Assignment: Structural estimation of the game of tennis

John Rust, Professor of Economics Georgetown University Spring 2023

I. Consider a simplified model of tennis, where the game is just a static repetition of “point games” where the server and receiver have two key decisions: 1) the server chooses whether to serve right or left, and 2) the receiver chooses whether to “focus” or “expecting” the serve to be to the right or the left. Here we are following the example in figure 1 of Walker and Wooders 2001 *American Economic Review* article “Minimax Play at Wimbledon”. I repeat the “payoff matrix” from their figure 1 below.

Figure 1: Stylized Tennis “Point Game” from Walker and Wooders 2001

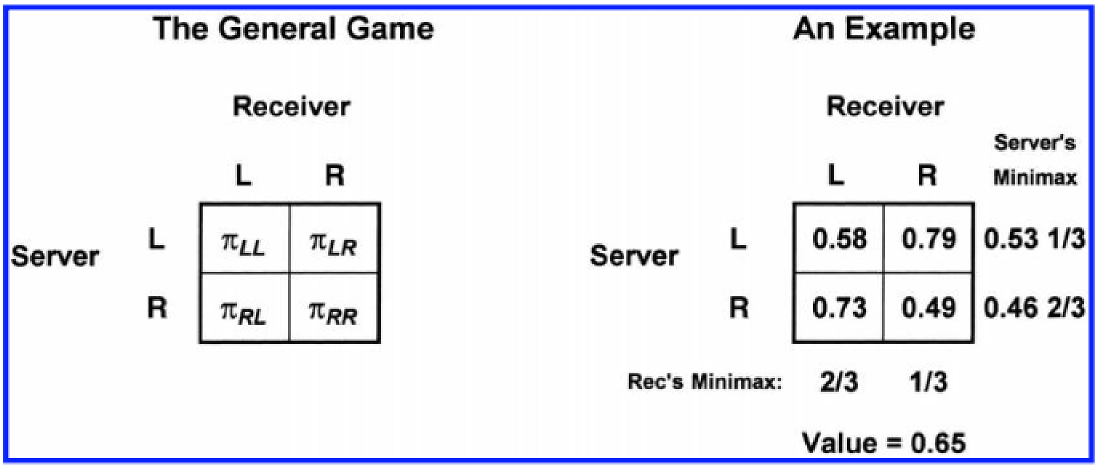

FIGURE 1. THE POINT GAME

Note: Outcomes (cell entries) are the probability the Server wins the point.

The “payoffs” in the matrix are actually conditional win probabilities we refer to as the “*POPs*” (for Point Outcome Probabilities). For example if the server chooses to serve to the left (L) and the receiver puts his/her attention on returning a ball hit to the left, then the probability the server wins is .58. If the receiver focused on returning a ball to the right when the server actually hit the ball to the left, then the probability the server wins is .79. This higher probability is due to the “surprise factor” that the receiver was expecting the serve to go R when in fact it was hit L. We also see this feature in the 2nd row of the payoff matrix, if the server hits to the R, but the receiver is expecting it to come L, then the probability the server wins the point is .73, whereas if the receiver expects the ball to come to his R, then the server will win with only probability .49 since in this case the receiver is more “mentally prepared” for the serve and thus not “surprised”. This is a “reduced form” model of tennis in that we do not explicitly model decisions made during the “rally” that follows a successful serve, nor do we model the option of a 2nd serve in the event of a faulted first serve. To incorporate faults on the first serve, we just treat a serve that was *intended for the L* but faults (i.e. ball hits the net and does not go over) as a loss for the server, and similarly for a serve to the R. So in this stylized example of tennis, we ignore the option of a second serve in the event of a faulted first serve.

1. Suppose the objective of both players is to win each point. Show that this stylized game-theoretic model of each point of tennis can be viewed as a *constant sum game.* Can you find a Nash equilibrium of this game and is it unique?

2. I generated data on choices made by a simulated server and receiver playing 2000 point games with each other. The matrix, `tennis_data.txt` is a 2000×3 matrix where there first column contains the server choices (1 for serving L, 0 for serving R), the second column is the receiver's choice (1 for expecting L and 0 for expecting R), and the third column is the outcome of the game (1 for server winning the point, 0 if the server loses the point). Can you estimate the 2×2 matrix for POPs in this case using these data?

3. Suppose you are asked to test whether the data in `tennis_data.txt` is consistent with Nash equilibrium play. Nash equilibrium in this case is also known as "Minimax" since the receiver's decision is designed to maximize the receiver's chance of winning, but since the probability the server plus the probability the receiver wins a point must sum to 1, this is a constant sum game and thus the receiver's act to maximize his/her chance of winning necessarily minimizes the server's chance of winning. The existence of a Minimax solution to 2 person simultaneous move constant or zero sum games was first proved by John von Neumann in 1928, and was a forerunner of the notion of Nash equilibrium before Nash came along and helped to introduce the general notion of non-constant sum games and an equilibrium concept that generalizes Minimax now known as *Nash equilibrium.* A Nash equilibrium can either be in *pure strategies* or *mixed strategies.* Can you find all Nash equilibria of the game if we impose a reasonable restriction on the POPS that $\pi_{LL} < \pi_{LR}$ and $\pi_{LR} > \pi_{RR}$ and characterize whether the equilibria are in pure or mixed strategies?

4. Using your answer to part 3 above, can you describe a *structural estimation algorithm* that imposes the constraint of Nash equilibrium and estimates the unknown structural parameters $\theta = (\pi_{LL}, \pi_{LR}, \pi_{RR}, \pi_{RR})$? **HINT:** Consider a Nested Fixed Point Maximum Likelihood estimation algorithm where you write out the likelihood function for the data, consistent of the choice probabilities for the server and receiver times the probability of the outcome of the game conditional on those choices, but also imposing Nash equilibrium on the choice probabilities of server and receiver.

5. Now consider estimating an *unrestricted* version of tennis where we no longer impose the constraint that play is necessarily according to Nash equilibrium strategies. How many parameters does this unrestricted model have and how would you estimate it?

6. Given your answers to parts 4 and 5, can you describe a *likelihood ratio test* of the hypothesis that play is consistent with Nash equilibrium? How many degrees of freedom does this test have and what is its asymptotic distribution under the null hypothesis that play is consistent with Nash equilibrium? **HINT:** Recall the likelihood ratio test equals 2 times the log-likelihood ratio of the unrestricted model versus the restricted model and the degrees of freedom equal the difference in the number of parameters of the unrestricted and restricted models.

7. Now consider a *Wald test* of the hypothesis of Nash equilibrium play. Can you write down how to calculate the Wald test statistic? What are its degrees of freedom and what is its asymptotic distribution under the null hypothesis that play is consistent with Nash equilibrium?

8. **Substantial credit for this part** Can you use the `tennis_data.txt` and write computer code to test the hypothesis of Nash equilibrium play? You can do this in any computer language you like. The pencil and paper answers you gave to parts 4 to 7 above should be your guide to writing the computer code necessary to carry out the Wald and likelihood ratio tests.

9. In reality, we typically do not observe a receiver's mental act of "focusing" on which direction the server will choose on any given serve to be mentally prepared to return it. Suppose we do not observe the receiver's choice. Can we still test the hypothesis that the server's choices (where we can typically observe the direction of each serve by a server) is consistent with Nash equilibrium? **HINT:** In a mixed strategy equilibrium, Nash equilibrium implies that there should be *equal win probabilities* for the server from choosing to serve to the L or the R. That is, the server should be indifferent about which direction to serve in a mixed stategy Nash equilibrium. Is this testable? If so can you describe either a Wald or Likelihood Ratio test for this that does not use the 2nd column of `tennis_data.txt` that contains the receiver's choice of which direction to focus on?

10. **Substantial credit for this part** Can you discuss the identification of the structural parameters θ in the case a) full information, we observe both the server's and receiver's decisions in every game, b) partial information, we only observe the serve direction but not the receiver's decision.

11. **Substantial extra credit for this part** Given one or both of the tests you constructed in part 9, can you implement them in computer code and carry out the test to determine whether the server's serve choices are consistent with the hypothesis of Nash equilibrium (Minimax) play?