

### The efficiency of top agents: An analysis through service strategy in tennis

Klaassen, Franc J.G.M.; Magnus, Jan R.

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

[www.peerproject.eu](http://www.peerproject.eu)

#### Empfohlene Zitierung / Suggested Citation:

Klaassen, F. J., & Magnus, J. R. (2009). The efficiency of top agents: An analysis through service strategy in tennis. *Journal of Econometrics*, 148(1), 72-85. <https://doi.org/10.1016/j.jeconom.2008.09.036>

#### Nutzungsbedingungen:

Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: <http://www.peerproject.eu> Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**gesis**  
Leibniz-Institut  
für Sozialwissenschaften

#### Terms of use:

This document is made available under the "PEER Licence Agreement". For more information regarding the PEER-project see: <http://www.peerproject.eu> This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der  
  
Leibniz-Gemeinschaft

## Accepted Manuscript

The efficiency of top agents: An analysis through service strategy in tennis

Franc J.G.M. Klaassen, Jan R. Magnus

PII: S0304-4076(08)00164-4  
DOI: [10.1016/j.jeconom.2008.09.036](https://doi.org/10.1016/j.jeconom.2008.09.036)  
Reference: ECONOM 3111

To appear in: *Journal of Econometrics*

Received date: 10 July 2007  
Revised date: 14 March 2008  
Accepted date: 21 September 2008

Please cite this article as: Klaassen, F.J.G.M., Magnus, J.R., The efficiency of top agents: An analysis through service strategy in tennis. *Journal of Econometrics* (2008), doi:10.1016/j.jeconom.2008.09.036

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# The efficiency of top agents: an analysis through service strategy in tennis

Franc J.G.M. Klaassen

*Tinbergen Institute and Faculty of Economics and Business,  
University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam,  
The Netherlands*

Jan R. Magnus

*CentER and Department of Econometrics & OR, Tilburg University,  
P.O. Box 90153, 5000 LE Tilburg, The Netherlands*

This revision: March 12, 2008

**JEL Classification:** C14; C15; D01; D21

**Key words:** Inefficiency, Frontier, Optimal strategy, Tennis

**Abstract:** We consider the question whether top tennis players in a top tournament (Wimbledon) employ an optimal (efficient) service strategy. While we show that top players do not, in general, follow an optimal strategy, our principal result is that the estimated inefficiencies are not large: the inefficiency regarding winning a point on service is on average 1.1% for men and 2.0% for women, implying that — by adopting an efficient service strategy — players can (on average) increase the probability of winning a match by 2.4%-points for men and 3.2%-points for women. While the inefficiencies may seem small, the financial consequences for the efficient player at Wimbledon can be substantial: the expected paycheck could rise by 18.7% for men and even by 32.8% for women. We use these findings to shed some light on the question whether economic agents are successful optimizers.

**Corresponding author:** Jan R. Magnus

**E-mail addresses:**

f.klaassen@uva.nl (Franc Klaassen), magnus@uvt.nl (Jan R. Magnus).

## 1 Introduction

Most economists believe that agents maximize something and that they are successful in doing so. Stigler (1976) provides a typical and outspoken example of such a view in his critique of Leibenstein's (1966) notion of X-efficiency. Any inefficiency claimed by Leibenstein or others is — according to Stigler — nothing but a failure to measure all relevant inputs, or to correctly specify what is being maximized. For example, John Capozzi's well-known business maxim:

*Only make a great deal if you have no intention of ever doing business with that person again. . . otherwise make a good deal,*

would not — if followed — indicate inefficiency. It might indicate that the agent does not maximize short-term profit, but he or she would still maximize long-term profit or, more vaguely, 'utility.' Førsund, Lovell and Schmidt (1980, p. 21) point out that such a view is essentially an act of faith, as it can be neither proved nor disproved.

Perhaps, however, we *can* prove or disprove the hypothesis that agents are successful maximizers. For this we would need a situation where (a) it is unambiguous what it is that the agent wishes to maximize, and (b) clean and complete data are available. Under these conditions any apparent sub-optimality must be true suboptimality. Such a situation can only be found in a very structured environment. One possibility for creating such an environment is through a laboratory experiment. This has the advantage of maximum control, but it also has disadvantages: in laboratory experiments, reported violations of optimality are often belittled by claiming that the incentives were insufficient or that the violations will be eliminated by learning or by market competition. Although Tversky and Kahneman (1986) agree that these factors are relevant, they question whether accounting for them would ensure fully optimal choices. In the end, this is an empirical issue.

Our environment is a field experiment: the service strategy of tennis players at Wimbledon. This is a real-life setting where high prizes can be won, competition is fierce, and the players (our agents) are highly trained and very experienced. They want to win matches on the professional tour, especially at the 'grand slam' tournaments of which Wimbledon is arguably the most important. It seems reasonable to assume that these agents wish to maximize the probability of winning a match. In addition, our data are clean. The tennis environment is therefore ideal to study the efficiency of human behavior, also because tennis has an unusual and archaic rule which does not exist in other comparable sports (table tennis, badminton, volleyball), namely that the server has two chances to bring the ball into play (first service, second service) rather than one. Even with one service the question needs to be answered how difficult this service should be: too easy and the

server loses the point in the subsequent rally, too difficult and the service will be a fault much of the time. Choosing the right balance is obviously important. In the current situation the server has two services, and this has two consequences. First, to determine the optimal strategy of choosing the strengths of the two services is more difficult than in the one-service case. Second, the existence of two services doubles the amount of information we possess about a player's strategy against a specific opponent.

Based on a simple model we can calculate the players' optimal strategy and we can compare this with their actual strategy. The discrepancy (ratio) between the two defines their efficiency. We shall show that tennis players are not entirely successful in maximizing their objective function. This is not surprising, because 100% efficiency is humanly unattainable given the continuity of the decision problem. Our main interest, however, is not a discrete test for (in)efficiency, but rather an estimate of the continuous measure of how close top players are to full efficiency. Since we also have a good measure of the quality differences among the players (position on world ranking), we can examine how the players' inefficiencies depend on their quality and competition within a match. This will shed some light on Tversky and Kahneman's (1986) question whether incentives, experience, and competition ensure fully optimal choices.

We emphasize that our interest is not in *relative* efficiency (how well does one agent perform relative to another agent, in particular relative to the most successful agent), but rather in *absolute* efficiency (how well do agents perform compared to the optimum achievable). Thus we shall not assume that top agents lie on the efficiency frontier. Instead we want to measure how far they are removed from the frontier. We also note that we do not measure the efficiency of *average* agents, but of *top* agents. We want to know whether top tennis players are efficient and, if not, how much room for improving their efficiency exists. Several methods have been proposed to allow for inefficiencies of firms, in particular stochastic frontier analysis; see the surveys by Førsund, Lovell and Schmidt (1980) and Schmidt (1985/86), the monograph by Kumbhakar and Lovell (2000), and a Bayesian perspective by Koop, Osiewalsky and Steel (1997). Nonparametric tests of optimizing behavior of consumers as well as firms have been introduced by Varian (1982, 1985). The main emphasis of these studies is, however, the measurement of efficiency (productivity) of an average agent, while our interest is on the efficiency of a top agent. The latter should also help us to better understand the relevance of high levels of experience and ability for efficiency.

Sports statistics (and sports economics) has developed from an anecdotal field where one collects statistics (so many double faults, so many aces), to an almost-respectable discipline. An important reason for this development is that sport statistics can help answer behavioral questions. Moreover, sports

data are readily available and they are measured much more precisely than most economic data. This has led to studies on racial discrimination (Gwartney and Haworth, 1974; Kahn and Sherer, 1988; Nardinelli and Simon, 1990; Stone and Warren, 1999; Szymanski, 2000; Kanazawa and Funk, 2001; and Goff, McCormick and Tollison, 2002), efficiency of the betting market (Zuber, Gandar and Bowers, 1985; Sauer, Brajer, Ferris and Marr, 1988; Golec and Tamarkin, 1991; Dixon and Coles, 1997; and Gray and Gray, 1997), comparison of betting markets and financial markets (Levitt, 2004), the effect of labor strikes on consumer demand (Schmidt and Berri, 2004), preferences under risk (Julien and Salanié, 2000), mixed strategy equilibria (Walker and Wooders, 2000, 2001; Chiappori, Levitt and Groseclose, 2002; and Palacios-Huerta and Volij, 2008), incentive effects (Ehrenberg and Bognanno, 1990), rationality (Gandar, Zuber, O'Brian and Russo, 1988), optimal labor contracts (Lazear and Rosen, 1981), control of externalities (Carlton, Frankel and Landes, 2004), favoritism (Garicano, Palacios-Huerta and Prendergast, 2005), maximizing behavior of firms (Romer, 2006; and Adams, 2006), and so on.

The studies most closely related to our paper are Walker and Wooders (2001), Chiappori, Levitt and Groseclose (2002), Palacios-Huerta and Volij (2008), Romer (2006), and Adams (2006). Walker and Wooders examine whether tennis players aim their first service to the receiver's left or right (only two options), in such a way that the probability of winning a point is equal for the two directions, as the theory of mixed-strategy equilibrium implies. Their results provide some evidence that the behavior of top players conforms closely to this theory, which contrasts to the conclusions in many experiments. Our set-up and analysis differs from theirs in three important respects. First, since the probability of winning a point depends not only on the direction of the first service (especially when the first service is a fault), but also on spin, speed, and many other factors, we concentrate on a broader concept, namely the probability of serving in, and we consider both the first and second service. Our analysis should therefore have higher power. Second, since our analysis is continuous rather than discrete, we not only test for efficiency, but also (and in particular) estimate the *magnitude* of the inefficiency. Third, both Walker and Wooders and we are interested in the relevance of a player's quality for optimal play. Using a different data set, involving inexperienced card players, Walker and Wooders reject the theory, and they take this as evidence that play by high-quality players conforms more closely to the theory than play by novices. In our analysis, we can test the relevance of a player's quality for optimality within a single data set.

Chiappori, Levitt and Groseclose (2002) also test mixed-strategy play, but now for penalty kicks in soccer rather than for tennis. Their results are also consistent with optimality, thus confirming the conclusions of Walker and Wooders (2001). Palacios-Huerta and Volij (2008) bring professional

soccer players to the laboratory to play card games, and find that they play close to optimality, in sharp contrast to college students.

Romer (2006) studies profit maximization of firms by focusing on coaches and their teams in professional (American) football. More specifically, he tests for optimality of the coach's decision on 'fourth down' between kicking and 'going for it.' In contrast to Walker and Wooders (2001), Chiappori, Levitt and Groseclose (2002), and Palacios-Huerta and Volij (2008), he overwhelmingly rejects optimality. Even though Romer studies teams and allows for interactions between agents, whereas we study individual behavior, our tennis data may help answer some of his questions. Romer gives two possible explanations for his rejection and the overconservative behavior of coaches. First, the coach's objective function may be more complicated than Romer assumes; second, coaches are not able to correctly maximize. Unfortunately, says Romer, there is little evidence which of the two explanations causes the suboptimal behavior. In tennis, however, there are two services, and the maximization for the second service is easier than for first service. We shall see that this fact can be exploited to shed some light on the true cause of suboptimal behavior.

Adams (2006) questions Romer's results, in particular the assumption that success rates on third down equal those on fourth down, and concludes that coaches may in fact make optimal decisions.

The literature thus reports mixed evidence on optimality. Our contribution is to provide some new and cleaner insights to help resolve this ambiguity, and also to estimate the level of efficiency, rather than testing for perfect efficiency, which is the focus in the existing literature.

The organization of this paper is as follows. In Section 2 we present the theoretical model, based on the relationship between the probability that a service is in  $(x)$  and the conditional probability that the server wins the point if the service is in  $(y(x))$ . We prove the existence and uniqueness of an optimal strategy and introduce the concept of efficiency of the service. In Section 3 we propose a functional form for  $y(x)$ . Since tennis allows two services, we have information on the (current, not necessarily optimal) strategy  $(x_1, x_2)$  and the corresponding 'yield'  $(y_1, y_2)$ . We do not, however, observe these probabilities directly, only the associated relative frequencies. In Section 4 we emphasize the importance of correcting for measurement error and the use of random effects. In Section 5 we estimate the key probabilities using a generalized method of moments (GMM) approach. Next, in Section 6 we discuss the identification and estimation of the curvature of the  $y$ -curve. Then, in Section 7, we estimate and discuss the (in)efficiencies of top tennis players. In our conclusion we try to relate these results to the inefficiency of economic agents.



## 2 Theoretical model

Consider a tennis match between two players  $i$  and  $j$ . Both  $i$  and  $j$  are maximizing the probability of winning the match. If points are independent, then each server chooses that service strategy which will maximize the probability of winning a point. We will develop a model that answers the question how difficult a player should make his or her service in order to maximize the probability of winning a point on service.

### 2.1 Existence and uniqueness of an optimal strategy

Given the strengths of both players, let  $x_1$  denote the probability that the first service is in, and let  $x_2$  denote the probability that the second service is in. Also, let  $y(x)$  denote the *conditional* probability that player  $i$  wins the point if the service has probability  $x$  of being in. So,  $y_1 := y(x_1)$  denotes the conditional probability that player  $i$  wins the point on his or her first service if the first service is in, and similarly for the second service  $y_2 := y(x_2)$ . We realize that both  $x$  and  $y$  will be influenced by other factors (speed, direction, spin, concentration, (mental) effort to determine the optimal strategy, emotions). For the moment we abstract from these unobservable factors, but we shall discuss them in Section 2.3. Implicit in our set-up is the assumption that a first service of probability  $x$  yields the same  $y$  as a second service of probability  $x$ . We therefore ignore the possibility that a receiver may be more aggressive on a second service of strength  $x$  than on a (relatively weak) first service of the same strength  $x$ . In reality, however, there is typically a gap in strength between the first and second service, which mitigates the influence of the single  $y$ -curve assumption. Moreover, in Sections 6.3 and 7.1 we examine the importance of this assumption by allowing for separate  $y$ -curves for the first and second services, and show that our results are robust.

Given  $x$  and  $y(x)$  we define  $w(x) := x \cdot y(x)$ , which transforms the conditional probability  $y$  into an unconditional probability  $w$ . The probability that player  $i$  wins the point is then given by

$$p(x_1, x_2) = w(x_1) + (1 - x_1)w(x_2). \quad (1)$$

If the functional form of  $y(x)$  is known, we can calculate the optimal service strategy  $(x_1^*, x_2^*)$  by maximizing  $p(x_1, x_2)$ . Under suitable regularity conditions and in the absence of a boundary solution, the optimal strategy satisfies the first-order conditions

$$w'(x_1^*) = w(x_2^*), \quad w'(x_2^*) = 0. \quad (2)$$

In what follows we shall, however, allow for the possibility of a boundary solution  $x_2^* = 1$ .



It seems reasonable to assume that the easier a player makes his or her service, the more likely it is that the service is in ( $x$  increases), but the less likely it is that the point is won if the service is in ( $y$  decreases). Hence, we impose the condition that  $y$  is a (locally) decreasing function of  $x$ .

**Condition R1 (monotonicity):** The real-valued function  $y$  is continuous and monotonically decreasing on  $[0, 1]$ , and satisfies  $0 \leq y(x) \leq 1$  for all  $x \in [0, 1]$ .

In order to achieve existence and uniqueness of the optimal strategy we need more than monotonicity. We shall also impose (local) concavity.

**Condition R2 (concavity):** The real-valued function  $y$  is twice differentiable on  $(0, 1)$  with  $y'(1) := \lim_{x \uparrow 1} y'(x) < 0$ , and  $w(x)$  satisfies  $w'(x) < 0$  for all  $x \in (0, 1)$ .

Condition R2 implies that  $w$  is strictly concave on the interval  $(0, 1)$  and reaches a unique maximum for some  $x \in (0, 1]$ . This reflects the fact that if a player's service is too easy he/she is unlikely to win the point, but if the service is too difficult he/she is also unlikely to win the point — in fact,  $w(0) = 0$ . There should be an optimal service, neither too difficult nor too easy, which maximizes the player's probability of winning the point on that service. Given these regularity conditions we can now prove Theorem 1.

**Theorem 1 (existence and uniqueness):** Assume that regularity conditions R1 and R2 hold. Then there exists a unique optimal service strategy  $(x_1^*, x_2^*)$  which maximizes  $p(x_1, x_2)$ .

*Proof.* Both steps of the proof use the fact that  $w'$  is monotonically decreasing. For the first step (regarding  $x_2^*$ ) we note that  $w'(0) := \lim_{x \downarrow 0} w'(x) = y(0) > 0$ . If  $w'(1) := \lim_{x \uparrow 1} w'(x) < 0$ , then there exists a unique  $x_2^* \in (0, 1)$  such that  $w'(x_2^*) = 0$ . If  $w'(1) \geq 0$ , then  $w(x)$  is monotonically increasing for all  $x$  and hence reaches its maximum at  $x_2^* = 1$ .

For the second step (regarding  $x_1^*$ ) we show first that  $w(x_2^*) > w'(x_2^*)$ . If  $w'(1) < 0$ , then  $w'(x_2^*) = 0$  and  $w(x_2^*) > 0$ , and hence  $w(x_2^*) > w'(x_2^*)$ . If  $w'(1) \geq 0$ , then  $x_2^* = 1$  and

$$w(x_2^*) = w(1) = y(1) > y(1) + y'(1) = w'(1) = w'(x_2^*),$$

because  $y'(1) < 0$ . In addition,

$$w(x_2^*) = x_2^* y(x_2^*) \leq y(x_2^*) < y(0) = w'(0).$$

Thus we find that

$$w'(x_2^*) < w(x_2^*) < w'(0),$$

and hence there exists a unique  $x_1^*$  with  $0 < x_1^* < x_2^*$  such that  $w'(x_1^*) = w(x_2^*)$ .  $\parallel$

We remark that both conditions can be much weakened. Regarding condition R1 (global monotonicity) it is sufficient that the  $y$ -curve is negatively sloped around  $x_1^*$  and  $x_2^*$ . Similarly, condition R2 (global concavity) need only hold locally for all  $x \in (x_0, 1)$ , where  $0 \leq x_0 < x_1^*$ , and this is in fact what we shall use in our estimation. (The choice  $x_0 = 0.4$  appears to be satisfactory.) In fact, it is possible to choose the  $y$ -curve such that it gives us maximum flexibility over the slopes at  $x_1^*$  and  $x_2^*$ , because all we need is that there exists a pair  $(x_1^*, x_2^*)$  with  $0 < x_1^* < x_2^* \leq 1$ , such that  $w(x)$  has a unique maximum at  $x_2^*$  and  $w(x) + (1 - x)w(x_2^*)$  has a unique maximum at  $x_1^*$ . The  $y$ -curve may thus have kinks. We shall use this generalization later when we discuss the robustness of the specification.

## 2.2 Implications

Theorem 1 has certain implications for the optimal service strategy  $(x_1^*, x_2^*)$ ; see also Gale (1971).

**Theorem 2 (implications):** If conditions R1 and R2 are satisfied, then the following relations must hold for the optimal strategy  $(x_1^*, x_2^*)$ :

- (a)  $x_1^* < x_2^*$ ,
- (b)  $y(x_1^*) > y(x_2^*)$ ,
- (c)  $w(x_1^*) < w(x_2^*)$ , and
- (d)  $w(x_2^*) - w(x_1^*) < (x_2^* - x_1^*)w(x_2^*)$ .

*Proof.* Statement (a) is implied in the proof of Theorem 1. Statement (b) then follows since  $y$  is a decreasing function. Statement (c) follows from the fact that if  $(x_1^*, x_2^*)$  is the optimal service strategy, then this strategy should lead to a higher probability of winning the point than the strategy  $(x_1^*, x_1^*)$  (always using the first service), that is,

$$w(x_1^*) + (1 - x_1^*)w(x_2^*) > w(x_1^*) + (1 - x_1^*)w(x_1^*),$$

and this simplifies to (c). Similarly, the optimal strategy  $(x_1^*, x_2^*)$  should lead to a higher probability of winning the point than the strategy  $(x_2^*, x_2^*)$  (always using the second service). This implies

$$w(x_1^*) + (1 - x_1^*)w(x_2^*) > w(x_2^*) + (1 - x_2^*)w(x_2^*)$$

which simplifies to (d).  $\parallel$

Relations (a) and (b) say that the first service should be more difficult than the second in two senses: (a) it is less often in, and (b) if it is in it is more likely to win the point. Relation (c) is less obvious. It says that the probability of winning a point at the first (difficult) service should be smaller than the probability of winning a point on the second (easier) service.

As a preliminary exercise, we calculate the observed frequencies for each player in each match using the data described in Section 5.1. If a player serves optimally, he or she should satisfy the four implications stated in Theorem 2. In fact, these implications are often not satisfied; see Table 1. The condition

Table 1: Empirical verification of the four consistency conditions in Theorem 2

	(a)	(b)	(c)	(d)	Total
Men	1.00	0.91	0.78	0.80	0.59
Women	0.98	0.72	0.77	0.64	0.42

$x_1 < x_2$  appears to be almost always satisfied, which means that almost all players take more risk on their first service than on their second service (as they should). However, this additional risk does not necessarily translate into higher productivity: the condition  $y(x_1) > y(x_2)$  is only satisfied for 91% of the men and 72% of the women. Condition (c) requires that  $(x_1, x_2)$  is a better service strategy than  $(x_1, x_1)$ , but this is only true for 77–78% of the players. Condition (d) requires that  $(x_1, x_2)$  is a better service strategy than  $(x_2, x_2)$ , but this is only true for 80% of the men and 64% of the women. For only 59% of the men and 42% of the women are all four consistency requirements satisfied.

It therefore seems that for many players the probability of winning a point can be increased by making small changes to their service strategy. Indeed, since the four consistency conditions are necessary but not sufficient for an optimal strategy (and hence even if all four conditions are satisfied a player may not follow the optimal strategy), the actual deviations from the optimum will be even larger. This conclusion, however, is too hasty and simplistic. For one, there will be measurement error: we are interested in probabilities but we observe relative frequencies, and this might lead to smaller deviations from the optimum. A more careful statistical analysis is required.

### 2.3 Other influences

It would be naive to believe that the  $y$ -curve only depends on  $x$ . In fact, both  $x$  and  $y$  will depend on a  $p \times 1$  vector  $\omega$  representing speed, direction, spin, concentration, (mental) effort to determine the optimal strategy, emotions, et cetera. We define an auxiliary vector  $v = (v_1, \dots, v_{p-1})'$  such that  $(x, v)$

and  $\omega$  are in one-to-one correspondence, and we write  $y = y(x, v)$ . Suppose we have only one service. Then we should maximize  $w(x, v) := x \cdot y(x, v)$  with respect to  $x$  and  $v$ . The first-order conditions are

$$y(x, v) + x \frac{\partial y(x, v)}{\partial x} = 0, \quad \varphi_h(x, v) := \frac{\partial y(x, v)}{\partial v_h} = 0 \quad (h = 1, \dots, p-1).$$

For given  $x$ , assume that the equations  $\varphi_h(x, v) = 0$  have a unique solution  $v = \zeta(x)$ , so that each  $\varphi_h(x, \zeta(x)) = 0$  for all  $x$ . At the optimum  $(x^*, v^*)$  we have  $\varphi_h(x^*, v^*) = 0$ , and hence

$$\begin{aligned} \left. \frac{\partial y(x, \zeta(x))}{\partial x} \right|_{x=x^*} &= \left. \frac{\partial y(x, \zeta(x^*))}{\partial x} \right|_{x=x^*} + \sum_{h=1}^{p-1} \varphi_h(x^*, v^*) \left. \frac{\partial \zeta_h(x)}{\partial x} \right|_{x=x^*} \\ &= \left. \frac{\partial y(x, \zeta(x^*))}{\partial x} \right|_{x=x^*}. \end{aligned}$$

The same reasoning applies when there are two services. We conclude that we may think of the  $y$ -curve as  $y(x, \zeta(x))$ , where  $v = \zeta(x)$  is chosen *optimally* for given  $x$ . Obviously some assumption is required on  $v$ , since no data are available on either  $v$  or  $\omega$ . The particular assumption of choosing  $v$  optimally implies a ‘conservative’ estimate of the efficiency. The same applies to the identification and estimation of  $\lambda$  (the curvature parameter of the  $y$ -curve), as we shall see in Section 6: no data and a conservative approach. We shall come back to the effects of the conservative approach when we discuss our efficiency results in Section 7.1.

In fact, it is not necessary that all  $p-1$  components of  $v$  are chosen optimally. We may partition  $v = (v_1, v_2)$ , where  $v_1$  is adjusted optimally (spin, direction) and  $v_2$  does not vary with respect to  $x$  (concentration).

## 2.4 Efficiency

In a given match between two players the probability that the server wins a point is given by (1):

$$p = x_1 y_1 + (1 - x_1) x_2 y_2. \quad (3)$$

Whether or not  $(x_1, y_1)$  and  $(x_2, y_2)$  are the *optimal* probabilities, the  $y$ -curve will pass through these two points. If the  $y$ -curve were linear then the two points would determine the curve, but we shall see that linearity is too restrictive. Suppose therefore that the  $y$ -curve depends on one (or more) curvature parameter  $\lambda$ . Given the  $y$ -curve, we obtain the optimal strategy  $(x_1^*, x_2^*)$  and the corresponding function values  $y_1^* = y(x_1^*)$  and  $y_2^* = y(x_2^*)$ , all of which depend on  $\lambda$ . The maximum probability of winning a point is thus

$$p^* = x_1^* y_1^* + (1 - x_1^*) x_2^* y_2^*, \quad (4)$$

and we define the efficiency of the server in this match as

$$\text{eff} := p/p^*, \quad (5)$$

which is a number between zero and one. The closer  $\text{eff}$  is to one, the higher is the efficiency. Note that the efficiency differs per player.

A natural question is whether we should measure efficiency at point level (as we propose) or at a higher level of aggregation: game, set, match, or even tournament or season. If we assume that points are independent and identically distributed, then the only difference between levels of aggregation is the counting system. For example, we may find that  $(p, p^*) = (0.65, 0.66)$  at point level, which translates to game winning probabilities  $(g, g^*) = (0.83, 0.85)$  (see Klaassen and Magnus, 2003), so that the inefficiency measure increases from  $1 - p/p^* = 1.5\%$  to  $1 - g/g^* = 1.9\%$ . If a game is defined differently, then the inefficiency measure will change, even if players behave in the same way. The inefficiency at a tiebreak will be measured differently than at a regular game; the inefficiency at a best-of-three match will be measured differently than at a best-of-five match. This seems undesirable. Hence we define efficiency at point level.

One might also argue that we should measure inefficiency in absolute terms  $(p^* - p)$  rather than in relative terms  $(1 - p/p^*)$ , as we do. We believe the relative definition is more appropriate. For example, for  $(p, p^*) = (0.9, 1.0)$  and  $(p, p^*) = (0.1, 0.2)$ , respectively, we find  $p^* - p = 10\%$ -points in both cases, but  $1 - p/p^* = 10\%$  in the first case and  $50\%$  in the second case, and the relative numbers seem intuitively more reasonable indications of inefficiency. All this relates to the *measurement* of (in)efficiency. Another matter is the *impact* of inefficiency. This will be discussed in Section 7.2.

In order to compute the efficiency measure (5) we need the probabilities  $(x_1, y_1)$  and  $(x_2, y_2)$  and the curvature parameter  $\lambda$ . Of course, we do not know these probabilities and this curvature. Their estimation is taken up in Sections 5 and 6. But first we shall discuss the specification of the  $y$ -curve.

### 3 Functional form for $y(x)$

It will be convenient to specify a functional form for the  $y$ -curve. The simplest specification is a linear function. This, however, does not work well in practice. For example, it forces  $x_1^* \leq 1/2$ , which is not realistic since the observed frequencies for  $x_1$  are 59.5% for men and 61.6% for women. Some curvature is required. We propose the following simple nonlinear function:

$$y(x) = \frac{\alpha - x^\lambda}{\tau}, \quad \lambda > 0. \quad (6)$$

Despite its simple form, this specification is already quite flexible over the relevant range  $[0.4, 1.0]$ , but it may still be too restrictive. We shall see

in Sections 6.3 and 7.1 that a more flexible generalization of (6) does not change the results. In fact, the generalization even strengthens our main result. Hence, (6) suffices.

The two regularity conditions R1 and R2 are satisfied if and only if

$$1 \leq \alpha \leq \tau + x_0^\lambda, \quad (7)$$

where  $0 \leq x_0 < x_1^*$ . The proposed  $y$ -curve allows  $y$  to be either concave or convex (or linear) on the  $(0, 1)$  interval, depending on whether  $\lambda > 1$  or  $0 < \lambda < 1$  (or  $\lambda = 1$ ). Given (6) we obtain  $w(x) = x(\alpha - x^\lambda)/\tau$ . If  $\alpha < \lambda + 1$ , then  $w'(1) < 0$  and  $x_2^*$  is obtained from the first-order condition

$$w'(x_2^*) = \frac{\alpha - (\lambda + 1)x_2^{*\lambda}}{\tau} = 0,$$

so that  $x_2^{*\lambda} = \alpha/(\lambda + 1)$ . If  $\alpha \geq \lambda + 1$ , then  $w'(1) \geq 0$  and  $x_2^* = 1$ . Hence,

$$x_2^{*\lambda} = \min\left(\frac{\alpha}{\lambda + 1}, 1\right). \quad (8)$$

Given  $x_2^*$ , we find  $x_1^*$  from the equation  $w'(x_1^*) = w(x_2^*)$ , that is

$$\alpha - (\lambda + 1)x_1^{*\lambda} = x_2^*(\alpha - x_2^{*\lambda})$$

from which we solve

$$x_1^{*\lambda} = x_2^{*\lambda} \left(1 - \frac{\lambda}{\lambda + 1} x_2^*\right). \quad (9)$$

From  $(x_1^*, x_2^*)$  we immediately obtain  $y_1^* = y(x_1^*)$  and  $y_2^* = y(x_2^*)$ .

The geometry of the solution is illustrated in Figure 1. On the horizontal axis we find the probability  $x$  that the service is in; on the vertical axis we find the conditional probability  $y(x)$  that server wins the point if the service is in, the unconditional probability  $w(x) = x \cdot y(x)$  that server wins the point on this service, and the derivative  $w'(x)$ . The optimal service strategy is found from Equations (8) and (9) which yields  $x_1^* = 60.4\%$  and  $x_2^* = 84.3\%$ . If the optimal strategy is employed, the probability of winning a point on service is  $p^* = 63.9\%$ , the probability of a double fault is  $6.2\%$ , and the probability of winning a service game is  $81.1\%$ .

Suppose we have observed  $x_1$ ,  $x_2$ ,  $y_1$ , and  $y_2$  for the server in a given match. This gives us two points  $(x_1, y_1)$  and  $(x_2, y_2)$  on the curve. For each  $\lambda$  we can solve  $\tau$  and  $\alpha$  from the two equations

$$\frac{\alpha - x_1^\lambda}{\tau} = y_1, \quad \frac{\alpha - x_2^\lambda}{\tau} = y_2. \quad (10)$$

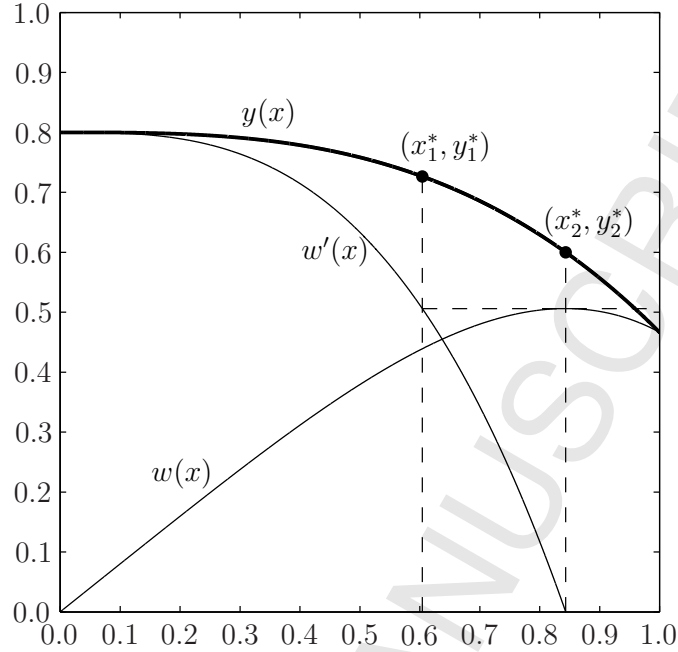


Figure 1: The  $y$ -curve as a function of  $x$  and the induced optimal service strategy (for  $\alpha = 2.4$ ,  $\tau = 3.0$ ,  $\lambda = 3.0$ ).

The solution reads

$$\tau = \frac{x_2^\lambda - x_1^\lambda}{y_1 - y_2}, \quad \alpha = \frac{y_1 x_2^\lambda - y_2 x_1^\lambda}{y_1 - y_2}. \quad (11)$$

Inserting (11) into (6) then gives

$$y(x) = \frac{y_1 x_2^\lambda - y_2 x_1^\lambda}{x_2^\lambda - x_1^\lambda} - \frac{y_1 - y_2}{x_2^\lambda - x_1^\lambda} \cdot x^\lambda. \quad (12)$$

Hence, in order to compute the  $y$ -curve and the optimal service strategy, we need to estimate  $\lambda$  and the probabilities  $(x_1, y_1)$  and  $(x_2, y_2)$ ; that is, the probabilities actually employed by the player.

## 4 Measurement error and random effects

In (12) we see that the  $y$ -curve depends on  $x$  (of course) and on five parameters: four probabilities  $x_1$ ,  $x_2$ ,  $y_1$ ,  $y_2$ ; and a curvature parameter  $\lambda$ , where the parameters may be match- and player-specific. It is not so clear how the curvature should be estimated, but the four probabilities can be estimated straightforwardly using relative frequencies. We shall see in this section that



this does not provide us with useful estimates of the slope (let alone of the curvature) of the  $y$ -curve, unless we correct for measurement error and use random effects.

Using the data described in Section 5.1 we calculate for both players  $i$  and  $j$  in each match the observed relative frequencies (over the whole match)  $f_{ij}(x_1)$  and  $f_{ij}(y_1)$  for the first service, and  $f_{ij}(x_2)$  and  $f_{ij}(y_2)$  for the second service, corresponding to the (unobserved) probabilities  $x_1$ ,  $y_1$ ,  $x_2$ , and  $y_2$ . This gives two points  $(f_{ij}(x_1), f_{ij}(y_1))$  and  $(f_{ij}(x_2), f_{ij}(y_2))$  for player  $i$  and also two points  $(f_{ji}(x_1), f_{ji}(y_1))$  and  $(f_{ji}(x_2), f_{ji}(y_2))$  for player  $j$ , hence four points per match. These points are plotted in Figure 2. In the left panel (men's singles) we see two point clouds centered around

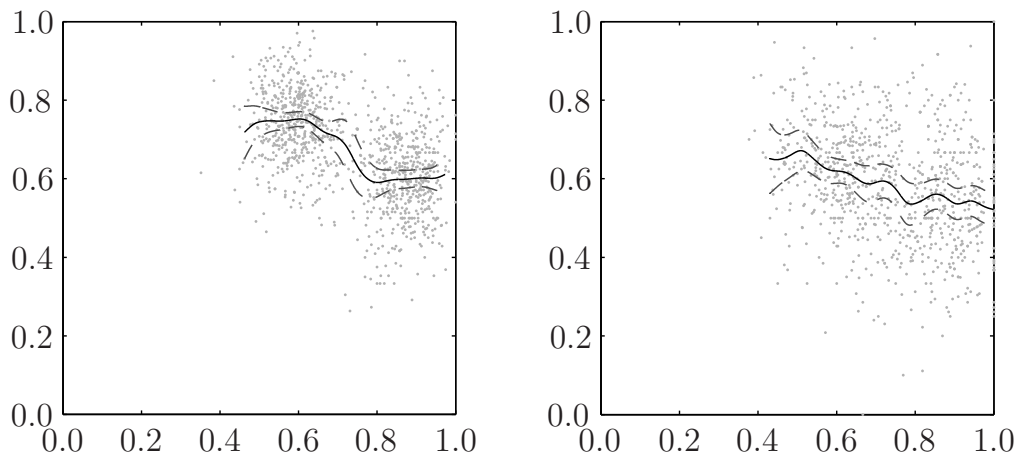


Figure 2: Empirical  $(x, y(x))$  plot based on relative frequencies, match data (men left, women right).

$(\bar{f}(x_1), \bar{f}(y_1)) = (0.59, 0.74)$  and  $(\bar{f}(x_2), \bar{f}(y_2)) = (0.86, 0.59)$  and a nonparametric regression line (with precision bounds). The curve is flat precisely where we are interested in the  $y$ -curve, namely around  $\bar{f}(x_1)$  and  $\bar{f}(x_2)$ . We know that the slope is negative at  $x_1$  and  $x_2$ , but we can not estimate it. If the slope can not be estimated, then to estimate curvature around these points will be fruitless. For the women (right panel) the points center around  $(0.62, 0.63)$  (first service) and  $(0.86, 0.53)$  (second service). There is some weak indication of a downward slope around  $f(x_1)$ , but not around  $f(x_2)$ . The reason that the slopes appear flat could be that we do not have enough data to obtain precise estimates given the little variation in  $x_1$  and  $x_2$  across players, but it could also be the bias caused by measurement error in the frequencies as a proxy for the probabilities  $x_1$  and  $x_2$ , or it could be that relevant variables have been omitted.

To assess the impact of omitted variables, we consider set averages (rather than match averages), so that we can control for all variables that are con-

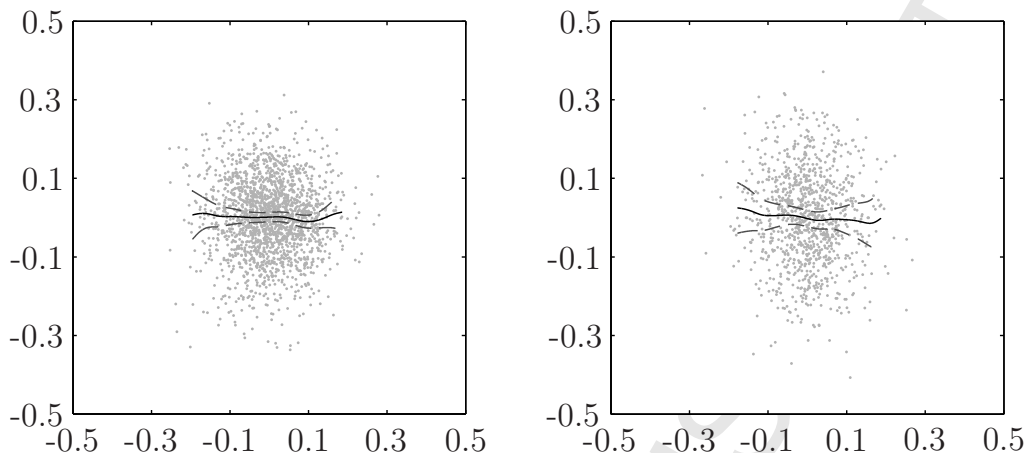


Figure 3: Empirical  $(x, y(x))$  plot based on relative frequencies (centered), first service, set data (men left, women right).

stant over sets by including fixed effects. Thus we compute  $(f_{ij}^{(s)}(x_1), f_{ij}^{(s)}(y_1))$  and  $(f_{ij}^{(s)}(x_2), f_{ij}^{(s)}(y_2))$  as the relative frequencies of the first and second services, observed in set  $s$  of a match of player  $i$  against player  $j$ . We then compute the averages over  $s$ :  $(\bar{f}_{ij}(x_1), \bar{f}_{ij}(y_1))$  and  $(\bar{f}_{ij}(x_2), \bar{f}_{ij}(y_2))$  and the centered values

$$\tilde{f}_{ij}^{(s)}(x_1) := f_{ij}^{(s)}(x_1) - \bar{f}_{ij}(x_1), \quad \tilde{f}_{ij}^{(s)}(y_1) := f_{ij}^{(s)}(y_1) - \bar{f}_{ij}(y_1)$$

for the first service, and similarly for the second service. The advantage of considering the centered relative frequencies lies in the fact that locational differences across server-service combinations have been removed, so that only the slopes around the first and second services remain. In Figure 3 we plot  $(\tilde{f}_{ij}^{(s)}(x_1), \tilde{f}_{ij}^{(s)}(y_1))$  for the first service for both men and women; the results for the second service (not reported here) are similar. The slopes are still very flat. What could be the reason? One possibility is that there are omitted variables, but that centering removes too much information. We could remedy this by using random effects instead of fixed effects. Another possibility is that omitted variables is not the main problem, but that the  $x$ -variables are measured with error. This would lead to a severe bias toward zero. (In general, measurement error in  $x$  leads to bias toward zero. The within transformation in panel models makes this bias even stronger, and the centering above is essentially a within transformation.) We now demonstrate that measurement error is indeed a serious factor here.

The sample standard deviation for the match frequencies across players is 6.4% for the first service in and 6.1% for the second service in. How much of this is noise due to the fact that frequencies differ from probabilities? In

other words, how much is due to measurement error? Let  $t_{1,ij}$  denote the number of first services and  $f_{1,ij}$  the proportion that is in. (For the men we have on average  $t_{1,ij} = 115$  and  $f_{1,ij} = 0.59$ . Similarly, for the second service,  $t_{2,ij} = 47$  and  $f_{2,ij} = 0.86$ .) This produces an approximate variance of the deviation of the frequencies from the probabilities of  $f_{1,ij}(1-f_{1,ij})/t_{1,ij}$  for the first service. Averaging over all players and taking the square root defines our measurement error indicator, which is 4.7% (for the second service, 5.1%). Hence, the noise in the relative frequencies for a single player is a substantial part of the variation of the frequencies over players. So, measurement error is relatively large and this explains to a large extent the flatness of the curves. (The same is true in the women's singles: 9.4% and 8.8% versus 6.2% and 7.0%.)

This is the situation at match level. To examine the possible benefit of going from match to set data we analyze the within-player variance of the set frequencies, averaged over all players. For the men the square root of this average is 9.3% for the first service, and 10.4% for the second service. The measurement error indicator for the first service is now the square root of the average  $f_{1,ij}^{(s)}(1-f_{1,ij}^{(s)})/t_{1,ij}^{(s)}$  over all sets and players, which is 8.9%. (For the second service it is 9.6%.) Hence, for a given player measurement error can almost fully explain the variation in frequencies across sets, which means that the variation in the underlying probabilities across sets is close to zero. (For the women the total variation 9.3% and 11.1% is also explained by the measurement error indicators 9.2% and 10.2%.) This again shows that moving from match to set data is not useful.

We conclude that we need to correct for measurement error in  $x$ , that we must use random effects in order to avoid efficiency loss associated with fixed effects, and that we may focus on match data. This is precisely what we shall do.

## 5 Estimation of the key probabilities

The previous section showed the importance of taking measurement error seriously and of using random effects. Our proposed method estimation is a two-step procedure motivated by the fact that we can estimate  $(x_1, x_2, y_1, y_2)$  and  $\lambda$  separately. The difficulty in estimating the curvature thus plays no role in this section, and will only be taken up in Section 6.

In Section 2 we highlighted the importance of the two probabilities  $x$  (probability that service is in) and  $y$  (conditional probability that server wins the point if service is in). For player  $i$  against player  $j$  the key probabilities are given by

$$z_{ij} := (x_{1,ij}, x_{2,ij}, y_{1,ij}, y_{2,ij})',$$

where  $x_{1,ij}$  denotes the probability that  $i$  (playing against  $j$ ) serves his or her

first service in, and similarly for the other three probabilities. We cannot observe these four key probabilities, but we can observe the associated relative frequencies, denoted by  $f_{ij}$ , and the associated number of observations underlying each of the relative frequencies, denoted by  $t_{ij}$ . If the  $t_{ij}$ -dimension were sufficiently large, then the probabilities in  $z_{ij}$  would be well approximated by the relative frequencies in  $f_{ij}$ , and there would be no need to model  $z_{ij}$  further. In that case we could jump directly to Section 6 and treat  $(z_{ij}, z_{ji})$  as given. However, the  $t_{ij}$ -dimension is not large, especially for the second service where (in our data set) it gets as low as 14 (men) and 5 (women). Hence important efficiency gains can be achieved by modeling  $z_{ij}$ .

## 5.1 The data

Our data consist of singles matches played at Wimbledon during 1992–1995: 508 matches for the men and 508 matches for the women. For each of these matches we know the two players, their rankings at the beginning of the tournament, and the match result. For almost half of the matches (258 matches in the men’s singles and 223 matches in the women’s singles) we know the complete sequence of points. The data are described in detail in Magnus and Klaassen (1999a).

The reason that we do not have detailed data on all matches played during the four years is that only matches played on one of the five ‘show courts’ (Centre Court and Courts 1, 2, 13, and 14) have been recorded. Typically, matches involving the most important players are scheduled on the show courts, and this causes an under-representation in the data set of matches involving weaker players. All results in this section have been corrected for this selection problem by weighting the matches by the inverses of the sampling percentages. The weighting procedure is discussed in detail in Magnus and Klaassen (1999b).

We shall use sixteen summary statistics per match, thus not the complete sequence of points; in fact we only use the data in Sections 4 and 5, and in Table 1. (In Section 4 we use match *and* set data, but from now on we only use match data.) For player  $i$  serving against player  $j$  we use the relative frequencies  $f_{ij}$  and the associated numbers of observations  $t_{ij}$ ; and the same for player  $j$  serving against  $i$ . For example, suppose player  $i$  serves 100 times against player  $j$ . Of the 100 first services, 60 are in and 40 are a fault; if the first service is in, player  $i$  wins the point in 44 cases and loses it in 16 cases. Of the 40 second services, 35 are in and 5 are a fault (double fault); if the second service is in, player  $i$  wins the point in 21 cases and loses it in 14 cases. This information allows us to compute  $f_{ij} = (0.600, 0.875, 0.733, 0.600)'$  and  $t_{ij} = (100, 40, 60, 35)'$ .

We do not have access to more matches of which all relevant summary statistics are known. If we had, this would have decreased the estimation

uncertainty from the GMM procedure described below. We could have increased the number of matches where we only require the rankings (currently 508 matches) by including Wimbledon singles matches from 1996 onwards. This would, however, have added nothing and it would also have raised the question of parameter instability over time.

## 5.2 The measurement equation

In what follows we assume that matches are independent (even when servers occur in several matches), and that points played in one match are independent and identically distributed (i.i.d.). The independence of matches seems a reasonable assumption, but the i.i.d. assumption of points within one match is a hot issue (not only in tennis), and was analyzed by Klaassen and Magnus (2001). They conclude that points in tennis are neither independent nor identically distributed, but that the deviation is small (though statistically significant) and that therefore the i.i.d. assumption will still be reasonable in many specific directions. In addition, in our case, we do not use the points themselves but summary statistics (averages) so that any possible harm caused by the wrong assumption is much reduced. Finally, any remaining error caused by the i.i.d. approximation will result in an overestimation of the inefficiency, which would strengthen rather than weaken our main conclusion that the inefficiency is small.

In measuring the (nonobservable) random probability  $z_{ij}$  with the (observable) random relative frequency  $f_{ij}$ , a measurement error occurs:

$$f_{ij} = z_{ij} + \eta_{ij}. \quad (13)$$

Since each  $t_{ij}^{(k)} f_{ij}^{(k)}$  ( $k = 1, \dots, 4$ ) follows a binomial distribution conditional on  $z_{ij}$ , we obtain

$$E(\eta_{ij} | z_{ij}) = 0, \quad \text{var}(\eta_{ij} | z_{ij}) = \Delta_{ij}, \quad (14)$$

where the diagonal elements of  $\Delta_{ij}$  are given by

$$\Delta_{ij}^{(k)} = \frac{1}{t_{ij}^{(k)}} z_{ij}^{(k)} (1 - z_{ij}^{(k)}) \quad (k = 1, \dots, 4). \quad (15)$$

We assume that the off-diagonal elements of  $\Delta_{ij}$  are zero, which is reasonable, because the components of  $\eta_{ij}$  represent ‘pure noise.’ In addition and for the same reason, we assume that  $E(\eta_{ji} | z_{ij}) = 0$  and that  $\eta_{ij}$  and  $\eta_{ji}$  are uncorrelated (conditional on  $z_{ij}$  and  $z_{ji}$ ).

### 5.3 A model for $z_{ij}$

Since the four probabilities in  $z_{ij}$  will be correlated and, in addition,  $z_{ij}$  and  $z_{ji}$  will be correlated, we need to model an eight-dimensional distribution. The probabilities depend on the characteristics of the server and his or her opponent. Some of these characteristics are observable, some are not. An important characteristic is a player's 'quality.' In tennis a player's quality is partly measured by his or her official ranking. Although this is not a perfect measure of quality (for example, it does not account for 'form of the day'), the ranking contains important information about the key probabilities, which we want to exploit.

Let  $\text{RANK}_i$  denote the ranking of player  $i$  at the moment of the tournament. Direct use of the ranking is not satisfactory, because quality in tennis is a pyramid: the difference between the top two players (ranked 1 and 2) is generally larger than between two players ranked 101 and 102. As motivated in Klaassen and Magnus (2001),  $\text{RANK}_i$  is transformed into a smooth version of the 'expected round' by defining

$$r_i = 8 - \log_2(\text{RANK}_i).$$

For example, if  $\text{RANK}_i = 4$  then  $r_i = 6.00$ , and if  $\text{RANK}_i = 3$  then  $r_i = 6.42$ , indicating that both are expected to lose in round 6 (the semifinal in grand slam tournaments like Wimbledon). The additional 0.42 indicates that the number 3 ranked player is somewhat better than the player ranked 4.

Apart from the ranking there are also quality components that we do not observe, such as form of the day, special ability on the court surface on which the match is played (grass for Wimbledon), and fear against a specific opponent. In addition, each player has his or her own style of play, a characteristic that may also affect the four key probabilities. None of these features is observed, but we correct for them by including two four-dimensional vectors  $\varepsilon_{ij}$  and  $\varepsilon_{ji}$  of unobservable effects.

We model the key probabilities in a given match between  $i$  and  $j$  as an eight-dimensional vector  $(z'_{ij}, z'_{ji})'$ , where

$$z_{ij} = \mu + r_i\beta_S + r_j\beta_R + \varepsilon_{ij}. \quad (16)$$

The linearity assumption underlying this equation is acceptable when we use  $r_i$  as a regressor, but not when we use  $\text{RANK}_i$ , as preliminary nonparametric regressions show. The vectors  $\beta_S$  and  $\beta_R$  contain the effects of the server's and receiver's rankings on each probability in  $z_{ij}$ , respectively. For the errors we assume homoskedasticity, which implies

$$\begin{pmatrix} \varepsilon_{ij} \\ \varepsilon_{ji} \end{pmatrix} \sim \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2 & \Sigma_1 \end{pmatrix} \right). \quad (17)$$



It is convenient to center  $r_i$  and  $r_j$ . The zero mean standardization of the errors then implies that  $\mu$  is equal to the mean of  $z_{ij}$ .

The assumption regarding the errors implies that  $\varepsilon_{ij}$  is a random (instead of fixed) effect. Because the rankings  $\text{RANK}_i$  and  $\text{RANK}_j$  are determined before the tournament starts, we assume that there is no correlation between  $\varepsilon_{ij}$  and the rankings. To account for the correlations between the key probabilities for a given server we include a variance matrix  $\Sigma_1$ . In addition, there may be correlations between the key probabilities across the two servers in a given match. For example, if player  $i$  is better on the present court surface (grass, say) than his or her ranking indicates, or if player  $i$  usually performs well against  $j$ , then especially the two last components of  $\varepsilon_{ij}$  and  $\varepsilon_{ji}$ , which concern the winning probabilities, are negatively correlated. Thus we introduce a covariance matrix  $\Sigma_2$ . Since the moments do not depend on the specific players ( $i$  and  $j$ ), we have, for example,  $\text{cov}(\varepsilon_{ij}^{(3)}, \varepsilon_{ji}^{(4)}) = \text{cov}(\varepsilon_{ji}^{(3)}, \varepsilon_{ij}^{(4)})$ , so that  $\Sigma_2$  is symmetric.

In summary, we have a model for the whole vector  $(z'_{ij}, z'_{ji})'$  of probabilities that govern the two service series within a match. These probabilities are restricted to be constant within a match. However, in all other respects, the probabilities are unrestricted, as we allow for full heterogeneity across players (servers and receivers) and for possible correlations within a match. There are four free parameters in each of  $\mu$ ,  $\beta_S$ , and  $\beta_R$ , and ten free parameters in each of  $\Sigma_1$  and  $\Sigma_2$ , in total 32 parameters.

## 5.4 Moment conditions

We employ the generalized method of moments (GMM) to estimate the 32 parameters. Hence we need to derive moment conditions, taking into account that the lengths of the service series across players vary, as do the lengths of the four service series for each player (in  $t_{ij}$ ). We define

$$B = (\mu : \beta_S : \beta_R), \quad r'_{ij} = (1, r_i, r_j),$$

and write (16) in matrix form as

$$z_{ij} = Br_{ij} + \varepsilon_{ij}. \quad (18)$$

Combining (18) with (13) then gives

$$f_{ij} = Br_{ij} + \varepsilon_{ij} + \eta_{ij},$$

where the fact that  $E(\eta_{ij} | z_{ij}) = E(\eta_{ji} | z_{ij}) = 0$  implies that  $\varepsilon_{ij}$  is uncorrelated with  $\eta_{ij}$  and  $\eta_{ji}$ . From the conditional moments (14) we find  $E(\eta_{ij}) = 0$  and  $\text{var}(\eta_{ij}) = E(\Delta_{ij})$ . Since  $z_{ij}$  (and hence  $\Delta_{ij}$ ) cannot be observed, we define the diagonal matrix  $\tilde{\Delta}_{ij}$  whose  $k$ -th diagonal element is given by

$$\tilde{\Delta}_{ij}^{(k)} = \frac{1}{t_{ij}^{(k)} - 1} f_{ij}^{(k)} (1 - f_{ij}^{(k)}) \quad (k = 1, \dots, 4), \quad (19)$$



and we note that  $E(\tilde{\Delta}_{ij} \mid z_{ij}) = \Delta_{ij}$  so that  $E(\tilde{\Delta}_{ij}) = E(\Delta_{ij})$ . (If  $t_{ij}^{(k)}$  equals one or zero, then (19) breaks down. This does not occur in our data set, because the minimum number of second services in is 14 (men) and 5 (women), and the minimum number of first services in is 31 (men) and 14 (women).) The following expectations are then implied:

$$\begin{aligned} E((f_{ij} - Br_{ij})r'_{ij}) &= 0 \\ \text{var}(f_{ij} - Br_{ij}) &= \Sigma_1 + E(\tilde{\Delta}_{ij}) \\ \text{cov}(f_{ij} - Br_{ij}, f_{ji} - Br_{ji}) &= \Sigma_2. \end{aligned} \quad (20)$$

The first moment is the usual least-squares orthogonality condition (12 restrictions), the second moment concerns the within-server variance (10 restrictions, because of the symmetry), and the third moment captures the correlation between the frequency vectors of the two servers in a match (10 restrictions).

## 5.5 Implementation

Let  $n_1$  denote the number of server series (twice the number of matches), the cross-section dimension of the panel. In our sample we have  $n_1 = 516$  in the men's singles and  $n_1 = 446$  in the women's singles. Because  $n_1$  is substantially larger than the elements of  $t_{ij}$  (especially the elements that concern the second service), our asymptotic justification is based on large  $n_1$  and finite  $t_{ij}$ . The moment conditions have thus been set up in a format that is standard in studies that rely on large- $n_1$  asymptotics.

The set of moment conditions (20) contains 32 different elements. Each server  $i$  in his or her match against  $j$  has exactly one observation of each element. Let  $h_{ij}$  denote the 32-dimensional vector containing all observations:

$$h_{ij} = \begin{pmatrix} \text{vec}((f_{ij} - Br_{ij})r'_{ij}) \\ \text{vech}((f_{ij} - Br_{ij})(f_{ij} - Br_{ij})' - \Sigma_1 - \tilde{\Delta}_{ij}) \\ \text{vech}((f_{ij} - Br_{ij})(f_{ji} - Br_{ji})' - \Sigma_2) \end{pmatrix},$$

where  $\text{vech}()$  denotes the half-vec operator stacking the nonrepeated elements of a symmetric matrix. Obviously,  $E(h_{ij}) = 0$  and one could use the sample average  $(1/n_1) \sum h_{ij}$  in a GMM procedure to estimate the parameters. There are however two sources of inefficiency that we wish to take account of. First, all players serve a different number of points, so that the precision of the frequencies varies. Taking an unweighted average across players does not take this into account. Second, for a given server the number of points varies across the key frequencies. To increase the efficiency of our estimates, we weight each element of  $h_{ij}$  by the number of observations used to compute

that element. Hence, we obtain a new moment vector

$$\tilde{h}_{ij} = \Omega_{ij} h_{ij},$$

where  $\Omega_{ij}$  is a diagonal  $32 \times 32$  matrix with the 32 weights on the diagonal. This leads to the following GMM objective function:

$$\min \left( \frac{1}{n_1} \sum \tilde{h}_{ij} \right)' W \left( \frac{1}{n_1} \sum \tilde{h}_{ij} \right),$$

where we minimize over all 32 parameters. We use the standard optimal weighting matrix  $W$ , that is, the inverse of the variance of the limiting distribution of  $\tilde{h}_{ij}$ . We begin with  $W = I_{32}$  to obtain a consistent estimate of the parameter vector, then we estimate  $W$  by  $\widehat{W}$ , and finally we minimize the objective function again using  $\widehat{W}$  instead of the identity matrix. This two-step procedure gives us consistent, asymptotically normal, and efficient estimates.

## 5.6 Estimation results

The estimates of  $\mu$ ,  $\beta_S$ ,  $\beta_R$ ,  $\Sigma_1$ , and  $\Sigma_2$ , together with their standard errors are reported in Table 2. The  $\mu$ -parameters give the results for the average player and are estimated very precisely. On average 59.5% of the first services are in for the men (61.6% for the women) and 86.4% of the second services (same for men and women). The similarity between men and women is remarkable. The scoring power of the service is of course rather different for men and women. If the first service is in then men score 74.0% (63.1% for women) on average, and if the second service is in they score 59.4% (52.6% for women).

The  $\beta_S$ -estimates show that the better a player is, the higher are the key probabilities. As expected, this is clearly true for  $y_1$  and  $y_2$  and (less strongly) also for  $x_1$  and  $x_2$ . The estimates of  $\beta_R$  show just the opposite.

The estimates in the variance matrix  $\Sigma_1$  are presented as standard deviations (for the four diagonal elements) and correlations (for the six off-diagonal elements). Of the ten estimated parameters in the covariance matrix  $\Sigma_2$ , seven are statistically and logically insignificant, and are set equal to zero. Yet we use 32 (rather than 25) moments at the estimation stage. This allows an overidentifying restrictions test (ORT), giving  $\text{ORT} = 11.77$  (0.11) for the men and  $\text{ORT} = 2.33$  (0.94) for the women ( $p$ -values in brackets), thus providing further support for restricting the seven parameters to zero. The three remaining estimates in  $\Sigma_2$  are correlations, also the estimate indicated by  $(y_1, y_1)$  (and similarly  $(y_2, y_2)$ ) as it measures the correlation between  $y_{1,ij}$  and  $y_{1,ji}$ . The three correlations are negative because if player  $i$  is ‘in form’ (relative to  $j$ ) as a server he or she is also likely to be ‘in form’ as a receiver.

Table 2: GMM parameter estimates of Equations (16) and (17)

Coefficients	Men		Women		
	Estimate	St. error	Estimate	St. error	
$\mu$	$x_1$	0.5947	0.0034	0.6157	0.0051
	$x_2$	0.8642	0.0031	0.8644	0.0047
	$y_1$	0.7403	0.0037	0.6308	0.0051
	$y_2$	0.5942	0.0044	0.5262	0.0077
$\beta_S$	$x_1$	0.0010	0.0015	0.0056	0.0029
	$x_2$	0.0028	0.0013	0.0108	0.0031
	$y_1$	0.0148	0.0017	0.0228	0.0025
	$y_2$	0.0110	0.0020	0.0144	0.0037
$\beta_R$	$x_1$	0.0007	0.0015	0.0042	0.0024
	$x_2$	-0.0011	0.0013	-0.0002	0.0025
	$y_1$	-0.0062	0.0017	-0.0198	0.0026
	$y_2$	-0.0088	0.0021	-0.0200	0.0039
$\Sigma_1$	$x_1$	0.0488	0.0049	0.0685	0.0059
	$x_2$	0.0317	0.0051	0.0505	0.0075
	$y_1$	0.0547	0.0042	0.0486	0.0083
	$y_2$	0.0540	0.0062	0.0754	0.0113
	$(x_2, x_1)$	0.6390	0.1540	0.2973	0.1343
	$(y_1, x_1)$	-0.0985	0.1226	-0.5923	0.1720
	$(y_2, x_1)$	0.0652	0.1377	0.2415	0.1782
	$(y_1, x_2)$	-0.2593	0.1671	-0.3052	0.1970
	$(y_2, x_2)$	0.2700	0.2112	0.1454	0.2308
	$(y_2, y_1)$	0.5752	0.1360	0.0919	0.2473
$\Sigma_2$	$(y_1, y_1)$	-0.1070	0.1547	-0.5843	0.3830
	$(y_2, y_2)$	-0.4106	0.2347	-0.3392	0.3889
	$(y_2, y_1)$	-0.2375	0.1319	-0.3370	0.2411

The correlations in the (not reported)  $25 \times 25$  variance matrix of the estimates are all small, so that the variables satisfy a high degree of orthogonality and there is no problem of multicollinearity. The estimates are all very plausible and the fact that all signs are the same for men and women underlines their significance.

## 6 Identification of $\lambda$

Now that we have estimated the distribution of  $(z_{ij}, z_{ji})$ , we know the distribution of  $(x_1, y_1)$  and  $(x_2, y_2)$  for server  $i$  in his/her match against  $j$ , but we do not yet know  $\lambda$ , a necessary ingredient for the estimation of the  $y$ -curve (12). The curvature parameter  $\lambda$  is essentially unidentified, because the  $y$ -

curve depends on three parameters and we only observe two points on each curve:  $(x_1, y_1)$  and  $(x_2, y_2)$ .

Thus we treat  $\lambda_{ij}$  as a random effect, in the same way as we have treated  $\varepsilon_{ij}$  in the estimation of  $z_{ij}$ , and assume

$$\lambda_{ij} \mid (z_{ij}, z_{ji}) \sim (\lambda, \sigma_\lambda^2). \quad (21)$$

We shall obtain ‘conservative’ estimates of  $\lambda$  and  $\sigma_\lambda$ , in line with our discussion in Section 2.3, in the sense that we shall solve

$$\max_{\lambda, \sigma_\lambda} \frac{1}{n} \sum E(\text{eff}_{ij}),$$

where  $\text{eff} = p/p^*$  is the efficiency defined in (5), and  $n$  denotes the number of players. We shall discuss the seriousness of this conservativeness in Section 7.1. We maximize the *average* of the expected efficiency (over the  $n$  players), because the elements over which we maximize are all positive, so that the average is equal to the mean absolute error and therefore takes account of the spread as well as the location. This measure is therefore preferable over the median (only location).

The (nontrivial) estimation procedure is described below. We shall see that the estimate of  $\sigma_\lambda$  is essentially zero, so that we may take  $\lambda_{ij} = \lambda$  (constant). Apparently we may assume that for each player the  $y$ -curve has the same curvature (but different height and slope).

We treat men and women separately. For both men’s singles and women’s singles we consider 508 matches, that is, *all* singles matches played at Wimbledon during 1992–1995. This gives  $n = 1016$  servers. For each match we require the rankings  $r_i$  and  $r_j$ , but not the summary statistics used in Section 5.

## 6.1 Monte Carlo

In Section 5 we estimated the 25 parameters in  $\mu$ ,  $\beta_S$ ,  $\beta_R$ ,  $\Sigma_1$ , and  $\Sigma_2$ , and also the  $25 \times 25$  variance matrix. We collect the 25 parameters in a vector  $\theta$ , denote its estimator by  $\hat{\theta}$ , the asymptotic variance matrix of  $\hat{\theta}$  by  $V$ , and its estimator by  $\hat{V}$ . Then, approximately,  $\hat{\theta} \sim N(\theta, V)$ .

The Monte Carlo procedure consists of two stages. In the first stage we draw, for  $r = 1, \dots, R$  ( $R = 50$ ), a vector  $\theta^{(r)}$  from the  $N(\hat{\theta}, \hat{V})$  distribution under the restriction that the matrix

$$\begin{pmatrix} \Sigma_1^{(r)} & \Sigma_2^{(r)} \\ \Sigma_2^{(r)} & \Sigma_1^{(r)} \end{pmatrix}$$

is positive definite. Hence we draw from a truncated normal distribution. The first stage of the Monte Carlo draw is independent of  $\lambda_{ij}$ .

In the second stage we simulate  $E(\text{eff}_{ij})$  by drawing  $(z_{ij}, z_{ji}, \lambda_{ij}, \lambda_{ji})$ , for each of the 508 matches and for each of the  $R$  draws of  $\theta$ . By (16) and (17) we have in each match

$$\begin{pmatrix} z_{ij} \\ z_{ji} \end{pmatrix} \sim \left( \begin{pmatrix} \mu^{(r)} + r_i \beta_S^{(r)} + r_j \beta_R^{(r)} \\ \mu^{(r)} + r_j \beta_S^{(r)} + r_i \beta_R^{(r)} \end{pmatrix}, \begin{pmatrix} \Sigma_1^{(r)} & \Sigma_2^{(r)} \\ \Sigma_2^{(r)} & \Sigma_1^{(r)} \end{pmatrix} \right).$$

Since the densities of the four frequencies in  $f_{ij}$  have the appearance of normal curves, it seems reasonable to assume normality of the distribution of  $(z_{ij}, z_{ji})$ , but truncated for two reasons. First, we must have  $x_0 \leq x_1 \leq 1$  and  $x_0 \leq x_2 \leq 1$ . In particular, the condition  $x_2 \leq 1$  must be imposed. Second, conditions R1 and R2 must hold, which is the case if and only if condition (7) holds. Using (11), the truncation thus depends on  $\lambda_{ij}$  because

$$1 \leq \frac{y_{1,ij} x_{2,ij}^{\lambda_{ij}} - y_{2,ij} x_{1,ij}^{\lambda_{ij}}}{y_{1,ij} - y_{2,ij}} \leq \frac{x_{2,ij}^{\lambda_{ij}} - x_{1,ij}^{\lambda_{ij}}}{y_{1,ij} - y_{2,ij}} + x_0^{\lambda_{ij}},$$

where we set  $x_0 = 0.4$ . We also assume normality of  $(\lambda_{ij}, \lambda_{ji})$ , but truncated because of the restriction  $\lambda > 0$ . We now draw  $S = 50$  ten-dimensional vectors of independent realizations  $(u_{ij}, u_{ji})$  from the uniform distribution defined on the  $[0, 1]$  interval. We transform them into  $S$  draws  $(z_{ij}, z_{ji}, \lambda_{ij}, \lambda_{ji})$  using the GHK procedure; see Hajivassiliou, McFadden and Ruud (1996). The resulting draws will satisfy regularity conditions R1 and R2, have positive  $\lambda$ , and all probabilities will be in the  $[0, 1]$  interval. We then compute  $\text{eff}_{ij}$  for each draw, and approximate  $E(\text{eff}_{ij})$  by taking the average over the  $S$  draws.

## 6.2 Estimation of $\lambda$

The result of the two-stage procedure is that we can calculate, for each of the  $R$  draws  $\theta^{(r)}$  and for each server, the expectation  $E(\text{eff}_{ij})$  as a function of  $\lambda$  and  $\sigma_\lambda$ , and hence also the average expected efficiency

$$\overline{\text{eff}}(\lambda, \sigma_\lambda) := \frac{1}{n} \sum E(\text{eff}_{ij}).$$

To estimate  $\lambda$  and  $\sigma_\lambda$ , we calculate the function  $\overline{\text{eff}}(\lambda, \sigma_\lambda)$  for different values of  $\lambda$  and  $\sigma_\lambda$ , where in each function calculation the draw of  $(z_{ij}, z_{ji}, \lambda_{ij}, \lambda_{ji})$  is based on the *same* draw of  $(u_{ij}, u_{ji})$ . Because of the concavity of the underlying functions, the function  $\overline{\text{eff}}(\lambda, \sigma_\lambda)$  will have a maximum in a suitably chosen parameter range.

The  $R$  maximizing parameter values have a mean (standard error) of  $\hat{\lambda} = 3.07$  (0.13) and  $\hat{\sigma}_\lambda = 0.002$  (0.003) for the men, and  $\hat{\lambda} = 3.77$  (0.35)

and  $\hat{\sigma}_\lambda = 0.005$  (0.008) for the women. Since  $\hat{\sigma}_\lambda$  is not significantly different from zero, we set  $\sigma_\lambda = 0$  and re-estimate  $\lambda$ . The  $R$  draws for  $\lambda$  have a mean (standard error) of  $\hat{\lambda} = 3.07$  (0.13) in the men's singles and  $\hat{\lambda} = 3.83$  (0.37) in the women's singles. (If we maximize the median instead of the mean of the expected efficiency, we obtain  $\hat{\lambda} = 3.06$  (0.13) for the men and  $\hat{\lambda} = 3.71$  (0.35) for the women.) The estimate of  $\lambda$  is significantly different for men and women, and is rather precise.

At the end of this admittedly complicated estimation procedure we thus have, for each of the  $R$  draws of  $\theta$ ,  $S$  feasible  $8 \times 1$  vectors  $(z_{ij}, z_{ji})$  for all 508 matches and one  $\lambda$ . In short, we have  $S = 50$  feasible  $9 \times 1$  vectors  $(z_{ij}, z_{ji}, \lambda)$ .

### 6.3 Sensitivity

A crucial element in this paper is the specification of the  $y$ -curve and the estimation of its curvature. It is therefore important to find out how sensitive our results are to (small) deviations from the chosen specification. In Table 3

Table 3: Sensitivity analysis of the curvature parameter  $\lambda$

$y$ -curve	Curvature	Men	Women
$C_0$	$\lambda$	3.0677 (0.1315)	3.7727 (0.3455)
$C_1$	$\lambda$	3.0553 (0.1246)	3.4338 (0.2751)
$C_2$	$\lambda_1$	3.1373 (0.1879)	3.5720 (0.2816)
	$\lambda_2$	2.8955 (0.2538)	2.9366 (0.6754)

we compare the  $\lambda$ -estimates from three different specifications of the  $y$ -curve. The curve  $C_0$  is our preferred specification (6), while the curves  $C_1$  and  $C_2$  are based on the idea that the most important aspect of the  $y$ -curve is the area around  $x_1$  and  $x_2$ . Suppose we allow two  $y$ -curves: one around  $x_1$  and one around  $x_2$ . The two curves are both power curves like (6), but with different sets of parameters  $(\alpha, \tau, \lambda)$ . Both curves are only specified locally, that is, between  $x_1$  and  $x_1^*$  and between  $x_2$  and  $x_2^*$ , and are essentially unrestricted elsewhere. In the curve  $C_1$  we set  $\lambda_1 = \lambda_2$ , while in the curve  $C_2$  we do not restrict  $\lambda$ . We see that for the men the estimate of  $\lambda$  is hardly affected; the value  $\hat{\lambda} = 3.1$  appears to be very stable. For the women slightly less so. Still, there is no statistical support for rejecting the specification (6).

## 7 Efficiency

For both men and women the previous section provides  $S$  feasible  $9 \times 1$  vectors  $(z_{ij}, z_{ji}, \lambda)$  for each of the 508 matches (1016 players). Thus we obtain a distribution of the  $50 \times 1016 = 50,800$  observations on  $p_{ij}$  and  $p_{ij}^*$  and thus on the efficiency  $\text{eff}_{ij}$ . In this section we analyze these observations in various directions.

### 7.1 Efficiency estimates

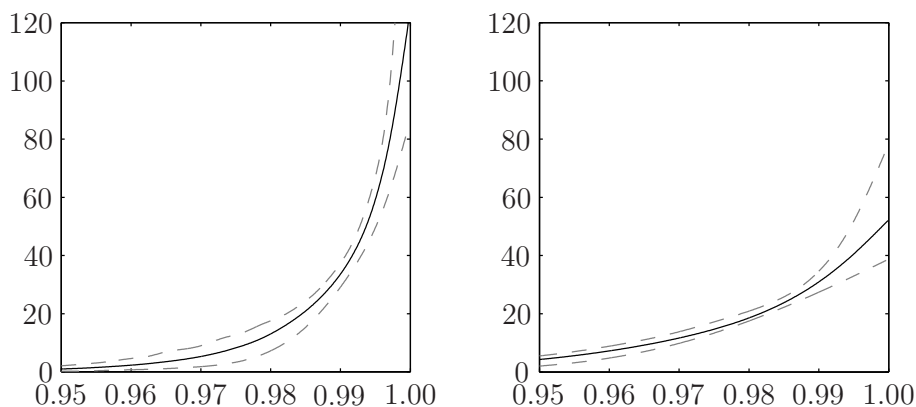


Figure 4: Distribution of the efficiency  $p/p^*$  across players (men left, women right).

The density of  $\text{eff}_{ij}$  is estimated nonparametrically using the quartic kernel. This standard approach, however, leads to a downward bias near the boundary of the support, in our case near one. To avoid the boundary effect we use a local linear fitting method, described in Karunamuni and Alberts (2005). Repeating this for all  $R$  replications we find the median and the 2.5 and 97.5 percentiles in Figure 4. The 95% band reflects uncertainty resulting from the GMM procedure of Section 5.

For the men, the mean of the (median) distribution is 98.9% (with a standard error of 0.2%) and the distribution can be characterized by the 5%, 25%, 50%, 75%, and 95% quantiles given by (96.7, 98.6, 99.3, 99.7, 99.9). Hence, on average the inefficiency is 1.1%, while 25% of the players have an inefficiency of more than 1.4% and 5% of the players an inefficiency of more than 3.3%. At the median, the men can thus increase the probability  $p$  of winning a point on service by 0.7%-points: from 64.8% to 65.5%.

For the women, the mean of the distribution is 98.0% (0.3%) and the 5%, 25%, 50%, 75%, and 95% quantiles are given by (94.2, 97.2, 98.6, 99.4, 99.8). Hence, on average the inefficiency is 2.0%, while 25% of the players have an



inefficiency of more than 2.8% and 5% of the players an inefficiency of more than 5.8%. The women are thus less efficient servers than the men. At the median, the women can increase the probability of winning a point on service by 1.2%-points: from 56.3% to 57.5%.

The inefficiency estimates are lower bounds because of our conservative approach. This conservativeness has two sources. The first source consists of the maximization over  $\lambda$  in Section 6.2. To analyze the sensitivity of the efficiency estimates with respect to  $\lambda$ , we compute the efficiency in the feasible range of  $2 \leq \lambda \leq 5$ . (The reason why this may be considered the feasible range is that  $\lambda < 2$  is unrealistic because it implies  $x_1^* < 57.7\%$  while the observed frequencies for  $x_1$  are about 60%, and  $\lambda > 5$  is unrealistic because it implies  $x_1^*/x_2^* > 0.7$  despite an average  $x_1/x_2$  of about 0.7.) For  $\lambda$  fixed at 2, 3, 4, and 5 we obtain mean efficiency estimates of 98.7, 98.9, 98.8, and 98.5 for the men, and 97.3, 97.8, 98.0, and 98.0 for the women. These are close to the conservative estimates above (98.9 and 98.0), so that maximization over  $\lambda$  does not appear to contribute much to the conservativeness of our efficiency estimates.

If we assume, instead of model  $C_0$ , one of the models  $C_1$  or  $C_2$ , then  $\hat{\lambda}$  changes little (see Table 3), but this does not necessarily imply that the efficiency  $p/p^*$  also changes little. This, however, is the case. For example, under model  $C_2$  the mean efficiency is estimated as 99.0% for men (89.9% under  $C_0$ ) and 98.1% for women (98.0% under  $C_0$ ). The (in)efficiency estimates are therefore not very sensitive to small misspecifications in the  $y$ -curve.

The second source of conservativeness is the assumed optimality of (some of the) service inputs apart from  $x$ , as discussed in Section 2.3. We now argue that this second source is minor. First, the variable  $x$  captures most of the relevant service inputs. Second, the potentially important other aspects (insofar they are not represented through  $x$ ) are speed, direction, and spin. These will vary in particular between first and second services. Model  $C_2$  allows for separate  $y$ -curves for first and second services, and shows that their impact on efficiency is small. Third, Walker and Wooders (2001) show that optimality regarding direction is justified. Fourth, other aspects (emotion, concentration) will have an even smaller effect on efficiency, as their dependence on  $x$  is small. Thus, the potential harm of the conservativeness of our approach is small, so that we may ignore the fact that our estimates are actually upper bounds.

One might ask how much of inefficiency is caused by aggregation. We assume that points are i.i.d. over a match, but suppose they are only i.i.d. over a set or a game? We have two answers to this objection. First, points are almost i.i.d. over a match, as shown in Klaassen and Magnus (2001). Second, disaggregation increases the measurement error and is therefore not feasible in our data. The discussion in Section 4 shows that moving from match to set data produces essentially no new information.

We conclude that the service strategy of top tennis players is not fully efficient. This is perhaps not surprising, because 100% efficiency is unattainable, even for top athletes. We thus differ from Walker and Wooders (2001), who do not reject that the probability of winning a point is the same for a service to the left and to the right. Our main result, however, is not the lack of full efficiency, but rather the fact that we obtain estimates of the (in)efficiency: 1.1% (0.2%) for men and 2.0% (0.3%) for women. These are small compared to the large suboptimalities found in laboratory experiments.

## 7.2 Impact of inefficiency

The fact that the inefficiency is small does not necessarily imply that the *impact* of the inefficiency is small too. We analyze the impact at point, game, set, and tournament level.

Although we *measure* efficiency in relative terms ( $p/p^*$ , see Section 2.4), the *impact*, we believe, is best captured by the absolute difference, that is,  $p^* - p$  at point level. For example, for  $(p, p^*) = (0.9, 1.0)$  and  $(p, p^*) = (0.1, 0.2)$ , respectively, we find  $p^* - p = 10\%$ -points in both cases, while the inefficiency measure  $1 - p/p^*$  is 10% in the first case and 50% in the second case. In the second case, the player can thus increase  $p$  by 50%, indicating a large inefficiency. But the *impact* is small: he/she still has only 20% probability of winning the point. In the first case, the player can increase  $p$  by 10% both absolutely and relatively. Although the inefficiency is lower than in the second case, the impact is equal, and this seems about right.

At point level the impact of inefficiency is that by serving efficiently, men can increase  $p$  by 0.7%-points (from 64.8% to 65.5%) on average, and women by 1.2%-points (from 56.3% to 57.5%). This again seems small. But a player serves many points in a tennis match. Thus we study the impact at higher levels of aggregation, applying the software developed in Klaassen and Magnus (2003) to all  $p$  and  $p^*$ .

If we consider a game, then the impact of inefficiency increases, not because the players perform differently but because of the structure of the tennis scoring system: from 0.7%-points at point level to 1.1%-points at game level for the men, and from 1.2%-points at point level to 2.5%-points at game level for the women.

At match level (arguably the most natural unit) the impact of inefficiency does not only depend on the inefficiency of player  $i$  but also of player  $j$ . What would be the efficiency gain for player  $i$  if he or she switches to serving efficiently while player  $j$  does not? The mean increase in the probability of winning the match is 2.4%-points (0.5%-points) for the men, and 3.2%-points (0.5%-points) for the women. These estimates can be related to Romer (2006), who finds a potential 2.1%-points increase of the probability of winning an (American) football match. Despite the fact that he studies

teams — for which one might expect larger inefficiencies than for individual players — Romer’s efficiency gain is lower than in our study, probably because his analysis considers only a small aspect of a match, whereas the service in tennis is of great importance.

The mean match efficiency gain averages out differences in efficiency gains across matches. In well-balanced matches, for example, serving efficiently is more important than in matches where one player is much stronger than the other. This issue is addressed in Figure 5. For all  $S$  draws of  $(z_{ij}, z_{ji}, \lambda)$

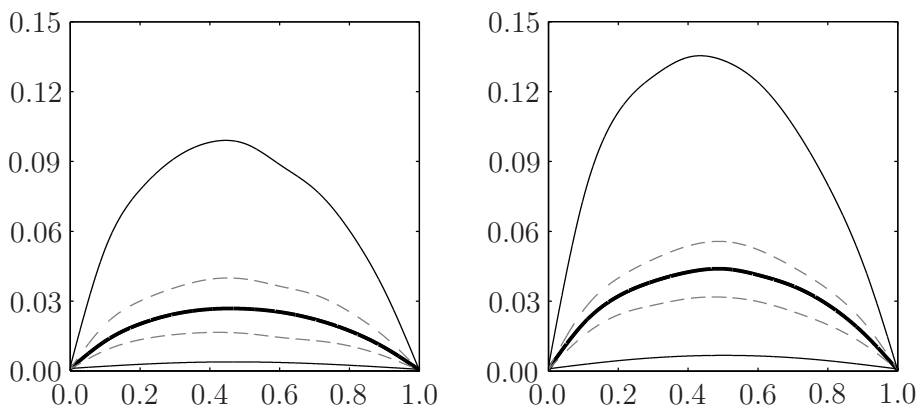


Figure 5: Efficiency gain  $m^* - m$  as a function of  $m$ , match level (men left, women right).

and for each match, we compute the probabilities  $m_{ij}$  ( $i$  wins the match against  $j$ , both serving normally) and  $m_{ij}^*$  ( $i$  wins the match against  $j$ ,  $i$  serves optimally,  $j$  serves normally). Then we regress  $(m_{ij}^* - m_{ij})$  on  $m_{ij}$  (50,800 observations) using nonparametric quantile regressions. We use the same kernel as above, including the boundary correction method, though now not only for  $m_{ij}$  near one but also near zero. The fact that the support is bounded also implies that the fitted regression curve exhibits flattening at the boundaries. We avoid this by using the local linear regression method instead of the more standard locally weighted averaging approach; see Fox (2000). Finally, to account for the skewness of  $m_{ij}^* - m_{ij}$ , we use quantile regression instead of the usual (mean) regression, taking the 10, 50 and 90% quantiles; see Koenker (2005, p. 222) on locally linear quantile regression. This whole procedure gives median and 10 and 90 percentiles regression curves, where the 80% band represents the variation of  $m_{ij}^* - m_{ij}$  across servers. These are the three solid lines in Figure 5. We repeat the procedure for each of the  $R$  replications, so that we can estimate the 95% confidence intervals around the three curves to measure the impact of the GMM estimation uncertainty. For simplicity, Figure 5 only contains the confidence band (dashed) around the median curve. If the two players are approximately of equal strength then

the median efficiency gain at match level for the efficient server is 2.7%-points for the men and 4.4%-points for the women. Moreover, 10% of the players will have an efficiency increase of more than 10–15%-points. In very uneven matches, however, serving efficiently is essentially irrelevant.

We now have estimates of the impact of inefficiency at point, game, and match level, but this does not yet answer the question what the *monetary* impact is. In order to answer this question, be it crudely, we run a hypothetical tournament of 128 players (seven rounds, like Wimbledon), where in each match both players have probability 50% to win the match, except one player who serves efficiently. The only efficient player has 52.7% (that is, an additional 2.7%-points) probability of winning a match in the men's singles (54.4% in the women's singles). What is the expected monetary gain for the efficient player? In grand slam tournaments the paycheck approximately doubles in each round. If we assume that this is exactly true, then the expected paycheck for the efficient player will rise by 18.7% for men and 32.8% for women. At Wimbledon this would mean an expected additional income of approximately \$10,000 for the efficient man and \$15,000 for the efficient woman. Hence, even though the inefficiency at point level may seem small, the monetary impact of inefficiency can be substantial.

Since a player has two services, we may ask whether an optimal first service is more or less important than an optimal second service. This question is answered in Figure 6, where we only consider the median. The graph

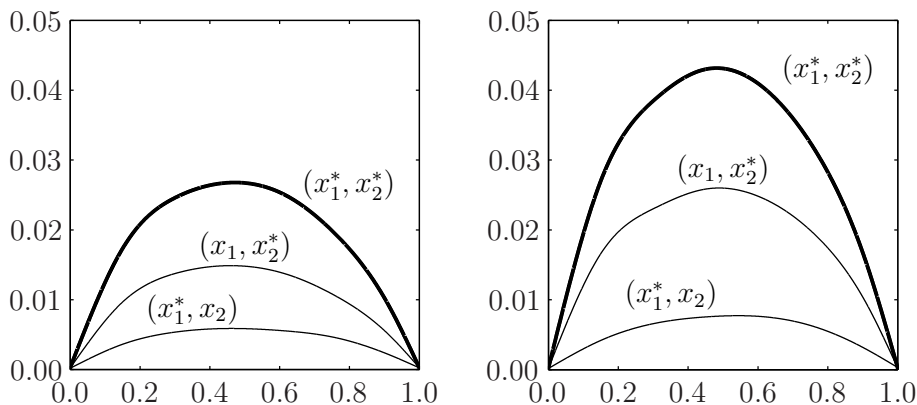


Figure 6: Decomposition of efficiency gain into first and second service, match level (men left, women right).

$(x_1^*, x_2^*)$  is the same as the median in Figure 5, but now decomposed into  $(x_1, x_2^*)$  where only the second service is optimal and  $(x_1^*, x_2)$  where only the first service is optimal. The figure shows that players can achieve a larger efficiency gain on their second service than on their first service, possibly because of a misguided (inefficient) fear for a double fault.

### 7.3 Efficiency and player's quality

Finally we ask whether better players are more efficient than weaker players, as we would perhaps expect. This is no tautology because better players could also be more talented and equally efficient as weaker players. Our basic equality is

$$\text{success} = \text{talent} + \text{efficiency}.$$

A talented but not so efficient player may be as successful as a less talented player who is more efficient. We observe success by the ranking of a player, and we can measure efficiency; talent is unobserved, but can be deduced.

We run the following simple linear regression:

$$\text{eff}_{ij} = \beta_1 + \beta_2 r_i + \beta_3 |r_i - r_j| + \beta_4 r_i r_j + \xi_{ij}.$$

To account for the random effects in  $\text{eff}_{ij}$ , we use for each server the draws  $(z_{ij}, \lambda)$  from Section 6.2, so that we have 50,800 observations. The regression is performed for each of the  $R$  replications of  $\theta$ . To combine the results across  $\theta$ , we draw 100 times from each estimated distribution of  $\hat{\beta}$  and, to account for skewness, the resulting 5000 draws are summarized in a 95% confidence interval around the median.

Two conclusions emerge. First, the confidence interval for  $\beta_2$  is (0.0004, 0.0015) for the men and (0.0004, 0.0023) for the women. Hence, higher-ranked players are indeed more efficient. This is consistent with Walker and Wooders (2001) who find strong rejections of mixed-strategy play in a data set involving inexperienced card players, in contrast to their results for experienced tennis players. Our evidence is even stronger since it comes from a single data set.

Second, the confidence interval for  $\beta_3$  is (−0.0010, −0.0001) for the men and (−0.0011, 0.0000) for the women. Hence, the closer the contest, the more efficient players are forced to be.

## 8 Conclusions and implications

In a tennis match between two players, the objective of each player is to win the match. If it is a match between two amateurs (possibly business partners), one player may be hesitant to beat his or her boss or to win too decisively. But top professionals do not act like this. If we were studying a minor tournament, even a top player might not fully commit because he or she wants to be fit for next week's major tournament. That is why we study only a major tournament, namely Wimbledon. Our 'utility' function to be maximized is the probability that a player (given his or her opponent and given the strengths of both players) wins a point while serving. It seems likely that this is indeed the function which players wish to maximize.

We asked the question whether the service strategy for top tennis players, playing in a top tournament, is efficient. The answer is that it is not. The rejection of perfect efficiency is not surprising. More important are our estimates of the magnitude of the inefficiency, and our conclusion that the inefficiencies are small. Our model and set-up are very general. Most assumptions have been subjected to extensive sensitivity analyses in order to check whether (small) deviations from the assumptions have a significant impact on the results. The results are robust, and even if the remaining restrictions were incorrect and thereby would inflate the estimated inefficiencies, this would only strengthen our conclusion that inefficiencies are small.

The inefficiency is measured by the potential relative increase in the probability of winning a point on service, and is on average 1.1% for men (2.0% for women). The impact of serving efficiently can be quantified at various levels of aggregation. At point level the impact is 0.7%-points (1.2%-points), at game level 1.4%-points (4.0%-points), and at match level 2.4%-points (3.2%-points). These differences do not reflect the players but the scoring system and the fact that at match level the impact of service efficiency depends on the quality difference between the players. In terms of expected monetary gains we have calculated that the expected paycheck for the efficient player could rise by 18.7% for men and 32.8% for women. So even small inefficiencies can have substantial financial consequences.

What is the reason for this inefficiency? Perhaps top tennis players know their  $y$ -curve, but are not able to solve the optimization problem. Or do they correctly solve the optimization, but on the wrong  $y$ -curve? The decomposition into first and second service (Figure 6) may help us here. From the point of view of achieving optimality it is much easier for a server to work out the optimal second service (maximize  $w(x)$ ) than to work out the optimal first service (maximize  $w(x) + (1 - x)w(x_2^*)$ ). Nevertheless, the second service appears to be less efficient than the first service. This provides evidence that, although players may be maximizers, *they do not maximize the correct function*.

We also find that higher-ranked players are more efficient than lower-ranked players, and that the closer the match, the more efficient a player serves.

Let us now try and relate these results to economics, and in particular to our original question whether economic agents are successful maximizers. First, our results show that economic agents and even top agents are not fully efficient, and that this inefficiency can be financially nontrivial. Second, the different inefficiency results at point-, game-, and match level show that the market form has an impact on the measurement of inefficiency. For example, in the case of Bertrand competition for homogeneous goods, a ‘winner-takes-all’ effect is induced, and this nonlinearity is not related to productivity



but to the nature of the competition. Third, the fact that tighter matches cause players to be more efficient suggests that in a more competitive market firms are forced to be more efficient; otherwise they will be driven out of the market. This supports the view of many policy makers that measures aimed at strengthening market mechanisms lead to a more efficient economy. Moreover, it corroborates Friedman's (1953, p. 22) point that only return-maximizing businesses survive in a competitive market.

What contribution, if any, does this paper make to the closely related subject of rational behavior? In rational choice theory the 'rational man' is assumed to (a) know his or her preferences over all relevant alternatives, and (b) choose the best alternative. In our tennis framework this corresponds to (a) a player who knows the pay-off structure  $p(x_1, x_2)$ , and (b) is able to maximize that pay-off.

We find inefficiency and hence the agent does not know the pay-off structure and/or is not able to solve the maximization problem. This would reject rationality in favor of bounded rationality, and therefore corroborates findings from the experimental economics literature.

The inefficiency is larger for the second than for the first service, even though the maximization problem for the second service is easier. This suggests that at least (a) is rejected, that is, the agent does not fully use his or her true preference relations.

The inefficiency is much smaller than typically found in studies using laboratory experiments. This can be explained by the higher motivation, larger experience, and stronger competition among top tennis players. The impact of competition is supported by our data. The impact of experience is confirmed by studies in experimental economics, which find that learning reduces violations of rationality; see van de Kuilen and Wakker (2006). An open question in that literature is whether in the limit (infinite learning) these violations of rationality disappear. Our results suggest that extensive learning leads to quite rational decisions. This comes close to an affirmative answer to Tversky and Kahneman's (1986) question whether incentives, experience, and competition ensure fully rational choices.

Another implication is that rationality might be an acceptable approximation in many applications, *at least for the top agents*. Friedman's (1953, p. 21) expert billiard players are like our top tennis players: they make their shots as if they know and can implement the complicated mathematical formulas underlying the optimal paths of the balls. This will be approximately true for expert billiard players, but not necessarily for the average billiard player. Since we find that the efficiency is smaller for weak players than for strong players, rationality may not be a good approximation for the typical (amateur) tennis player, billiard player, or economic agent. Hence, in spite of Friedman's (p. 21) assertion that

*It is only a short step from these examples to the economic hypoth-*



*esis that under a wide range of circumstances individual firms behave as if they were seeking rationally to maximize their expected returns,*

it is not justified to think of ‘individual firms’ (that is, not the top firms) as rational.

## Acknowledgements

We thank IBM UK and The All England Club at Wimbledon for their kindness in providing the data. Preliminary versions of this paper were presented at seminars at the University of Amsterdam, Erasmus University Rotterdam, the First International Conference on Mathematical Modelling in Sport (Manchester, UK, June 2007), the New Economic School of Moscow, and at ASSA 2008 in New Orleans. We are grateful to all participants for their many comments. In particular we thank Jan Boone, Tijmen Daniëls, George Deltas, Bas Donkers, Kees Jan van Garderen, Noud van Giersbergen, Wouter den Haan, Philip Jung, Siem Jan Koopman, Knox Lovell, Theo Offerman, Thijs ten Raa, Joep Sonnemans, Mark Steel, and Koen Vermeylen for help and constructive discussions, and two referees and Arnold Zellner for their encouraging and useful comments.

## References

- Adams, C.P., 2006. Estimating the value of ‘going for it’ (when no one does). Mimeo, Federal Trade Commission.
- Carlton, D.W., Frankel, A.S., Landes, E.M., 2004. The control of externalities in sports leagues: an analysis of restrictions in the National Hockey League. *Journal of Political Economy* 112(1), S268–S288.
- Chiappori, P.-A., Levitt, S., Groseclose, T., 2002. Testing mixed-strategy equilibria when players are heterogeneous: the case of penalty kicks in soccer. *American Economic Review* 92(4), 1138–1151.
- Dixon, M.J., Coles, S.G., 1997. Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics* 46(2), 265–280.
- Ehrenberg, R.G., Bognanno, M.L., 1990. Do tournaments have incentive effects? *Journal of Political Economy* 98(6), 1307–1324.

- Førsund, F.R., Lovell, C.A.K., Schmidt, P., 1980. A survey of frontier production functions and of their relationship to efficiency measurement. *Journal of Econometrics* 13(1), 5–25.
- Fox, J., 2000. *Nonparametric Simple Regression: Smoothing Scatterplots*. Sage Publications, Thousand Oaks, CA.
- Friedman, M., 1953. *Essays in Positive Economics*. Chicago University Press, Chicago.
- Gale, D., 1971. Optimal strategy for serving in tennis. *Mathematics Magazine* 44, 197–199.
- Gandar, J., Zuber, R., O’Brian, T., Russo, B., 1988. Testing rationality in the point spread betting market. *The Journal of Finance* 43(4), 995–1008.
- Garicano, L., Palacios-Huerta, I., Prendergast, C., 2005. Favoritism under social pressure. *The Review of Economics and Statistics* 87(2), 208–216.
- Goff, B.L., McCormick, R.E., Tollison, R.D., 2002. Racial integration as an innovation; Empirical evidence from sports leagues. *American Economic Review* 92(1), 16–26.
- Golec, J., Tamarkin, M., 1991. The degree of inefficiency in the football betting market. *Journal of Financial Economics* 30, 311–323.
- Gray, P.K., Gray, S.F., 1997. Testing market efficiency: evidence from the NFL sports betting market. *The Journal of Finance* 52(4), 1725–1737.
- Gwartney, J., Haworth, C., 1974. Employer costs and discrimination: the case of baseball. *Journal of Political Economy* 82(4), 873–881.
- Hajivassiliou, V., McFadden, D., Ruud, P., 1996. Simulation of multivariate normal rectangle probabilities and their derivatives — theoretical and computational results. *Journal of Econometrics* 72, 85–134.
- Julien, B., Salanié, B., 2000. Estimating preferences under risk: the case of racetrack bettors. *Journal of Political Economy* 108(3), 503–530.
- Kahn, L.M., Sherer, P.D., 1988. Racial differences in professional basketball players’ compensation. *Journal of Labor Economics* 6(1), 40–61.
- Kanazawa, M.T., Funk, J.P., 2001. Racial discrimination in professional basketball: evidence from Nielsen ratings. *Economic Inquiry* 39(4), 599–608.

- Karunamuni, R.J., Alberts, T., 2005. A generalized reflection method of boundary correction in kernel density estimation. *The Canadian Journal of Statistics* 33(4), 497-509.
- Klaassen, F.J.G.M., Magnus, J.R., 2001. Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association* 96, 500–509.
- Klaassen, F.J.G.M., Magnus, J.R., 2003. Forecasting the winner of a tennis match. *European Journal of Operational Research* 148, 257–267.
- Koenker, R., 2005. *Quantile Regression*, Econometric Society Monographs No. 38. Cambridge University Press, New York.
- Koop, G., Osiewalsky, J., Steel, M.F.J., 1997. Bayesian efficiency analysis through individual effects: hospital cost frontiers. *Journal of Econometrics* 76, 77–105.
- van de Kuilen, G., Wakker, P.P., 2006. Learning in the Allais paradox. *Journal of Risk and Uncertainty* 33, 155–164.
- Kumbhakar, S.C., Lovell, C.A.K., 2000. *Stochastic Frontier Analysis*. Cambridge University Press, New York.
- Lazear, E.P., Rosen, S. 1981. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy* 89(5), 841–864.
- Leibenstein, H., 1966. Allocative efficiency vs. ‘X-efficiency’. *American Economic Review* 56(2), 392–415.
- Levitt, S.D., 2004. Why are gambling markets organised so differently from financial markets? *The Economic Journal* 114, 223–246.
- Magnus, J.R., Klaassen, F.J.G.M., 1999a. On the advantage of serving first in a tennis set: four years at Wimbledon. *The Statistician (Journal of the Royal Statistical Society, Series D)* 48, 247–256.
- Magnus, J.R., Klaassen, F.J.G.M., 1999b. The effect of new balls in tennis: four years at Wimbledon. *The Statistician (Journal of the Royal Statistical Society, Series D)* 48, 239–246.
- Nardinelli, C., Simon, C., 1990. Customer racial discrimination in the market for memorabilia: the case of baseball. *The Quarterly Journal of Economics* 105(3), 575–595.

- Palacios-Huerta, I., Volij, O., 2008. *Experientia docet: professionals play minimax in laboratory experiments.* *Econometrica* 76(1), 71–115.
- Romer, D., 2006. *Do firms maximize? Evidence from professional football.* *Journal of Political Economy* 114(2), 340–365.
- Sauer, R.D., Brajer, V., Ferris, S.P., Marr, M.W., 1988. *Hold your bets: another look at the efficiency of the gambling market for National Football League games.* *Journal of Political Economy* 96(1), 206–231.
- Schmidt, M.B., Berri, D.J., 2004. *The impact of labor strikes on consumer demand: an application to professional sports.* *American Economic Review* 94(1), 344–357.
- Schmidt, P., 1985/86. *Frontier production functions.* *Econometric Reviews* 4(2), 289–328.
- Stigler, G.J., 1976. *The Xistence of X-efficiency.* *American Economic Review* 66(1), 213–216.
- Stone, E.W., Warren Jr, R.S., 1999. *Customer discrimination in professional basketball: evidence from the trading-card market.* *Applied Economics* 31, 679–685.
- Szymanski, S., 2000. *A market test for discrimination in the English professional soccer leagues.* *Journal of Political Economy* 108(3), 590–603.
- Tversky, A., Kahneman, D. 1986. *Rational choice and the framing of decisions.* *Journal of Business* 59, 251–278.
- Varian, H.R., 1982. *The nonparametric approach to demand analysis.* *Econometrica* 50, 945–973.
- Varian, H.R., 1985. *Non-parametric analysis of optimizing behavior with measurement error.* *Journal of Econometrics* 30, 445–458.
- Walker, M., Wooders, J., 2000. *Equilibrium play in matches: binary Markov games.* Mimeo, Department of Economics, University of Arizona.
- Walker, M., Wooders, J. 2001. *Minimax play at Wimbledon.* *American Economic Review* 91(5), 1521–1538.
- Zuber, R.A., Gandar, J.M., Bowers, B.D., 1985. *Beating the spread: testing the efficiency of the gambling market for National Football League games.* *Journal of Political Economy* 93(4), 800–806.