

Home Advantage in Professional Tennis*

Ruud H. Koning[†]

May 9, 2009

Abstract

Home advantage is a pervasive phenomena in sport. It has been well-established in team sports as basketball, baseball, and American football in the US, and soccer in Europe. Less attention has been given to home advantage in individual sports. In this paper we examine the existence of home advantage in professional tennis. We use match-level data to measure home advantage. Our test is based on logit-models, and consistent specification is addressed explicitly. Depending on the interpretation of home advantage, restrictions on the specification of the model need to be imposed. We find that significant home advantage exists for men, but such an advantage cannot be found for women.

Keywords: home advantage, tennis, logit model, consistent specification.

1 INTRODUCTION

Home advantage is a well established phenomena in different professional sports; a recent overview is given in Stefani (2008). Most of the published results concern home advantage in team sports, although it is well known that the medal count of a country organizing Olympic Games tends to be higher than expected (Balmer, Nevill, and Williams, 2001, 2003). Because most of the elements of the Olympic competition are individual events, this suggests that home advantage exists in individual sports as well. Only a few studies focus explicitly on home advantage in individual sports. A small but significant home advantage in speed skating is established in Koning (2005), and Nevill, Holder, Bardsley, Calvert, and Jones (1997) document little evidence of home advantage in professional tennis and golf. In this paper, we estimate home advantage

*This paper has benefitted from research assistance by Wim Siekman, Stefanie Bouw, and Tirza de Rond.

[†]Department of Economics & Econometrics, Faculty of Economics and Business, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands, r.h.koning@rug.nl.

in professional tennis, using a different methodology from the ones used before. Our approach uses individual match outcomes as opposed to individual player performance at a tournament level. The contribution of this paper is three fold. First, we propose a logit model to analyze outcomes of tennis matches. Home advantage is defined in the context of this model, and special care is taken to obtain a logically consistent model specification. Second, we measure not only home advantage in individual matches, but also as improved access to tournaments for home players. Third, we use a much expanded data set compared to earlier studies. This allows us to assess whether home advantage has changed over time and/or varies between type of tournament.

The plan of the paper is as follows. In section 2 we present a literature review and introduce our measures of home advantage. Specification of the model is presented in section 3. A description of the data and empirical results are given in section 4 and section 5 concludes.

2 LITERATURE REVIEW

Home advantage is a well established phenomena in different professional sports. Recent overviews documenting and interpreting its existence are given in Nevill, Balmer, and Wolfson (2005) and the articles following in that special issue, and Stefani (2008). Usually, home advantage is defined as ‘the consistent finding that home teams in sports competitions win over 50% of the games played under a balanced home and away schedule’ (Courneya and Carron, 1992). However, in the context of individual sports, where winners are usually determined in a tournament held in a fixed location, this definition is not applicable. Instead, a definition of home advantage as in Koning (2005) is more useful: ‘Home advantage is the performance advantage of an athlete, team or country when they compete at a home ground compared to their performance under similar conditions at an away ground’.

In this paper, we consider tennis so we need to define ‘performance’ more precisely. Also, we need to be more explicit about ‘similar conditions at an away ground’ since this is a unobserved (counterfactual) situation. We deal with this last issue in the context of the empirical models we use in the next two sections. As far as performance is concerned, one can think of two different measures of performance of a tennis player in a given tournament. The first is the tournament ranking of a player. The winner of a tournament receives tournament ranking 1, the losing finalist tournament ranking 2, the losing semi-finalist tournament ranking $(3 + 4)/2$, and so on. The other measure of performance is winning a particular match. The first measure is related to the sum (during a tournament) of the second measure. Since we want to test for the existence of home

advantage, we also need to make ‘similar conditions at an away ground’ operational. We discuss that issue in the context of the statistical models below.

Home advantage is usually attributed to four factors:

- crowd support;
- familiarity with local circumstances;
- fatigue due to travel time;
- specific rules that favor either the home or away team.

Crowd support may help a home player to perform better, to prevent him or her giving up sooner, and intimidate the opponent (assuming that that is not a home player). It is less likely that crowd support influences the decisions made by referees in tennis, as these are usually not from the home country. Moreover, the introduction of the ‘challenge’, where a player can ask for a video replay to see if a ball is in or out, has decreased reliance on decisions by the referee and line judges. Unfortunately, we don’t have data on the consistency of the crowd (that is, the ratio of home supporters to away supporters) so we cannot assess the influence of consistency of the crowd on home advantage as in Nevill, Newell, and Gale (1996). The argument of crowd support in the case of tournament style competitions can be broadened to include local sponsor interests. Individual tennis players are frequently sponsored by sponsors from their home country. Such sponsors may be present at the tournament, or the player may be involved in promotional activities during the tournament. In any case, good performance of a player on home ground makes it easier to renew existing contracts or find new sponsors. The importance of home sponsor support will be greater for subtop players than for top players, who are usually sponsored by sponsors operating globally. Familiarity with local circumstances may be relevant if players from, say Spain, have had the opportunity to practice much more on gravel tennis courts than other players. As tournament venues are usually not used for practice, only the type of surface may be a factor in determining home advantage. It is unlikely that home advantage is caused by travel fatigue for away players. All players travel from tournament to the next tournament, travel time does not vary very much between players. Moreover, during the tournament the players reside in hotels close to the venue. Finally, rules favor home players to the extent that tournament organizers have some leeway to allow a player to enter the tournament for other reasons than his or her world ranking. In particular, most tournaments offer wild-cards to local players, giving home players home advantage through improved access to the tournament. In section 4 we will establish whether this is indeed the case.

Home advantage in tennis has been studied before, mainly in Nevill et al. (1997) and Holder and Nevill (1997). In these papers, the measure of performance is the tournament ranking of a player and the central question that is answered is whether or not home players achieve a better than expected tournament ranking given their world ranking at the time of the tournament. The data used to test for home advantage are the four Grand Slam tournaments of 1993 (the Australian Open, the French Open, Wimbledon, and the US Open). Home advantage is estimated by first estimating a model where the log of tournament rank is regressed on the log of the world rank of a player. This baseline model is then extended with a dummy measuring home advantage (allowing for a different intercept for home players), and an interaction term between home advantage and log world rank (allowing for a different slope for home players). In both papers, the authors find little evidence of home advantage. Home players perform significantly better only in the Wimbledon tournament, there is no evidence of home advantage in the other three tennis tournaments. Holder and Nevill attribute this finding to ‘an anomaly of the data collection’ (p. 556): in their study ranks are missing for away players at Wimbledon with a rank lower than 100, but world rank is available for all home players. Our data used below do not suffer from this issue. The approach by Holder and Nevill is based on an implicit comparison between the performance of a home player with a certain world rank and an away player with that same rank. Since the strength (world rank) of the opponents does not enter as a covariate, one has to assume that the tournament draws for both players are identical.

In our approach below, we use match level data, and we relate the outcome of a match to observed covariates. Other studies have identified relevant covariates that we can use in our baseline model or as variables to moderate the effect of home advantage.

In a recent paper, del Corral (2009) estimates match uncertainty in tennis. He shows that Grand Slam tournaments have become less competitive since 2002 when the number of seeded players was increased from 16 to 32. According to his empirical results, competitiveness in tennis varies by sex, round, and surface. Also, he shows that upsets are mainly determined by quality differences between players (as measured by the difference in their rankings). In his empirical analysis, he does not take home advantage into consideration.

Klaassen and Magnus (2001, 2003) model the probability of winning a tennis match differently. In Klaassen and Magnus (2001) they focus on winning a point in a game, instead of winning a complete match. In that paper, the probability of winning the next point by the player who serves depends on time-constant covariates (quality difference between players, absolute quality of the match) and on time-varying covariates (who won the previous point, first point of the game, importance of the next point). Klaassen

and Magnus (2001) that the probability of winning the next point depends positively on the quality difference between both players, and positively on the absolute quality of the match. Their focus is on testing whether or not points are independently identically distributed, so their attention for other covariates is limited. They do report that they experiment with some other covariates, but home advantage is not one of them. Klaassen and Magnus (2003) follow up on that paper and consider the probability that a player wins a match (instead of the next point). This problem is close to the one we consider in the next section. They estimate the probability that the better player (as measured by world ranking) wins against the weaker player. Difference of quality is the main covariate. Again, they do not take home advantage into account.

3 MODEL SPECIFICATION

In this paper, we take ‘winning a match’ as the measure of performance in tennis. The main reason to prefer this measure over tournament ranking is that it allows us to measure home advantage while taking a given draw into account. When estimating the performance of a home player, it matters if a home player with world rank, say, 50 plays a top ten player in the first round, or a lucky qualifier. Considering the availability of individual match data, we can estimate the probability of winning a match easily, and let this probability depend on certain covariates. Obviously, home advantage and quality of the players are important covariates to be included in a model. However, the use of match data imposes restriction on the specification that can be estimated. This issue we address first.

Consider two arbitrary tennis players, A and B . Since A and B are chosen arbitrarily, we do not know which one is the higher ranked player, or which one plays at home. One can consider these players as being drawn randomly from the pool of participants in a tournament. The individual characteristics of A and B are measured as vectors x_A and x_B . This vector x may include world ranking, height, weight, age, number of career wins, etc. Also, there are variables that are common to both players, such as the time of the match, or the surface, and these common characteristics are denoted by the vector z_{AB} . The probability that A wins against B is denoted as $\Pr(A > B)$. We model this probability by a logit model:

$$\Pr(A > B) = \frac{1}{1 + e^{-f(x_A, x_B, z_{AB})}} \equiv \Lambda(f(x_A, x_B, z_{AB})). \quad (1)$$

The term $f(x_A, x_B, z_{AB})$ is also known as the index of the logit model. From equa-

tion (1), the probability that A does not win, and therefore B wins is

$$\begin{aligned}\Pr(B > A) &= 1 - \Pr(A > B) = \frac{e^{-f(x_A, x_B, z_{AB})}}{1 + e^{-f(x_A, x_B, z_{AB})}} \\ &= \frac{1}{1 + e^{f(x_A, x_B, z_{AB})}} = \Lambda(-f(x_A, x_B, z_{AB})).\end{aligned}$$

Alternatively, it follows directly from equation (1) that

$$\Pr(B > A) = \frac{1}{1 + e^{-f(x_B, x_A, z_{AB})}} = \Lambda(f(x_B, x_A, z_{AB})).$$

Hence, consistent model specification requires that

$$f(x_A, x_B, z_{AB}) = -f(x_B, x_A, z_{AB}) \quad \text{for all } x_A, x_B, z_{AB}. \quad (2)$$

If we assume that the index is linear

$$f(x_A, x_B, z_{AB}) = \beta'_1 x_A + \beta'_2 x_B + \beta'_3 z_{AB},$$

with β_1 , β_2 , and β_3 vectors of parameters to be estimated, restriction (2) takes the form

$$\begin{aligned}\beta'_1 x_A + \beta'_2 x_B + \beta'_3 z_{AB} &= -\beta'_1 x_B - \beta'_2 x_A - \beta'_3 z_{AB} \\ \beta'_1(x_A + x_B) + \beta'_2(x_B + x_A) + 2\beta'_3 z_{AB} &= 0\end{aligned}$$

If this is to hold for all x_A and x_B , clearly β_3 must be zero. In other words, covariates common to both players (such as surface, or day of the week) cannot enter the index as main effects. Also, the first two terms in the sum collapse to

$$(\beta_1 + \beta_2)'(x_A + x_B) = 0.$$

If this is to hold for all x_A and x_B , we have $\beta_2 = -\beta_1$ and the index must satisfies

$$f(x_A, x_B, z_{AB}) = \beta'_1 x_A + \beta'_2 x_B + \beta'_3 z_{AB} = \beta'_1(x_A - x_B). \quad (3)$$

In other words, only differences between covariates can enter the index. From an intuitive point of view, this is straightforward. Suppose the only covariate is a dummy h_A taking the value 1 if the player A plays at home. Also, suppose home advantage is positive. If player B also plays at home (so player A and B are of the same country, playing in their home country), it is not possible that the winning probabilities of both player A and player B increase due to the home advantage. Note that another implication of

restriction (3) is that the model cannot have an intercept.

Klaassen and Magnus (2001, 2003) do not address this identification issue explicitly. In Klaassen and Magnus (2001) a restriction as in (2) does not appear in their linear probability model because there is no restriction between the probability that player A wins the next point when he serves, and that player B wins the next point when he serves. Both players do not serve simultaneously for the next point. In Klaassen and Magnus (2003) they estimate a logit model as implied by (3), without discussing the general nature of the identification problem above. They do impose one restriction based on the observation that $\Pr(A > B) = \frac{1}{2}$ if both players are equally strong: they do not include an intercept in the logit model.

In this framework we measure home advantage by estimating the coefficient γ in

$$\beta'(x_A - x_B) + \gamma(h_A - h_B).$$

If both players are playing at home, $h_A - h_B$ is zero, and this is also the case if both players meet away. If A plays at home and B does not, A is expected to have a home advantage of γ (on the logit scale), if B plays at home and A does not, A has a *disadvantage* of γ . In other words, B has an advantage of γ . If home advantage exists, we should find $\gamma > 0$ and significant.

In this setup it is easy to moderate the effect of home advantage. Home advantage may vary with common covariates, for example, it may be weaker in Grand Slam tournaments. There are two possibilities to include moderating variables. First, one can condition on values of $x_A - x_B$ or z_{AB} and estimate the model for each subgroup. In other words, one estimates a model as (1) for each value of the moderating variable (for example, tournament is a Grand Slam tournament or tournament is not a Grand Slam tournament). Second, one can include interactions as covariates in the index, for example $z_{AB}h_A - z_{AB}h_B$. In this case, it is not necessary to include main effects in the index since these will have zero coefficient according to (3).

4 DATA AND TESTS

We assess participation in tournaments and estimate the model of the previous section using data on both professional men and women tennis tournaments. We have data on 22811 matches for men, over the period 2000 to 2008 (partially). The dataset contains information on the outcome of the match, the tournament, the date of play, round within the tournament, surface of the court, and the final result. Also, the dataset has information on the world ranking of both players at the beginning of the tournament,

Table 1: Access to tennis tournaments, men (2000-2008) and women (2007-2008)

| | home | ER_H | ER_A | tail | number |
|---------------|------|--------|--------|------|--------|
| men | | | | | |
| Grand Slam | 0.10 | 221.4 | 178.7 | 0.18 | 37 |
| Masters | 0.11 | 140.8 | 116.1 | 0.21 | 90 |
| International | 0.18 | 206.0 | 85.7 | 0.80 | 474 |
| women | | | | | |
| Tier I | 0.13 | 152.8 | 88.3 | 0.28 | 19 |
| Tier II | 0.11 | 129.8 | 63.4 | 0.49 | 29 |
| Tier III/IV | 0.10 | 204.8 | 90.9 | 0.88 | 62 |

home indicators, and (fixed) betting odds before the match. The men dataset has three tournament types (in decreasing order of importance): Grand Slam, Masters, and International tournaments. The dataset for women has the same information, but is much smaller in size: 2896 matches played in 2007 and 2008. Tournament types that we can distinguish are Tier I and Grand Slam tournaments, Tier II and Tier III/IV tournaments.

Our empirical strategy to test the existence of home advantage is as follows. First, we test for home advantage through better access to tournaments for home players. Then, we focus of home advantage given the set of participants of a tournament. We start by finding a baseline model that explains the outcome of a match satisfactorily. Of course, this specification satisfies consistency condition (3) of the previous section. Then, we extend this baseline specification with a home advantage dummy to assess whether or not home advantage exists. Finally, we establish whether a home advantage effect, if any, is moderated by other variables.

First, we consider improved access to tournaments as a type of home advantage. In tennis, the organization of a tournament has the option to give so called ‘wild cards’ to eligible players (see for example Chapter VII, article 7.12 of the ATP Rule Book). These wild cards are entries in the main draw of the tournament, and are awarded at ‘the sole discretion of the tournament’. Usually, they are given to players from the home country, young and talented players, or players who make a comeback after an injury. Wimbledon winner of the men’s tournament in 2001, Goran Ivanišević, participated in that tournament on a wild card, while being ranked only 125th on the World Ranking at that moment. High ranked players are guaranteed entry to a tournament, lower ranked players may have to play a qualification tournament or depend on a wild card.

There are no standard measures for home advantage through improved access to a tournament. We propose to measure improved access by looking at the world ranks

of weaker participants of a tournament. A tournament with k rounds has 2^k participants. Hence, the strongest draw possible would have players with world ranking 1 to 2^k . In practice, players with a weaker ranking than 2^k will also participate, for example because they receive a wild card from the organization of the tournament, or because they qualify through a qualification tournament that precedes the main tournament. Especially tournaments of lower importance are played simultaneously and therefore one would expect that more lower ranked players participate in such tournaments. Organizers have to contract enough players to have a complete schedule, and a player can participate in only one tournament at a given moment in time. Should home players have improved access to tournaments through this mechanism, we should find that among the weaker participants, home players are weaker than away players. We measure this as follows. If the best players participate in a tournament, the weakest observed rank is 2^k . The number of players with a lower rank (i.e., ranked on a higher position than 2^k), and their ranks, indicate the strength of the tournament. For each tournament in our sample, we calculate the median position of home and away players ranked lower than 2^k . Then, this median is averaged over all tournaments, and the results are in columns labeled ER_H for home players and ER_A for away players in table 1. The numbers should be interpreted as follows. In Grand Slam tournaments for men, 128 players participate. In an average Grand Slam tournament, the median rank of home players whose rank is weaker than 128, is 224.5. The same number is 178.2 for away players. Participating home players not in the top 128 of the world ranking are of lesser quality than participating away players not in the top 128 of the world ranking. This effect is found for all types of tournaments, both for men and women. The difference is significant as can be established using a permutation test approach. For each tournament, we condition on the observed ranks of the weaker players (those with a world rank weaker than 2^k) and the number of home players. The ‘home’ label is then permuted over the observed ranks and we calculate the average median ranks for home players and away players, and its difference, for this permutation. We repeat this procedure a large number (10000) of times to obtain the distribution of the difference of the average median ranks, under the null hypothesis that the ranks of home players and away players are distributed similarly (that is, according to the observed ranks of the weaker players for each tournament). The p -values of the observed differences obtained from Table 1 is 0 in all six cases. Moreover, notice that the difference is largest in tournaments of lowest importance, as expected. In the fifth column we give the fraction of players with a lower rank than 2^k . 18% of the players in Grand Slam tournaments for men are ranked lower than 128. Clearly, this fraction is much higher for the least important types of tournament in our dataset, both for men and women.

The last column in Table 1 gives the number of tournaments of each type in our dataset. From Table 1 we conclude that there is clear evidence that home players have improved access to tournaments, especially to tournaments of lower importance.

Now that we have established that there is home advantage by improved access to participation in tournaments, we proceed to measure home advantage within the tournament. We follow the approach discussed in the previous section, so first we specify a baseline model that depends on individual specific covariates only, possibly interacted with moderating variables. In our data set, only three individual specific covariates are available: home advantage, odds of winning as offered by different bookmakers, and world ranking. We do not want to include odds offered by bookmakers as a covariate, even though that information will generate a very good baseline model, see McHale and Forrest (2005) and Koning (2009). The reason is that home advantage will be priced into such fixed odds already. In fact, if betting odds contain all relevant information on the result of a tennis match, we could use betting odds as the dependent variable. However, McHale and Forrest (2007) and Koning (2009) show that betting odds in tennis are not unbiased predictors of the outcome of a tennis match so we do not use betting odds as the dependent variable. This leaves us with world ranking as the individual specific covariate. Variables common to both players are type of tournament, quality of the match (measured by the sum of the rankings of both players, this variable is categorized in four quartiles), surface, and year. The choice between entering the difference of world ranks in equation (1) in levels, or in logarithms with base 2 as suggested by Klaassen and Magnus (2001) is left to the data. As our baseline specification, we use the model that provides the best fit.

So first, we estimate two models:

$$\Pr(A > B) = \frac{1}{1 + \exp(-\beta(\log WR_A - \log WR_B))}, \quad (4)$$

and

$$\Pr(A > B) = \frac{1}{1 + \exp(-\beta(WR_A - WR_B))}. \quad (5)$$

To choose between these specifications we draw calibration plots in Figure 1, where the actual outcome of a match is regressed on the predicted probabilities obtained from models (4) and (5). If the estimated probabilities from these models are unbiased estimators of the actual outcome, the calibration curve would be coinciding with the 45-degree line. Clearly, the model with difference of the logarithm of world ranks (model (4), top panels in Figure 1) provides a much better fit than the model with dif-

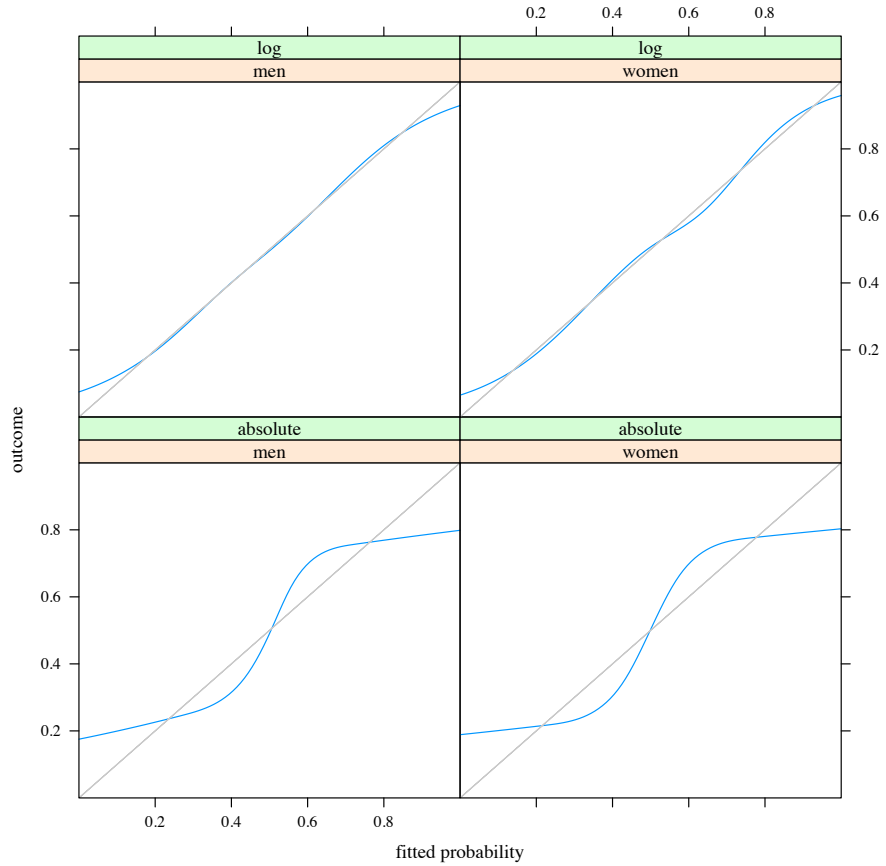


Figure 1: Calibration of baseline model, for men and women, world ranking difference in logs and levels.

ference of absolute levels of world ranks (model (5), lower panels in Figure 1). For this reason, we use the difference of log world ranks only, and denote this variable as $\log WR \equiv \log WR_A - \log WR_B$.

In a second step, we test whether the effect of world rank in our baseline specification needs to be moderated by interacting it with absolute quality of the match (measured by the sum of world ranks of both players), round, and type of tournament. We test significance of these interactions at a 99% confidence level, because of the size of the data set. For men, it turns out that the type of tournament is significant. The estimation results of the preferred specifications are given in Table 2. The first four lines in that table give the estimate of β as in model (4) for each of the tournament types

Table 2: Home advantage and baseline specifications for men and women

| | $\hat{\beta}$ | s.e.($\hat{\beta}$) | $\hat{\beta}$ | s.e.($\hat{\beta}$) |
|----------------------------------|---------------|-----------------------|---------------|-----------------------|
| men | | | | |
| (log <i>WR</i>) × GrandSlam | −0.453 | 0.016 | −0.456 | 0.017 |
| (log <i>WR</i>) × Masters | −0.301 | 0.014 | −0.304 | 0.014 |
| (log <i>WR</i>) × International | −0.388 | 0.010 | −0.392 | 0.010 |
| (log <i>WR</i>) × ATP | −0.457 | 0.061 | −0.447 | 0.066 |
| <i>HA</i> | | | 0.143 | 0.029 |
| women | | | | |
| log <i>WR</i> | −0.510 | 0.018 | −0.517 | 0.020 |
| <i>HA</i> | | | 0.122 | 0.083 |

indicated. The estimated coefficient is negative, as expected: if player *A* has a better ranking than player *B*, he is more likely to win. The relation between world rank and the probability of winning is slightly stronger for women, as is seen from the more negative coefficient in the first line of the second part of Table 2. In the case of women, none of the potentially moderating variables has a significant effect.

It is now easy to test for the presence of home advantage in individual tennis matches, by extending the model with a home advantage dummy. The estimates of the coefficient are given in the rows labelled *HA* in Table 2. The effect is positive and significant at any reasonable level of significance for men, and positive but insignificant for women. From now on, we focus on home advantage for men only. Similar to our search for a baseline specification, we interact home advantage with variables that are common to both players to assess whether other variables moderate the home advantage effect. Home advantage does not vary by type of tournament. We do find that home advantage varies significantly with the absolute quality of the match, as measured by the sum of the world rankings of both players. We recode the quality variable into four quartiles, so that the first quartile corresponds to matches of high quality (the sum of both rankings is low), and the fourth quartile corresponds to matches of lowest quality. The estimation results are given in Table 3 and it appears that home advantage is stronger in games of high quality, and decreases to zero for games of lowest quality. Home advantage may tip the balance in a match between two very good players, and is not important in deciding the outcome in a match between two weak players. In fact, home advantage does not seem to exist in a match between two weak players. That outcome is determined by relative world ranking and the type of tournament. We also examine whether home advantage varies with other common characteristics of the

Table 3: Home advantage interacted with quality of match (men)

| | $\hat{\beta}$ | s.e.($\hat{\beta}$) |
|------------------------------------|---------------|-----------------------|
| (log WR) \times GrandSlam | -0.455 | 0.017 |
| (log WR) \times Masters | -0.305 | 0.014 |
| (log WR) \times International | -0.387 | 0.010 |
| (log WR) \times ATP | -0.441 | 0.066 |
| $HA \times Q_1$ | 0.341 | 0.067 |
| $HA \times Q_2$ | 0.261 | 0.063 |
| $HA \times Q_3$ | 0.109 | 0.057 |
| $HA \times Q_4$ | -0.006 | 0.048 |

match, as the type of surface, year, country, round in the tournament, or being the favorite. Not one of these variables turns out to be significant, so we conclude that our final specification in Table 3 is quite robust.

To interpret the relative magnitude of home advantage compared to the effect of differences in world ranking, we use an equivalence scale. Consider two players A and B who play a match away. Suppose we were to substitute player A by another player, A^* , who plays at home. Because player A^* plays at home, he can be a lower ranked player than player A , and still have the same winning probability against player B . This, of course, is due to home advantage. Let δ_{HA} be the effect of home advantage. Then the world ranking of player A^* playing at home that is equivalent to the world ranking of player A playing away follows from

$$\delta_{HA} + \beta(\log WR_{A^*} - \log WR_B) = \beta(\log WR_A - \log WR_B).$$

A player with world rank $WR_{A^*} = 2^{-\delta_{HA}/\beta} WR_A$ who plays at home has an equal winning probability against player B as a player with world rank WR_A who does not play at home. The estimates of Table 3 are translated into the factor $2^{-\delta_{HA}/\beta}$ in Table 4. It is clear that this factor decreases for all tournament types when matches of lesser quality is considered. The magnitude of the estimates are large. Consider a match in a Grand Slam tournament., between the world rank number 3 and the world rank number 7, both playing away. This match is in the first quartile of the quality distribution. According to the results in Table 4, a home player with world rank 12 ($\approx 1.68 \times 7$) has the same probability of winning against the world rank 3 player as the away player with world rank 7. Home advantage amounts to five places on the world ranking in this case.

Table 4: Home advantage expressed as equivalence factor (men)

| | GrandSlam | Masters | International | ATP |
|-------|-----------|---------|---------------|-------|
| Q_1 | 1.680 | 2.169 | 1.842 | 1.708 |
| Q_2 | 1.488 | 1.810 | 1.597 | 1.507 |
| Q_3 | 1.181 | 1.282 | 1.217 | 1.188 |
| Q_4 | 1.000 | 1.000 | 1.000 | 1.000 |

5 CONCLUSIONS

In this paper, we have tested for the existence of home advantage in professional tennis, both for men and women. Home advantage was measured along two dimensions: as improved access to tournaments, and as an increased probability of winning a tennis match, given the relative strength of both players. Our main findings are as follows. First, both in men tennis and women tennis we see that home players have significant improved access to tournaments, for example through wild cards, qualification tournaments preceding the main draw, or direct invitation by the organization of the tournament. This is especially prevalent in tournaments of lesser importance. Second, using a consistent specification to model outcomes of individual matches, we found in men tennis a significant and quantitatively important home advantage effect. This effect is strongest in matches between highly skilled opponents, and absent when we consider a match between two weak players. No such home advantage effect in individual matches is found for women.

REFERENCES

- Balmer, N.J., A.M. Nevill, and A.M. Williams (2001). Home advantage in the Winter Olympics (1908–1998). *Journal of Sports Sciences* 19(2), 129–139.
- Balmer, N.J., A.M. Nevill, and A.M. Williams (2003). Modelling home advantage in the Summer Olympic Games. *Journal of Sports Sciences* 21(6), 469–478.
- del Corral, J. (2009). Competitive balance and match uncertainty in grand slam tennis: Effects of seeding system, gender, and court surface. *Journal of Sports Economics* forthcoming.
- Courneya, K.S. and A.V. Carron (1992). The home advantage in sport competitions: A literature review. *Journal of Sport and Exercise Physiology* 14, 13–27.

- Holder, R.L. and A.M. Nevill (1997). Modelling performance at international tennis and golf tournaments: Is there a home advantage? *The Statistician* 46(4), 551–559.
- Klaassen, F.J.G.M. and J.R. Magnus (2001). Are points in tennis independent and identically distributed? evidence from a dynamic binary panel data model. *Journal of the American Statistical Association* 96(454), 500–509.
- Klaassen, F.J.G.M. and J.R. Magnus (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research* 16(2), 257–267.
- Koning, R.H. (2005). Home advantage and speed skating: Evidence from individual data. *Journal of Sports Sciences* 23(4), 417–427.
- Koning, R.H. (2009). Betting odds and market efficiency. Manuscript.
- McHale, I.G. and D. Forrest (2005). The importance of recent scores in a forecasting model for professional golf tournaments. *IMA Journal of Management Mathematics* 16(2), 131–140.
- McHale, I.G. and D. Forrest (2007). Anyone for tennis (betting)? *European Journal of Finance* 13(8), 156–166.
- Nevill, A., N. Balmer, and S. Wolfson (2005). The extent and causes of home advantage: Some recent insights. *Journal of Sports Sciences* 23(4), 335 – 336.
- Nevill, A.M., R.J. Holder, A. Bardsley, H. Calvert, and S. Jones (1997). Identifying home advantage in international tennis and golf tournaments. *Journal of Sports Sciences* 15, 437–443.
- Nevill, A.M., S.M. Newell, and S. Gale (1996). Factors associated with home advantage in English and Scottish football matches. *Journal of Sports Sciences* 14, 181–186.
- Stefani, R.T. (2008). Measurement and interpretation of home advantage. In J. Albert and R.H. Koning (Eds.), *Statistical Thinking in Sports*, Chapter 12, pp. 203–216. Boca Raton: Chapman & Hall/CRC.