.org/unsd/nationalaccount/.

United States Census Bureau (2008) International Database (IDB). Last accessed July 2008, http://www.census.gov/ipc/www/idb/index.html.

Waguespack DM, Sorenson O (2011) The ratings game: Asymmetry in classification. *Organ. Sci.* 22(3):541–553.

Weiss HJ (1986) The bias of schedules and playoff systems in professional sports. *Management Sci.* 32(6):696–713.

White HC (2002) *Markets from Networks: Socioeconomic Models of Production* (Princeton University Press, Princeton, NJ).

Wooldridge JM (2002) *Econometric Analysis of Cross-Section and Panel Data* (MIT Press, Cambridge, MA).

World Bank (2008) World Development Indicators (WDI) Database. Last accessed July 2008, http://go.worldbank.org/9EQZCKCE00/.

Yogev T (2010) The social construction of quality: Status dynamics in the market for contemporary art. *Socio-Econom. Rev.* 8(3):511–536.

Zitzewitz E (2006) Nationalism in winter sports judging and its lessons for organizational decision making. *J. Econom. Management Strategy* 15(1):67–99.

Zuckerman EW (2012) Construction, concentration, and (dis)continuities in social valuations. Cook KS, Massey DS, eds. *Annual Review of Sociology*, Vol. 38 (Annual Reviews, Palo Alto, CA), 223–245.