Online Appendix

Romain Gauriot Lionel Page John Wooders

January 24, 2023

Abstract

1 Appendix D: Robustness of Power Simulations

As a robustness check, here we reproduce the simulation results reported in Section 6, but where now the simulated data matches the characteristics of the observed data, point game by point game, rather than just in aggregate. Specifically, if point game i has n_R^i serves to the right, n_L^i serves to the left, and an empirical winning frequency of $\hat{p}^i$, then the simulated data for point game i has n_R^i serves to the right, n_L^i serves to the left, and the probability of winning a point is $\hat{p}^i$ for serves in each direction (and hence the null hypothesis that $p_L^i = p_R^i$ is true). The number of winning serves to the right and left are therefore distributed, respectively, $B(n_R^i, \hat{p}^i)$ and $B(n_L^i, \hat{p}^i)$ in the simulated data for point game i.

The Power of Our Test

The subsection "The Power of Our Test" in Section 6 provided the power functions for the Pearson joint test and the KS tests based on the Pearson p-values and the Fisher exact t-values. It demonstrated that for "small" samples of 40 point games, the test based on the t-values was substantially more powerful than the other two. In addition, for "large" samples of 7000 point games, the test based on the t-values was especially powerful – the joint null hypothesis of equality of winning probabilities is almost surely rejected for even small departures from equilibrium play.

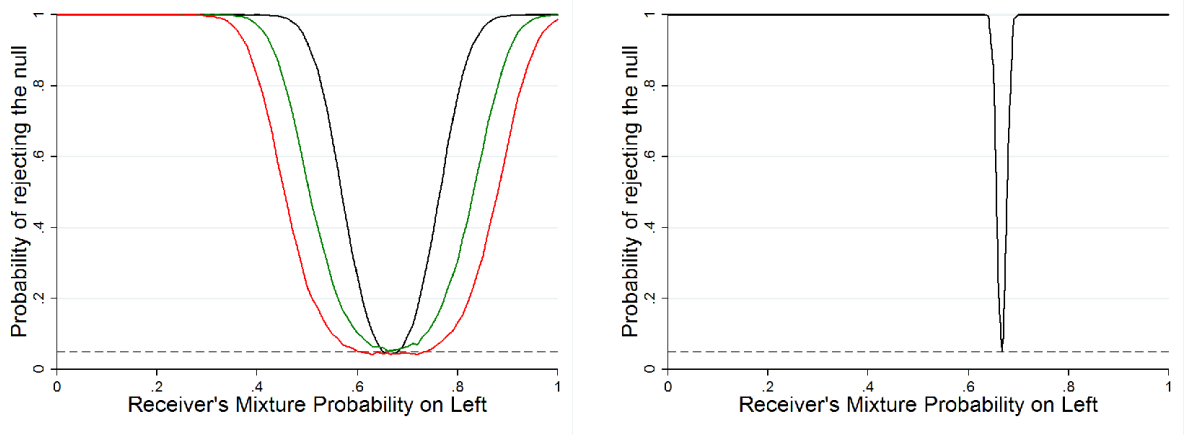
Table D1 is the analogue Table 6. Comparing to the two tables reveals that the KS test based on the t's has similar power when the simulated data matches the characteristics of the Hawk-Eye data. Table D1 shows that the (true) joint null

hypothesis of equality of winning probabilities is rejected, at the 5% significant level, for sure by the KS test based on the p's and it is rejected with probability .726 by the Pearson joint test. These results reaffirm our conclusion that these tests are not valid for large samples.

True θ	KS based on t's	KS based on p's	Pearson joint test
0.65	0.834	1	0.746
0.66	0.212	1	0.716
2/3	0.051	1	0.726
0.67	0.093	1	0.726
0.68	0.675	1	0.764

Table D1: Rejection rate for H_0 at the 5% level, $N = 7198$

Figure D1 is the analogue to Figure 14 and shows that the power functions in Figure 14 are largely unchanged when the data is simulated (under the null hypothesis) to match characteristic of the WW data (Figure D1(a)) or the Hawk-Eye data (Figure D1(b)).



(a) $N = 40$ (b) $N = 7198$

Figure D1: Power Functions for KS test based on t-values (black), p-values (red), and Pearson joint (green)

2 Appendix E: Ball Bounces

Figure E1 below shows actual and imputed ball bounces for male second serves from the deuce court.

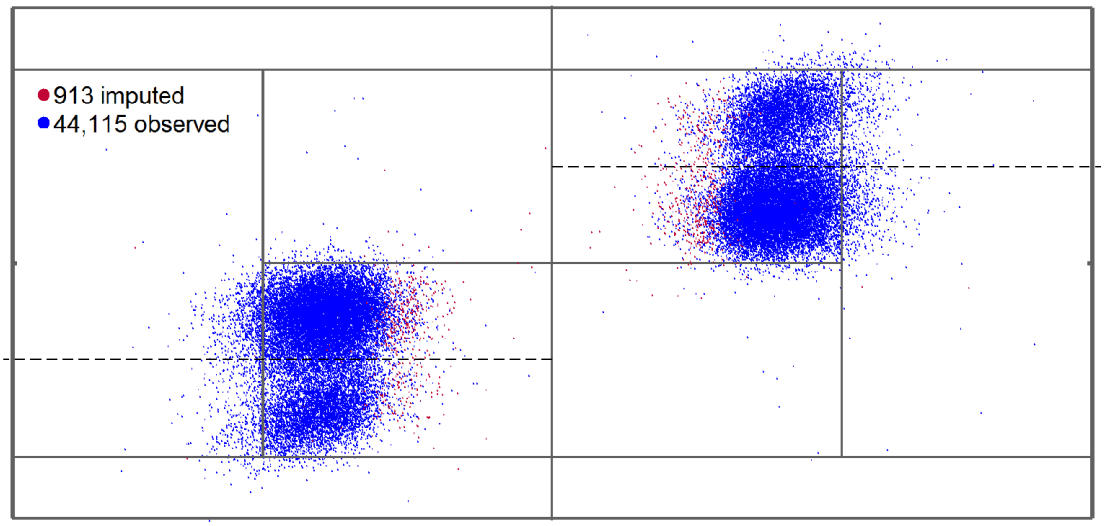


Figure E1: Ball Bounces for Deuce Court Second Serves by Men

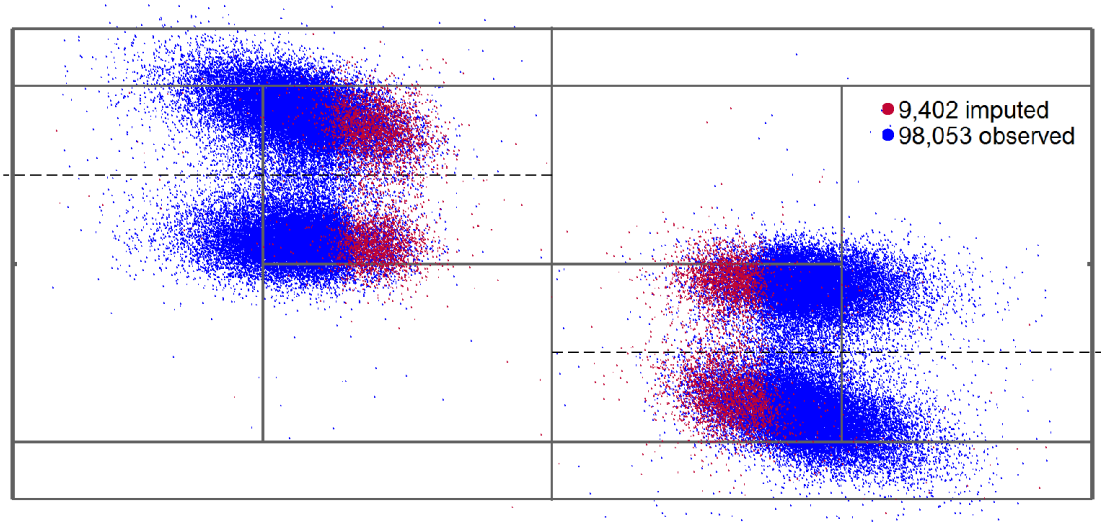


Figure E2: Ball Bounces for Ad Court First Serves by Men

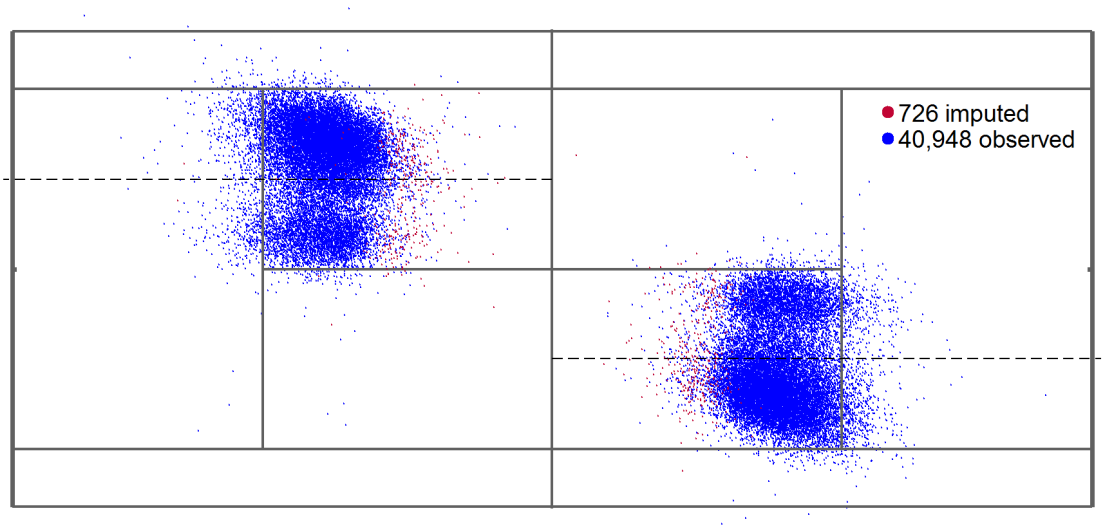


Figure E3: Ball Bounces for Ad Court Second Serves by Men

Ball bounces for first and second serves by women are below.

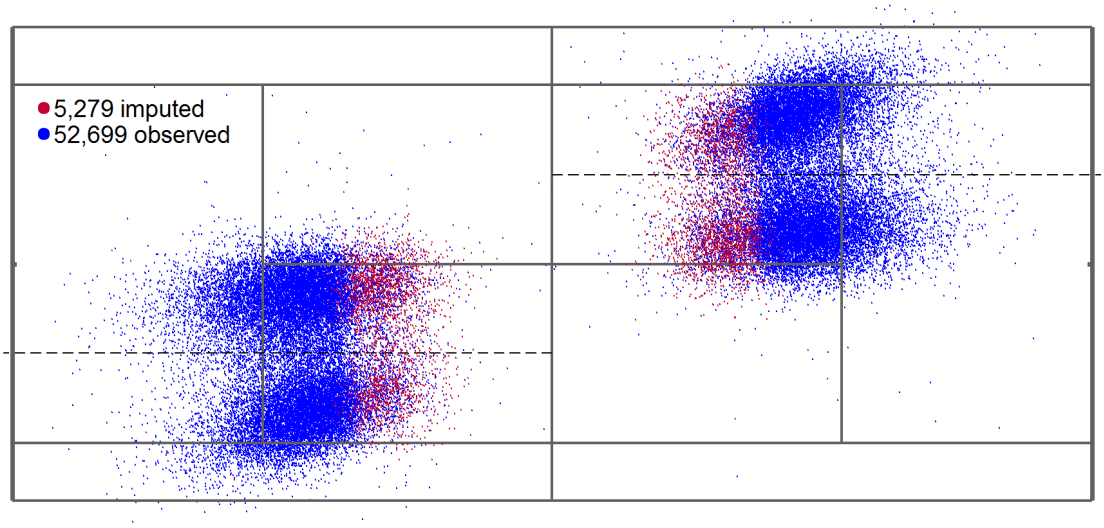


Figure E4: Ball Bounces for Deuce Court First Serves by Women

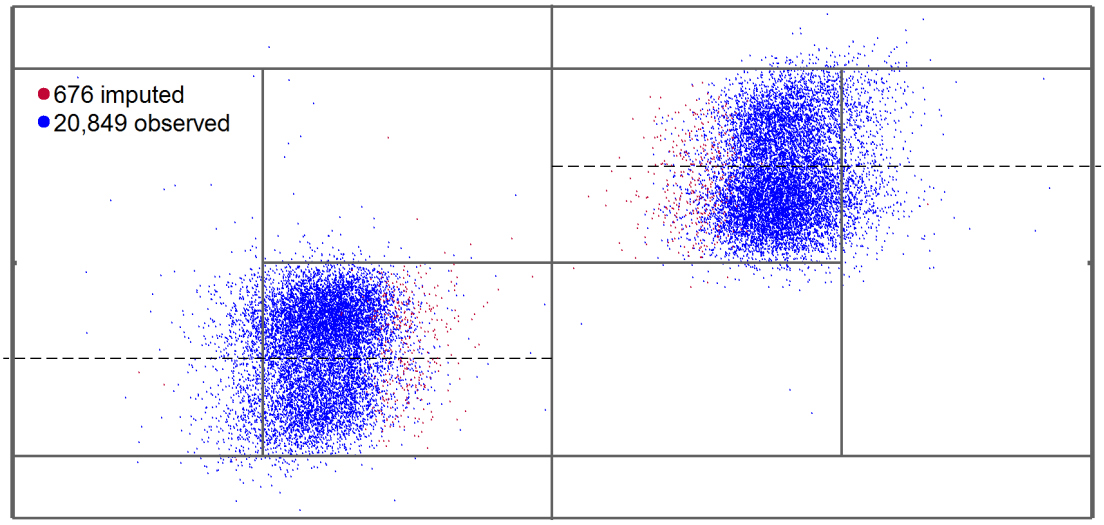


Figure E5: Ball Bounces for Deuce Court Second Serves by Women

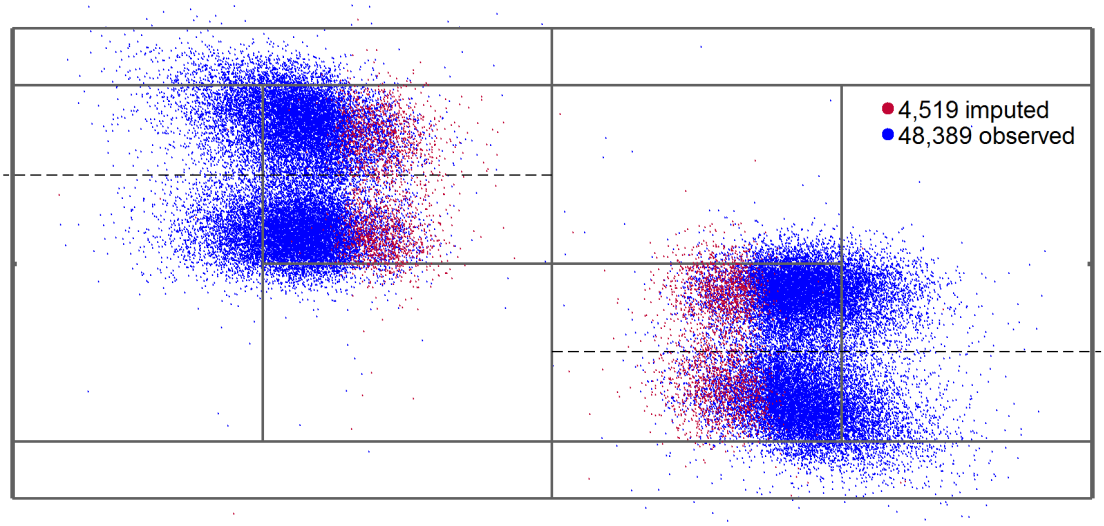


Figure E6: Ball Bounces for Ad Court First Serves by Women

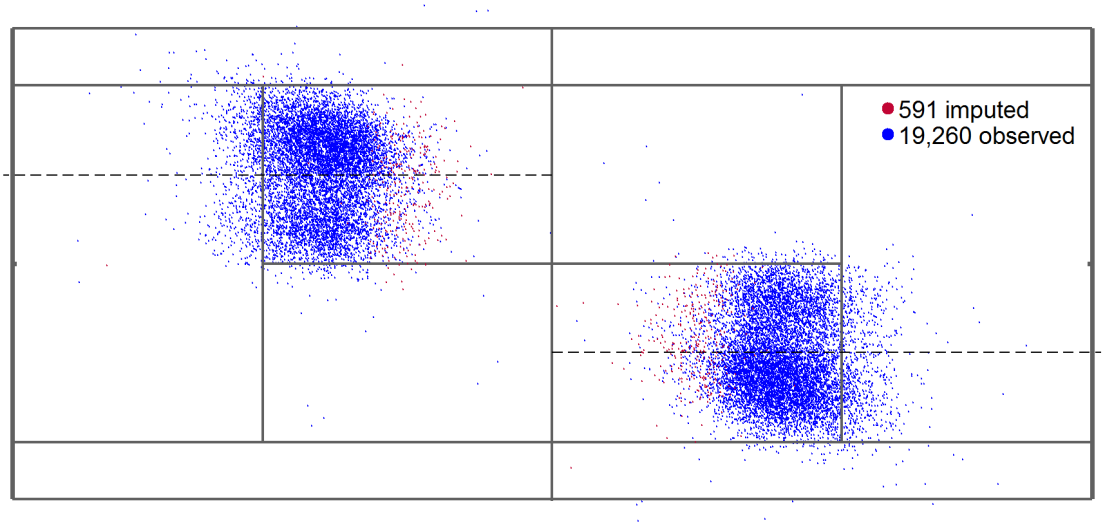


Figure E7: Ball Bounces for Ad Court Second Serves by Women