# Data and code for: Expertise, Gender, and Equilibrium Play.

Romain Gauriot, Lionel Page and John Wooders

February 13, 2023

We extract information from a large proprietary dataset to which we were granted access by Dr Paul Hawkins, managing director of the Hawk Eye Innovation company at the time of the request. The proprietary nature of the data creates some limitations about what can be made publicly available.

The whole raw data set itself cannot be made public. It includes point-by-point ball-tracking data from more than 3,000 tennis matches. This data has commercial value for the company, and Hawk Eye Innovation is unwilling for it to be made public. Researchers can contact the company managing director, Steve Carter, to request access to a similar dataset for research purposes (Steve.Carter@hawkeyeinnovations.co.uk).

The specific subset of data used in our analysis can be made public. The code used Stata we include two datasets: *pointgames.dta* and *Pointbypointforcorrelation.dta*. We also provide those dataset in csv format.

In *pointgames.dta* each observation is a pointgame. This dataset is used to test the equality of winning probabilities and do the run test of Section 5. The variables included are:

- gender: players' gender.

- player: each player is given the id 1 or 2 in the pointgame.

- court: ad or deuce court.

- servesR1, servesL1, servesR2, servesL2: number of first (second) serves on the right (left).

- pointwonR1, pointwonL1, pointwonR2, pointwonL2: number of first (second) serves won on the right (left).

- run1: number of runs when considering only first serves.

- run2: number of runs when considering only second serves.

- rrank: receiver ranking.

- rrank: server ranking.

In *Pointbypointforcorrelation.dta* each observation is a serve. This dataset is used to look at serial correlation. The variable included are:

- idpointgame: id of the pointgame.

- idinpointgame: id of the serve in the pointgame.

- rightserve: dummy variable equal to 1 if the serve is on the right.

- totalserves1: number of first serves in the pointgames.

- serve: 1st or 2nd serve.

- court: ad or deuce court.

- gender: players' gender.

- rankservermagnus: receiver ranking using the Klaassen and Magnus (2001)'s ranking transformation.

- rankreceivermagnus: server ranking using the Klaassen and Magnus (2001)'s ranking transformation.

In addition, we include two other datasets. *AER.dta* includes the pointgames' data from Walker and Wooders (2001) and Hsu, Huang, and Tang (2007) and is used for Figure 15 and 16. *Oneil.dta* includes the data from O'Neill (1987) and is used for Figure 8.

The Stata code used to perform the estimations is provided in do files in this folder. The names of the do file are self-explanatory.

When a statistical test is run multiple times (e.g., Figure 4 Right panel), the results from each iteration are stored in a temporary dataset saved in the folder *Results*. When the code produces a figure, the figure is saved in the folder *Figure*.

The power simulations (e.g., Figure 14) were run on the New York Abu Dhabi's High Performance Computer (HPC). *Power Simulation to run on the HPC(Figure 14 and Table 6).do* is the Stata code run on the HPC. There are 101 $\theta$'s (see footnote 32). The simulations for each $\theta$ are run as an independent job on the HPC (so those 101 jobs are run in parallel). *Appending computation from the HPC (Figure 14 and Table 6).do* aggregates the results of those 101 jobs and produces the figures.

For more information, please email romain.gauriot@gmail.com.

# References

**Hsu, S.H., C.Y. Huang, and C.T. Tang.** 2007. "Minimax Play at Wimbledon: Comment." *American Economic Review* 97 (1): 517–523.

**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.

**O'Neill, Barry.** 1987. "Nonmetric test of the minimax theory of two-person zerosum games." *Proceedings of the National Academy of Sciences* 84 (7): 2106–2109.

**Walker, M., and J. Wooders.** 2001. "Minimax Play at Wimbledon." *American Economic Review* 91 (5): 1521–1538.