

Appendix to 'Home Advantage in Professional Tennis'*

Ruud H. Koning[†]

1 DESCRIPTIVE STATISTICS

The data set we use are match results on professional tennis tournaments for both men and women. We have data on 25266 tennis matches for men (3 January 2000 to 1 March 2009), and 5113 tennis matches for women (31 December 2006 to 1 March 2009). These data were obtained from the website `www.tennis-data.co.uk`. The datasets contain results from all Grand Slam tournaments during the periods indicated, as well as results from most other major tournaments. The data set for men contains all results in the main draw from 615 different tournaments, for women it has all results in the main draw of 131 tournaments.

Within each match, one player is randomly labelled 'player A' and the other is labelled 'player B'. Because the assignment to these labels is random, the win percentage of player A is 0.50 in both datasets. Definitions of the variables are as follows:

location location of the match,

tournament name of the tournament,

*Journal of Sports Sciences, forthcoming.

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

date date of the match,

series type of tournament,

court indoor/outdoor court,

surface type of surface of the court,

round round within the tournament of the match,

best.of indicator best of three match or best of five match,

country country where the match is played,

player.a name of player labelled as A,

player.b name of player labelled as B,

rank.a rank of player A at start of tournament (world rankings are updated weekly)
(WR_A),

rank.b rank of player B at start of tournament (WR_B),

country.a country of player A,

country.b country of player B,

win.a indicator that player A wins the match (if not, player B wins),

year year of the match,

ha.a indicator whether player A plays at home (**country** equals **country.a**),

ha.b indicator whether player B plays at home (**country** equals **country.b**),

ha difference of home advantage: $HA = \mathbf{ha.a} - \mathbf{ha.b}$,

log.rank difference of log rankings: $\Delta \log WR = \log_2 WR_A - \log_2 WR_B$,

log.quality quality of the match: $\log_2 WR_A + \log_2 WR_A$.

The original data sets contain more variables, we list descriptive statistics only of variables actually used in the analysis.

data set men
22 Variables 25226 Observations

location

n missing unique
 25226 0 94

lowest : 's-Hertogenbosch Acapulco Adelaide Amersfoort Amsterdam
 highest: Vienna Vina del Mar Warsaw Washington Zagreb

tournament

n missing unique
 25226 0 145

lowest : AAPT Championships Abierto Mexicano
 highest: US Open Verizon Tennis Challenge

ABN AMRO W
 Vietnam Op

date

n missing unique
 25226 0 1809

lowest : 2000-01-03 2000-01-10 2000-01-17 2000-02-07 2000-02-14
 highest: 2009-02-25 2009-02-26 2009-02-27 2009-02-28 2009-03-01

series

n missing unique
 25226 0 4

Grand Slam (4510, 18%), Masters (5058, 20%), International (15251, 60%)
 ATP (407, 2%)

court

n missing unique
 25226 0 2

Indoor (4610, 18%), Outdoor (20616, 82%)

surface

n missing unique
 25226 0 4

Carpet (1633, 6%), Clay (8416, 33%), Grass (2621, 10%), Hard (12556, 50%)

round

n missing unique
 25226 0 7

Frequency	1st Round	2nd Round	3rd Round	4th Round	Quarterfinals	Semifinals	The Final
	12064	6841	1791	389	2344	1195	602
%	48	27	7	2	9	5	2

best.of

n missing unique
 25226 0 2

best of three (20625, 82%), best of five (4601, 18%)

country

n missing unique
25137 89 35

lowest : Argentina Australia Austria Brazil Canada
highest: The Netherlands UAE USA Uzbekistan Vietnam

player.a

n missing unique
25226 0 785

lowest : Abel M. Acasuso J. Agassi A. Agenor R. Ager J.
highest: Zib T. Zimonjic N. Zitko R. Zovko L. Zverev M.

player.b

n missing unique
25226 0 761

lowest : Abdulla M. Abel M. Acasuso J. Adaktusson J. Agaev E.
highest: Zib T. Zimonjic N. Zitko R. Zovko L. Zverev M.

rank.a

n missing unique Mean .05 .10 .25 .50 .75 .90 .95
25218 8 662 79.6 5 10 26 56 96 153 225

lowest : 1 2 3 4 5, highest: 1497 1533 1554 1564 6530

rank.b

n missing unique Mean .05 .10 .25 .50 .75 .90 .95
25221 5 655 79.11 5 10 26 56 96 151 223

lowest : 1 2 3 4 5, highest: 1517 1543 1734 2250 6530

country.a

n missing unique
25152 74 73

lowest : Algeria Argentina Armenia Australia Austria
highest: Uruguay USA Uzbekistan Venezuela Zimbabwe

country.b

n missing unique
25153 73 71

lowest : Algeria Argentina Armenia Australia Austria
highest: Uruguay USA Uzbekistan Vietnam Zimbabwe

win.a

n missing unique Sum Mean
25226 0 2 12607 0.4998

year

n missing unique
25226 0 10

Frequency 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
% 11 12 11 11 11 11 11 10 10 2

ha.a

	n	missing	unique
25071		155	2

FALSE (21402, 85%), TRUE (3669, 15%)

ha.b

	n	missing	unique
25072		154	2

FALSE (21333, 85%), TRUE (3739, 15%)

ha

	n	missing	unique	Mean
25007		219	3	-0.003159

-1 (2951, 12%), 0 (19184, 77%), 1 (2872, 11%)

log.rank

	n	missing	unique	Mean	.05	.10	.25	.50	.75
25213		13	11666	0.005032	-3.67711	-2.78427	-1.38466	-0.01294	1.39232
2.82290		3.77259							

lowest : -12.673 -9.950 -9.572 -9.493 -9.219
highest: 8.755 8.892 8.931 10.074 12.673

log.quality

	n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
25213		13	7905	11.04	6.615	7.807	9.562	11.281	12.738	13.931	14.639

lowest : 1.000 1.585 2.000 2.322 2.585
highest: 18.853 18.965 19.252 19.359 19.389

data set women
22 Variables 5139 Observations

location

n missing unique
5139 0 62

lowest : 's-Hertogenbosch Acapulco Amelia Island Antwerp Auckland
highest: Tashkent Tokyo Vina del Mar Warsaw Zurich

tournament

n missing unique
5139 0 70

lowest : Abierto Mexicano Acura Classic
highest: Toronto US Open

date

n missing unique
5139 0 591

lowest : 2006-12-31 2007-01-01 2007-01-02 2007-01-03 2007-01-04
highest: 2009-02-25 2009-02-26 2009-02-27 2009-02-28 2009-03-01

series

n missing unique
5139 0 4

Tier I (1015, 20%), Tier II (859, 17%), Tier III/IV (1810, 35%)
NaN (1455, 28%)

court

n missing unique
5139 0 2

Indoor (610, 12%), Outdoor (4529, 88%)

surface

n missing unique
5139 0 4

Carpet (161, 3%), Clay (1424, 28%), Grass (464, 9%), Hard (3090, 60%)

round

n missing unique
5139 0 7

	1st Round	2nd Round	3rd Round	4th Round	Quarterfinals	Semifinals	The Final
Frequency	2453	1410	329	101	480	241	125
%	48	27	6	2	9	5	2

best.of

n missing unique
5139 0 2

best of three (5138, 100%), best of five (1, 0%)

country

n	missing	unique
5010	129	35

lowest :	Australia	Austria	Belgium	Canada	Chile
highest:	The Netherlands	Turkey	UAE	USA	Uzbekistan

player.a

n	missing	unique
5139	0	358

lowest :	Abduraimova N.	Adamczak M.	Afinogenova K.	Albanese L.	Amanmuradova A.
highest:	Zakopalova K.	Zec Peskiric M.	Zhang S.	Zheng J.	Zvonareva V.


player.b

n	missing	unique
5139	0	377

lowest :	Abramovic I.	Adamczak M.	Afinogenova K.	Ahn K.H.	Albanese L.
highest:	Zhang S.	Zheng J.	Zhou Y.M.	Zoric N.	Zvonareva V.

rank.a

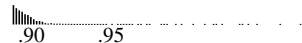
n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
5126	13	350	74.04	4.00	9.00	22.25	53.00	94.00	152.00	204.00



lowest : 1 2 3 4 5, highest: 992 999 1061 1063 1093

rank.b

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
5126	13	367	75.03	5.0	9.0	23.0	52.0	93.0	148.0	204.8



lowest : 1 2 3 4 5, highest: 1137 1172 1216 1323 1431

country.a

n	missing	unique
5025	114	60

lowest :	Argentina	Australia	Australian Nationality	Austria	Belarus
highest:	Ukraine	USA	Uzbekistan	Uzbekistan	Venezuela

country.b

n	missing	unique
5020	119	61

lowest :	Argentina	Australia	Australia	Austria	Belarus
highest:	Ukraine	USA	Uzbekistan	Uzbekistan	Venezuela

win.a

n	missing	unique	Sum	Mean
5139	0	2	2572	0.5005

year

n	missing	unique
5139	0	4

2006 (6, 0%), 2007 (2396, 47%), 2008 (2294, 45%), 2009 (443, 9%)

ha.a

n	missing	unique
4899	240	2

FALSE (4446, 91%), TRUE (453, 9%)

ha.b

n	missing	unique
4897	242	2

FALSE (4454, 91%), TRUE (443, 9%)

ha

n	missing	unique	Mean
4794	345	3	0.0008344

-1 (381, 8%), 0 (4028, 84%), 1 (385, 8%)

log.rank

n	missing	unique	Mean	.05	.10	.25	.50
5113	26	3671	-0.007575	-3.49368	-2.67199	-1.34395	0.01577
.75	.90	.95					
1.32193	2.64331	3.45943					

lowest : -7.304 -7.180 -6.870 -6.759 -6.755
highest: 7.285 7.401 7.555 7.836 8.000

log.quality

n	missing	unique	Mean	.05	.10	.25	.50	.75	.90	.95
5113	26	3086	10.86	6.170	7.267	9.299	11.196	12.676	13.819	14.520

lowest : 1.000 1.585 2.000 2.322 2.585
highest: 18.004 18.099 18.115 18.350 19.515

2 FAMILIARITY WITH SURFACE

One potentially important component of home advantage is familiarity with a certain surface. Typically, there are four different types of surface available in tennis: carpet, clay, grass, and hard court. Not all types are readily available in every country, and hence, a player may not have been able to develop an extensive experience with that surface. For example, courts with a grass surface (usually these courts are outdoor courts) are hardly available in Nordic countries. As discussed in section 4 we measure the supply of available courts within each country by the distribution of matches played over these types of surface. This distribution is given in table 1. We assume that this distribution is representative of the tennis infrastructure in a country and measures the possibility to develop experience on a particular surface to a player of that country. Ideally, we would have experience on a surface for each player, but we do not have any information on the surfaces played on before the player entered our sample.

Table 1: Distribution of matches by country and surface

	Carpet	Clay	Grass	Hard	Matches
Argentina	0.000	1.000	0.000	0.000	246
Australia	0.000	0.000	0.000	1.000	1754
Austria	0.000	0.706	0.000	0.294	929
Brazil	0.000	0.777	0.000	0.223	265
Canada	0.000	0.000	0.000	1.000	534
Chile	0.000	1.000	0.000	0.000	270
China	0.000	0.000	0.000	1.000	368
Colombia	0.000	1.000	0.000	0.000	59
Croatia	0.319	0.681	0.000	0.000	379
Denmark	0.000	0.000	0.000	1.000	121
France	0.297	0.480	0.000	0.223	2273
Germany	0.000	0.762	0.178	0.060	1516
Great Britain	0.000	0.000	0.969	0.031	1910
India	0.000	0.000	0.000	1.000	356
Italy	0.169	0.831	0.000	0.000	890
Japan	0.000	0.000	0.000	1.000	441
Mexico	0.000	1.000	0.000	0.000	295
Monaco	0.000	1.000	0.000	0.000	531
Morocco	0.000	1.000	0.000	0.000	265
New Zealand	0.000	0.000	0.000	1.000	297
Poland	0.000	1.000	0.000	0.000	207
Portugal	0.000	0.989	0.000	0.011	266
Qatar	0.000	0.000	0.000	1.000	300
Romania	0.000	1.000	0.000	0.000	269
Russia	0.772	0.000	0.000	0.228	543
San Marino	0.000	1.000	0.000	0.000	30
Spain	0.000	0.700	0.000	0.300	1055
Sweden	0.000	0.475	0.000	0.525	507
Switzerland	0.529	0.471	0.000	0.000	507
Thailand	0.000	0.000	0.000	1.000	176
The Netherlands	0.000	0.322	0.316	0.363	827
UAE	0.000	0.000	0.000	1.000	295
USA	0.000	0.057	0.038	0.905	6338
Uzbekistan	0.000	0.000	0.000	1.000	89
Vietnam	0.000	0.000	0.000	1.000	29