

PERFORMING BEST WHEN IT MATTERS MOST: EVIDENCE FROM PROFESSIONAL TENNIS

JULIO GONZÁLEZ-DÍAZ*, OLIVIER GOSSNER, AND BRIAN W. ROGERS

ABSTRACT. Stakes affect aggregate performance in a wide variety of settings. At the individual level, we define the critical ability as an agent's ability to adapt performance to the importance of the situation. We identify individual critical abilities of professional tennis players, relying on point-level data from twelve years of the US Open tournament. We establish persistent heterogeneity in critical abilities. We find a significant statistical relationship between identified critical abilities and overall career success, which validates the identification procedure and suggests that response to pressure is a significant factor for success.

1. INTRODUCTION

In many contexts, the choices faced by a decision maker vary widely in their importance. Most of a company's commercial deals may be part of the day-to-day routine, whereas a few other deals may have a huge impact on the company's success. Traders experience successive periods of low and high market volatility, and take positions of varying magnitude. Political speeches address larger or smaller audiences, and some are given at crucial times. In most professionals' lives, performance at just a few exams and job interviews has a dramatic impact on career development. From the point of view of a decision maker, performance in critical moments matters much more than in other moments, and it is beneficial to adjust performance accordingly.

This paper establishes persistent heterogeneity in individuals' responses to the importance of situations, and finds a significant relationship between individual adjustment capacities and career success. Although the question of performance

Date: August 24, 2012.

*Corresponding author. Phone: (+34) 881813207, fax: (+34) 881813197, e-mail: julio.gonzalez@usc.es.

under pressure has received considerable attention both in economics and psychology, heterogeneity in agents' reactions to pressure has, to the best of our knowledge, not yet been convincingly demonstrated.¹ When matching individuals according to their abilities, it is precisely this heterogeneity, and not the overall effect of pressure on aggregate performance, that is relevant.

Key to our approach is distinguishing two different notions of skill. The first is the general ability to perform well, resulting in relatively good outcomes *on average*. The second is the ability to adjust one's performance to the importance of the situation, without necessarily improving one's average performance. The latter ability is desirable when it is not possible to always maintain peak performance, since it trades better performance in high stakes situations for worse performance in unimportant ones. We focus primarily on this second skill, which we call *critical ability*.

There are at least three reasons why, in many situations, it is not possible to perform consistently at one's best, and thus why heterogeneity of critical ability may be a significant determinant of overall success. The first is that an agent may have limited resources to allocate toward a sequence of decisions. In this case, performing better in one situation comes at the opportunity cost of performing worse in other situations. Critical ability then manifests as skill in allocating resources optimally across a set of decisions. The second reason is that important situations may involve feelings induced by high pressure which can affect performance, in which case high critical ability corresponds to an agent's psychological ability to respond well to such pressure. Third, it may be difficult to identify which situations are in fact the most important. Part of high critical ability may be the ability to accurately assess the importance of various situations.

¹For example, the "clutch hitting" literature (see, e.g. Albert and Bennet, 2003; Albert, 2007, and references therein) studies whether professional baseball players over-perform in important situations. Despite many studies involving considerable amounts of data, no conclusive evidence of heterogeneity has been obtained.

We explore these questions of general interest in the context of men's professional tennis. This setting offers a number of special empirical advantages, commonly exploited in the economics literature since the pioneering work of Walker and Wooders (2001). Paserman (2010) shows how these advantages can be used to study psychological reactions to pressure.

First, the agents are elite, well-trained, and highly motivated. For relatively low-ranked players, a good showing in a tournament can revolutionize their career. For top players, the amount of purse money and prestige on the line are substantial. Perhaps even more significantly, the outcomes affect their professional rankings and their ability to secure lucrative sponsorships.

Second, the tennis scoring system allows for an unambiguous definition of the importance of each point, based on an estimate of the impact of the point on the probability of winning the match. The point-level data we use allows for a particularly precise paired measurement of outcomes and importance. There is a clear winner of each and every point in the match. We thus have high quality information about the importance of various situations and an unambiguous means of measuring players' relative performance. Relating this importance to the outcomes of successive points is precisely what allows us to test for the heterogeneity in the players' critical ability and to relate these abilities to career success.

Finally, in a single tournament, each participant plays many points, and the points typically vary rather dramatically in their importance for determining the outcome of the match. Given the size of our data, it becomes possible to identify the relative critical abilities of players in a precise way.

Professional tennis thus provides an ideal context in which to assess how individuals adjust their performance to the importance of a specific situation. If these highly experienced and motivated professionals differ in this adjustment, then there is reason to believe that the phenomenon occurs in other professional settings as well.

The first result of our econometric analysis is to establish that critical abilities do indeed differ significantly across players. That is, relative to their opponents,

some players exhibit the capacity to play their best game in the most important situations, while others exhibit a decrease in performance in these situations. Since our analysis covers 12 years of US Open tournaments, this result is a first indication that some players are consistently better than others at winning important points over time.²

Given individual-level estimates of players' critical abilities, we ask how relevant these abilities are, in the sense of how strongly related they are to the career success of the players. Specifically, we are interested in the relationships between a player's critical ability and his professional tennis rating. To interpret these results, it is important to recall that a higher critical ability does not capture an effect of winning more points. Instead, it corresponds to a shift in the distribution of points a player is likely to win. In particular, a player with high critical ability is more likely to win the more important points in a match, and less likely to win the less important points; however, he wins the same number of points on average as a player with lower critical ability, all else being equal. That is, this relationship is separate from the conventional notion of being a stronger tennis player, in terms of being more likely to win any given point.

An analysis of players' performance in the tournaments separate from those used for the measurement of critical abilities shows a significant relationship between these abilities and individual performance. This result validates our identification of critical abilities. The fact that the critical abilities obtained from US Open tournaments have some explanatory power to identify the better players in the remaining tournaments demonstrates persistence in the skill we call critical ability across tournaments. We then conclude that the capacity to win important points is an individual characteristic that is persistent over time.

²Following the influential work of Walker and Wooders (2001), who study mixed strategy equilibrium in sports, Chiappori et al. (2002) study mixed strategy equilibrium using soccer penalty kicks, paying particular attention to aggregation problems arising from heterogeneity in skills of players. While their starting point is an observed physical heterogeneity, our analysis differs in that it demonstrates that tennis players do indeed have heterogeneous skills relating to how they respond to important situations.

We estimate that a one standard deviation increase in critical ability leads to the same impact on career success as 55 percent of a standard deviation increase in serving ability, and the same as 86 percent of a standard deviation increase in returning ability. This illustrates that the effect of critical ability is qualitatively significant.

The idea that psychological factors are relevant for understanding human decision making goes back to Hume (1739), and is central to the influential work of Tversky (see, for instance, Kahneman and Tversky, 1979). Empirical studies of the effect of pressure on performance come both from economists and psychologists. Baumeister and Steinhilber (1984) show that home teams in the baseball World Series tend to win games early in the series but to lose decisive seventh games. They find a similar effect in the National Basketball Association (NBA) when comparing semi-final to final series. Experimental evidence by Ariely et al. (2009) shows that subjects exhibit decreased performance when stakes are extremely high, compared to their usual range of income. Dohmen (2008) shows that social environment affects the performance of soccer players in penalty kicks. Apesteguia and Palacios-Huerta (2010) find a significant advantage for the team shooting first in soccer penalty shootouts, which they interpret as the effect of increased psychological pressure on the team shooting second. Paserman (2010) finds that players' performance deteriorates on important points and estimates that, when facing an opponent of equal quality, a player able to avoid this detrimental performance would increase his probability of winning from 0.5 to 0.75-0.8. Otten (2009) shows that videotaping free-throw attempts decreases performance. While these studies establish that pressure is generally detrimental to performance, our paper is the first, to the best of our knowledge, to demonstrate heterogeneity in individual responses.

The remainder of the paper proceeds as follows. The next section formally defines the notion of the importance of points. Section 3 explains our econometric approach, Section 4 presents the main results, and Section 5 concludes.

2. IMPORTANCE OF POINTS

Variation in the importance of points within a tennis match is the vehicle we use to identify the critical abilities of the players. This variation arises under a very natural definition first proposed by Morris (1977), and employed in Klaasen and Magnus (2001), Paserman (2010), and Abramitzky et al. (2011). The fundamental idea behind this definition is that a point's importance represents the extent to which the (stochastic) outcome of the match hinges on the outcome of the given point.

To understand the notion of importance, it is necessary to first describe the rules by which tennis is scored. The US Open is a single-elimination tournament consisting of matches between pairs of players. Each player's objective is to win the match. In a match, the two players alternate service games, with a randomly chosen player electing who serves first. The winner of a game is the first player to reach four points, with a margin of victory of at least two points.³ The first player to win six games in this fashion is awarded a set. Again, a margin of two games is necessary to win a set, but there is a tie-break played if the game score reaches 6-6 within a set. The winner of the match is the player who first wins three sets.

The importance of a point is defined as follows. Assuming that, conditional on which player is serving, the outcomes of points are i.i.d. random variables with known Bernoulli probabilities, one can compute, at any score, the probability that a player will go on to win the match. The (absolute) difference in this probability, as a function of the outcome of the current point, is the point's importance.

As shown by Walker et al. (2010), absent psychological or effort-allocation effects, a game-theoretical analysis predicts that outcomes of points in tennis are i.i.d. conditional on the serving side and on the server. The i.i.d. case corresponds to our null hypothesis, and should be rejected in the presence of heterogeneity in

³The game continues for as many points as necessary to declare a winner. Throughout, we denote the point score within a game by 0, 1, 2, 3, instead of the equivalent conventional language of tennis, which uses love, 15, 30, 40.

critical abilities. We still rely on the i.i.d. assumption for computing point importance values when estimating the magnitude of the effect of critical abilities. This approach has the advantage of being standard, easily computable, and yielding a good approximation of what one could obtain under a more elaborate approach to point importance.

Arbitrarily name the players of a given match 1 and 2. Denote the score by $\{(p_i, g_i, s_i)\}_{i=1,2}$, with the interpretation that (p_i, g_i, s_i) denotes the number of points, games, and sets, respectively, that player i currently holds. Denote by θ the state of a match, which specifies the current score and the identity of the current server. Let $W(\theta)$ denote the event in which 1 wins the point at θ , and let $L(\theta)$ denote the the complementary event in which 1 loses the point at θ . Finally, denote by W^* the event in which 1 wins the match. The importance of the point at θ is given by

$$I(\theta) = \Pr(W^* | W(\theta)) - \Pr(W^* | L(\theta)).$$

The importance of a point can be decomposed into the constituent probabilities of winning at the various hierarchical levels of the match. In particular, $I(\theta)$ is the product of (i) the importance of the point for determining the outcome of the current game, (ii) the importance of the game in determining the outcome of the current set, and (iii) the importance of the set in determining the outcome of the match.

We introduce notation to clarify this idea. First, let $I(\theta)$ be alternatively represented as $PiM(\theta)$, short for *Point in Match*. Denote the importance of the point for determining the outcome of the current game by $PiG(\theta)$, for *Point in Game*. That is, $PiG(\theta)$ is the difference in the probability that player 1 wins the current game as a function of whether he wins or loses the current point. Analogously, define the importance of a game for determining the outcome of the current set by $GiS(\theta)$ for *Game in Set*; and finally define the importance of the set by $SiM(\theta)$, for *Set in Match*. We then have the following:

PROPOSITION 1. *For any point θ ,*

$$PiM(\theta) = PiG(\theta) \cdot GiS(\theta) \cdot SiM(\theta).$$

Proof. See Appendix A. □

This result expresses the independence of outcomes across levels of the scoring hierarchy. That is, conditional on the outcome of any particular game (set), the marginal importance of a point (game) beyond that game (set) is zero. While the idea is intuitive, it allows us to compute point importance more efficiently than the recursive approach used in the previous literature.

A crucial advantage of adopting this definition is that importance varies in substantial and subtle ways across the points in a match. To give a rough idea of the distribution of importance in our data, the median importance of a point is 1.7 percent, but the distribution is skewed, and the mean is 2.9 percent, with a standard deviation of 3.9 percent.

To begin with, independent of any effects beyond the current game, a point's importance in the game crucially depends on the player's probability of winning a serve, which we denote by q . For each of four scores, Figure 1 (left panel) illustrates the importance of the point for determining the outcome of the current game (PiG) as a function of q . Notice that the ranking of importance values for the scores in a game varies with this parameter, even ordinally. Figure 1 (right panel) reports the PiG for several scores of a game at the global average value of $\bar{q} = 0.63$ in our data.

Given that our main objective is to measure players' ability to respond to changes in importance, much of the power of our approach relies on accurately representing the variation in points' importance during a match. For instance, as we demonstrate below, the obvious binary indicators for importance, such as "break points",⁴ do not adequately capture variation in point importance.

⁴A break point is a point that, if won by the returner, results in the returner winning the game, or "breaking" the serve of his opponent. Given that typically $q > \frac{1}{2}$, so that the server typically wins the game, break points tend to be important.

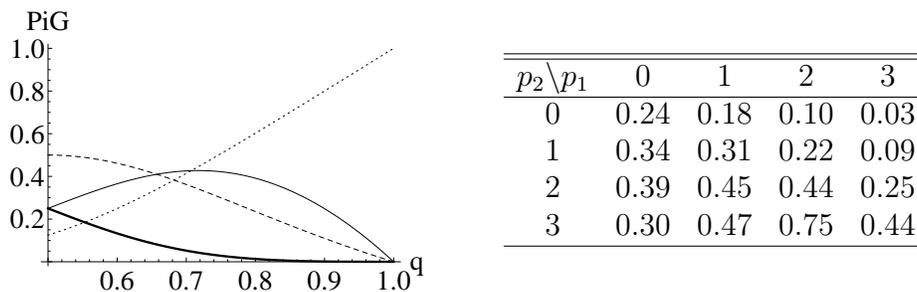


FIGURE 1. Left panel: heavy 2-0, dashed 3-3 (deuce), solid 0-2, dotted 0-3. Right panel: Game-importance of a point at $\bar{q} = 0.63$.

Before proceeding to the empirical sections of the paper, we make a few simple observations about point importance. First, closer matches have more points with high importance.⁵ This is because in unequal matches, the eventual outcome of the match is known with high probability early on, rendering the outcomes of individual points relatively unimportant. Thus, as players' abilities diverge, most points converge to zero importance. Second, we can summarize which points tend to be more important. Break points tend to be important whenever $q > 1/2$. When players are equally matched, the most important games are those for which the current set is close, and the most important sets are those in which the match is close. Finally, when one player is significantly stronger than the other, the most important points are those at which the weaker player has the (surprising) opportunity to go ahead. For a more detailed discussion of the properties of the importance variable, see Appendix B.

It is important to recognize that psychological or effort-allocation effects can cause points to be won in a non-i.i.d. way. In fact, we will demonstrate that players' heterogeneous responses to the importance of a point indeed have such an effect, and this could in turn influence our quantitative estimates of players' abilities. One observation to mitigate this concern is that Klaasen and Magnus (2001) specifically investigate the i.i.d. assumption in professional tennis empirically and conclude that it is a good approximation. More directly, we can verify that the discrepancy

⁵By "closer", we mean matches that have more points played at nearly-equal scores, so that the outcome of the match cannot be predicted with high probability until near the end, if then.

from an i.i.d. point distribution implied by the estimates from our model are indeed very small.⁶ Thus, we are confident that our results are not unduly influenced by the i.i.d. assumption that we use to compute point importance.

We emphasize that the assessment of importance at various stages in the match crucially depends on the relative abilities of the players, as measured by the values of q for each server. Any method that ignored the context of the match and defined importance as a function only of the score, would necessarily leave out important information, and would not yield reliable results. It is important, therefore, that we are able to get good approximations of the values of q for all the matches in our data set.

3. EMPIRICAL STRATEGY

Our data set, provided by the United States Tennis Association (USTA), consists of detailed point-by-point data of twelve annual US Open tournaments from 1994-2006.⁷ The US Open is one of the four annual professional Grand Slam tournaments. We focus on the men’s singles competition which includes, collectively, 1009 matches. Table 1 provides descriptive statistics on the data under the column “Principal data.” In order to identify the relative abilities of players with sufficient accuracy, we use a subset of this data to perform our main econometric analysis, on which we elaborate below. Additionally, our approach requires an auxiliary data set consisting of matches played in other tournaments in order

⁶Accounting for the effects of point importance in an entire match is computationally impracticable, which is why we use the i.i.d. framework. But we illustrate the effect using PiG in the context of a single game. We first compute the importance at every score in the game assuming an i.i.d. point distribution. We then recompute the importance values taking account of critical abilities. The resulting differences depend on (i) the importances computed under the iid assumption, (ii) the difference in critical ability between the two players, and (iii) the value of GiM – the importance of the game in the match. For representative values, we take $q = 0.65$, a difference in critical abilities of 5, and $GiM = 0.1$. The correlation between the two lists of importance across all scores in the game is 0.98. Of course, as the difference in critical abilities and the value of GiM increase, the vectors become less similar. But even so, the correlations remain high, and the relative orderings of point importance in the game are very similar.

⁷This data represents all matches for which information was made available to us by the USTA. We were unable to obtain point-level data from any other professional tennis matches.

to independently measure players' relative strengths. We obtained the auxiliary data set, which is publicly available, directly from the website of the Association of Tennis Professionals (ATP). This data includes aggregate information at the match level but does not contain point-by-point data.

TABLE 1. Descriptive statistics of the data

	Principal data	Reduced data	Auxiliary data
Tournaments	12	12	117
Grand Slams	12	12	36
Players	355	94	766
Matches	1,009	494	9,013
Points	223,139	110,033	2,137,238

Notice that we can hope to measure only relative, but not absolute, abilities of the players. To illustrate, if there is a match where player 1 wins most of his service points, one cannot identify whether this is because he is a strong server, or his opponent is a weak returner. If we had access to additional information, such as service velocity or frequency of aces, one could hope to identify an absolute level of performance. Notice, though, that identifying absolute abilities is not relevant to our approach, as it is only relative abilities that matter.⁸ Similarly, even though it is true that players differ in their critical abilities, there would still be no difference in the outcomes of points even if all players were to simultaneously improve their critical abilities in a uniform manner. In any case, the aim is to express abilities *relative* to the average ability in the population. In the process, we will also be able to determine cardinal rankings across players' abilities.

Since we are mainly interested in demonstrating that players differ meaningfully in their ability to perform well in critical situations, rather than in capturing the subtle changes in abilities over time that a player may undergo, we make the assumption that each player can be described by a set of abilities that are constant

⁸Using additional information to estimate absolute performance would also require making additional assumptions about the production function of winning a point. One advantage of our approach is that such assumptions are not needed.

across all the matches we observe. We thus pool all tournaments into a single set of matches instead of analyzing them separately.⁹

In regard to identification, each tournament of the US Open allows for direct comparison of pairs of players, and each has a tree structure of matches. This makes the comparison of players who play against each other an easy task, but noise terms accumulate when comparing players who are far apart in the tree. Pooling tournaments has a double advantage in this respect: not only does it add data points for each player, but it also enhances the connectivity of the graph that allows us to compare all pairs of players, thereby permitting us to compare each player’s characteristics against the pool of others.

When one tests for heterogeneity among players’ characteristics, pooling players across tournaments is more demanding, since heterogeneity arising from time-dependent characteristics of players is now ignored. In order to conclude that heterogeneity exists, not only do we need to observe asymmetric behavior in matches, but we also need these asymmetries to arise from characteristics that are consistent over time. In particular, this precludes finding heterogeneity if it exists only in a time-inconsistent manner, as shown to be an issue in the “clutch hitting” literature (Albert and Bennet, 2003; Albert, 2007).

Even after the data are pooled, there are still players for whom we have too few observations to estimate their characteristics well. We proceed by eliminating those players (and their matches) for whom we observe fewer than five matches.¹⁰ After this is done, some players remain who have fewer than five matches remaining in the sample, even though we originally observed at least five matches for them. The pruning is therefore iterated until all players left in the sample have at least five matches in the data. In the end, this leaves us with a reduced data set of ninety-four players and roughly 110,000 points. This is the sample that we use to

⁹As we argue below, while this pooling approach rules out potentially interesting patterns of career development, it presents important methodological advantages and enhances the robustness of our conclusions.

¹⁰While somewhat arbitrary, using data for five matches turns out to be sufficient to obtain estimates with tight enough standard errors for our purposes.

report our results in the next section, and which is summarized under “Reduced Data” in Table 1.¹¹

Pooling and pruning allow us to identify players’ abilities (Section 4.2) with enough accuracy to relate these abilities to players’ success (Section 4.3). These procedures are, however, not necessary when testing for heterogeneity in these abilities. Our heterogeneity tests in Section 4.1 are confirmed by a nonpooling analysis.

3.1. Service probabilities. Within a match, the importance of each point depends not just on the current score but, crucially, on the estimated probabilities with which each player wins a given service point. Denote by $q = (q_1, q_2)$ the probabilities with which each player wins a given service point. We begin the empirical analysis by constructing, for each match, a measure of q .

To avoid technical problems in the regression analysis, we want to have the importance of a point depend only on events that happened before the point. So in the beginning of a match, we require estimates of the q_i that reflect the relative abilities of the two players, but do not depend at all on the match at hand. To accomplish this, we collect data on the players from a set of tournaments leading up to the US Open each year. This data is referred to as the auxiliary data in the last column of Table 1. For each match, the auxiliary data consists of the number of points won by each player while serving and while returning.¹² We use this to compute, via maximum likelihood, the players’ serving and returning abilities by regressing the aggregate point outcomes on indicators for whether a given player is serving or returning. The estimated coefficients on the indicators from this regression provide the measure of abilities, which we denote by s_i and r_i for the serving and returning abilities, respectively, of player i . Then we use these maximum-likelihood coefficients to obtain estimates of the probability that

¹¹With respect to the original data set, note that we have reduced the number of players by a factor of four, while the number of matches and points has only been reduced by a factor of only two, which is helpful for the identification of individual characteristics.

¹²We do not have point-level information in the auxiliary data, so we cannot compute point importance values in these matches.

i wins the point when serving against j , by:

$$q_i^0 = \frac{1}{1 + \exp(-(s_i - r_j))}.$$

We need q to reflect players' beliefs about point outcomes at each stage of the match. At the beginning of the match, these beliefs should be captured by the outcomes of points played in previous tournaments. Thus, for each match, we take the initial value of q to be given by q^0 . As the match unfolds, the players should then update their beliefs according to the outcomes of the points played. In order to avoid a complex specification of Bayesian updating, we adopt a simple approach to this updating at the level of the set. Specifically, we assume that the values of q in set $s > 1$ of a match are the simple average of the initial values, q^0 , and the realized winning frequencies at service from the previous $s - 1$ sets. In this manner, players' beliefs, starting from the basis of historical averages, respond to data from the current match and, as the match progresses, get closer to the realized service winning probabilities of the current match.

Given these values of q to be used for each point in the match, we can proceed to construct importance values according to the definition of the previous section. At any particular point, the importance is defined relative to the current values of q , assuming that all remaining points are won and lost with the relevant probabilities.

We emphasize, however, that the results below do not hinge on this exact construction of q , and its implied values of importance. The qualitative results are the same under alternative specifications. In particular, we obtain very similar results when the service winning probabilities are fixed throughout the match and equal to the realized frequencies in the match itself.¹³

3.2. The model. The basis of our approach is to model the outcome of each point as depending on the server's ability to serve, the returner's ability to return, and the interaction of the importance of the point with each player's critical ability.

¹³The reason we do not adopt this similar approach is that it creates simultaneity problems with the estimation. The issue is that this procedure requires one to use the point-by-point data first to construct the measure of importance as a function of q , and then to regress players' abilities on the importance values so obtained.

Each point in our sample corresponds to a single observation in our primary regression.

Each match is a competition between two particular players, i and j . For each match, one of the players is arbitrarily labeled “Player 1” and his opponent is labeled as “Player 2.” Thus, a given player may be labeled as Player 1 in some matches and Player 2 in other matches. We introduce the following notation. For each player i , let δ_i^S (δ_i^R) denote i ’s serving (returning) dummy variable. It takes the value of zero for those observations, i.e., points, in which i does not serve (return). It takes value +1 for observations in which i serves (returns) in the role of Player 1, and value -1 when i serves (returns) in the role of Player 2. Let δ_i^C denote player i ’s critical-ability dummy, which takes value +1 (-1) for observations in which i has the role of Player 1 (Player 2), whether he is serving or returning, and value zero in observations not involving i . Notice that $\delta_i^C = \delta_i^S + \delta_i^R$ but, whenever δ_i^C appears, it interacts with the point’s importance, so that the model is full rank. We set the dependent variable to be the probability with which (the arbitrarily designated, but fixed) Player 1 wins the point. Recall that θ denotes a score in the tennis match, $W(\theta)$ the probability that Player 1 wins the point to be played at score θ , and $I(\theta)$ its importance. Finally, we use a demeaned vector of point importance values, denoted by $\hat{I}(\theta)$.¹⁴

The basic form of our regression is then

$$(1) \quad \Pr(W(\theta) | \theta) = \Phi \left(\beta_0 + \sum_{i=1}^n \left(\beta_i^S \delta_i^S + \beta_i^R \delta_i^R + \beta_i^C \delta_i^C \hat{I}(\theta) \right) \right)$$

for all points in the sample, where Φ is the logistic cumulative distribution; in other words, we run a logit regression.

The abilities of each player i are described by the triple of coefficients $(\beta_i^S, \beta_i^R, \beta_i^C)$.

4. CRITICAL ABILITIES

We have two main goals in this section. First, we demonstrate that there is significant heterogeneity in critical abilities across professional tennis players.

¹⁴In the next section we explain why demeaning is important for our approach.

Second, we show that this variation is significantly related to these players' career success, as measured by their professional rankings and ratings.

Recall that we require that the measure of critical ability be independent of the general ability to win more points on average. However, given our econometric specification in equation (1), the fact that the importances of all points are by definition (strictly) positive would mean that a player with high critical ability would tend to win more points on average than an otherwise identical player with a low critical ability. In order to circumvent this effect, we demean the importance variable. When we do so, a higher critical ability has roughly no effect on the average number of points won, but instead implies that more important points are won with higher probability, and less important points are won with lower probability, all else being equal.

We demean importance match by match. Thus, a high critical ability means that a player is likely to win the relatively more important points in the match, independent of their absolute level of importance. Demeaning at the match level ensures that all players have played points of the same average importance.¹⁵

4.1. Heterogeneity. Serving and returning abilities are clearly crucial to tennis performance and, not surprisingly, in all the regression specifications, our tests show that players are highly heterogeneous in these variables. In regard to critical abilities, we adopt the approach described in Section 3.1, whereby serving probabilities are estimated exogenously and updated during the match based on observed point outcomes set by set (and the corresponding importance values are denoted simply by PiM).

The first two columns in Table 2 show that players' critical abilities differ significantly and are important in explaining point outcomes. In the first column,

¹⁵Demeaning importance at the match level implies that every match has the same average point importance (of zero). Thus even if a player plays several matches, the average importance of all those points is still zero. Therefore, critical ability does not affect the number of points a player is likely to win.

TABLE 2. Joint significance for ability-identification specifications

	Regression Specifications			
Serve	0*** (551.5)	0*** (550.0)	0*** (510.2)	0*** (550.7)
Return	0*** (1009.3)	0*** (1016.5)	0*** (949.2)	0*** (1010.0)
Critical (<i>PiM</i>)	0.062* (114.8)	0.01** (127.6)		
Critical (BP)			0.79 (81.8)	
Critical ($q = 0.63$)				0.38 (96.6)
Endurance		0.0002*** (149.1)		

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Joint significance tests are based on a χ^2 distribution with d.f. 93, except for Return, which has d.f. 94, because of how the regressions are normalized. The table reports p -values for joint significance of the relevant set of independent variables, with χ^2 statistics reported in parentheses. For joint significance values below 10^{-10} , we report zero.

point outcomes are regressed on the serving and returning variables along with *PiM*, and we find that critical abilities are jointly significant at the 6 percent level.

In the second column, we explore this finding further by adding a variable that corresponds to the point number in the match. This allows us to estimate an endurance effect, whereby a player may perform better or worse as the match goes on. While perhaps interesting in its own right, it is also important to include as a robustness check because it may be correlated with *PiM*. If later points tend to be more important, then we could be incorrectly attributing to critical abilities some effects on point outcomes that are actually due to differential endurance.¹⁶ When point outcomes are regressed on the serving and returning variables, *PiM*, and *endurance*, critical abilities are jointly significant at the 1 percent level, and *endurance* is highly significant, indicating that it is an important determinant of point outcomes, and also that players differ significantly on that dimension.

¹⁶We demean the endurance variable for the same reason that we demean importance.

Two extra regressions confirm our assertion that properly measuring the importance of points depends crucially on taking into account the relative abilities of the players. These results show that other less informative measures of importance do not adequately capture a meaningful notion of importance. For the third regression in Table 2, we classify points as either important or unimportant according to whether or not the point is a break point. For this specification, we get a p -value of 0.79 for the joint significance of the critical abilities. In the last regression, we take q , the probability of the server winning a given point, to be fixed at the global frequency of service points won in the full data. This frequency is $q = 0.63$. When we regress point outcomes on the serving and returning abilities along with the importance values obtained in this way, we find again that critical abilities are not jointly significant (p -value 0.38).

Since all the regressions confirm that players differ significantly in terms of their serving and returning abilities, the lack of significance in critical abilities in the latter two specifications is rather interesting. The heterogeneity in the former abilities implies that there is a lot of variation across matches (and between the two players within each match) in terms of the likelihood of winning any given service point. This variation should not be ignored when estimating players' critical abilities. A score that corresponds to an important point between two given players may be relatively unimportant between two other players of different (relative) abilities. Players not only react to the score, but must put this score in the context of their relative strengths when assessing its importance for the match.

These results allow us to conclude strongly in favor of heterogeneity in critical abilities, using the parsimonious approach in which each player is assumed to have constant characteristics among all tournaments. Furthermore, our findings are confirmed by a separate analysis in which we allow players' characteristics to

be tournament dependent.¹⁷ Note finally that the observed heterogeneity is not a product of a large set of players alone, since the same conclusion does not hold under more naive measures of the point's importance.

One may wonder whether the observed heterogeneity holds any meaningful significance. We answer this question next by relating players' individual characteristics to measures of their career success.

4.2. Player estimates. Given the evidence for heterogeneity in all players' abilities, we proceed to examine the estimates at the individual level. Table 3 contains the top twenty-five servers and returners, as well as those with the highest critical ability.¹⁸ It also contains the top twenty-five players according to their average ATP ratings in the period under study.¹⁹ When using ATP ratings, we always take them on a log scale. Furthermore, we average each player's ATP rating across those years in 1994-2006 in which he played the US Open, so that we get a representative measure of his average level of success in our data set.

In regard to serving and returning estimates, tennis experts would probably agree that these lists fairly well describe the tennis players who are known to be particularly good servers and returners. Table 4 shows how serving, returning, and critical abilities, as well as ratings, are all positively correlated.

Examining the list of players with the highest critical ability, we find many of the same top players, but also a significant number of somewhat lesser known players. The correlations between serving and critical ability and between returning and critical ability are 0.34 and 0.20, respectively. Most relevant for our purposes, the correlation between critical ability and overall rating is 0.37.

¹⁷Because of the tree structure of tournaments, this approach amounts to treating all matches independently. In our sample of 494 matches, all groups of variables are significant at the 0.001% level, both with and without the inclusion of endurance.

¹⁸We report estimates from our baseline regression, represented in the first column of Table 2. The results do not depend much on the exact specification. For example, the correlation between abilities with and without the inclusion of the endurance variable are very high: 0.999 for serving and returning abilities, and 0.979 for critical abilities.

¹⁹The ATP points system is the generally accepted metric for rating the performance of professional tennis players. Points are accumulated based on tournament outcomes according to a fixed points scheme, and reflect a player's results from the previous twelve months of play.

TABLE 3. Top players according to our estimates and average ATP ratings.

Top 25	Serving Ability		Returning Ability		Critical Ability		ATP Rating	
1	A.RODDICK	0.485*** (0.063)	L.HEWITT	0.316*** (0.050)	T.ROBREDO	7.700*** (1.774)	P.SAMPRAS	8.109
2	P.SAMPRAS	0.421*** (0.053)	R.FEDERER	0.288*** (0.050)	A.CORRETJA	6.165*** (2.008)	R.FEDERER	8.077
3	R.KRAJICEK	0.362*** (0.073)	K.KUCERA	0.272*** (0.081)	J.FERRERO	4.566*** (1.481)	M.STICH	7.856
4	R.FEDERER	0.282*** (0.054)	A.AGASSI	0.254*** (0.045)	A.COSTA	4.305 (3.506)	L.HEWITT	7.813
5	M.MIRNYI	0.254*** (0.071)	J.BJORKMAN	0.239*** (0.060)	M.ROSSET	3.109 (2.676)	A.AGASSI	7.78
6	M.STICH	0.234** (0.099)	M.YOUZHNY	0.224*** (0.087)	M.ZABALETA	3.105 (2.066)	A.RODDICK	7.779
7	A.AGASSI	0.217*** (0.000)	N.ESCUDE	0.214** (0.094)	G.POZZI	2.999 (2.708)	Y.KAFELNIKOV	7.766
8	P.RAFTER	0.216*** (0.062)	Y.KAFELNIKOV	0.211*** (0.055)	R.SCHUETTLER	2.932 (3.115)	G.KUERTEN	7.742
9	G.RUSEDISKI	0.212*** (0.060)	D.NALBANDIAN	0.198*** (0.065)	A.RODDICK	2.718* (1.627)	T.MUSTER	7.702
10	N.ESCUDE	0.193** (0.099)	G.CORIA	0.196*** (0.074)	G.IVANISEVIC	2.547 (2.452)	J.FERRERO	7.583
11	G.KUERTEN	0.178*** (0.067)	J.BLAKE	0.190*** (0.063)	B.BLACK	2.516 (2.157)	P.RAFTER	7.573
12	L.HEWITT	0.171*** (0.053)	P.KORDA	0.186** (0.080)	M.WOODFORDE	2.474 (4.129)	R.NADAL	7.523
13	M.LARSSON	0.170* (0.101)	A.RODDICK	0.170*** (0.059)	B.KARBACHER	2.436 (2.907)	P.KORDA	7.431
14	M.SAFIN	0.136*** (0.057)	G.CANAS	0.156* (0.097)	P.SAMPRAS	2.330*** (1.099)	T.HENMAN	7.407
15	B.BECKER	0.131 (0.089)	D.HRBATY	0.139** (0.072)	L.HEWITT	2.311** (1.312)	C.MOYA	7.377
16	T.MARTIN	0.130*** (0.053)	P.RAFTER	0.135*** (0.059)	N.ESCUDE	2.243 (2.246)	D.NALBANDIAN	7.37
17	J.BLAKE	0.125** (0.064)	V.SPADEA	0.122* (0.071)	A.MEDVEDEV	1.935 (1.903)	A.CORRETJA	7.293
18	X.MALISSE	0.117* (0.071)	H.LEE	0.118 (0.095)	P.RAFTER	1.826 (1.598)	M.SAFIN	7.281
19	W.ARTHURS	0.113 (0.080)	S.SARGSIAN	0.109 (0.077)	S.DOSEDEL	1.772 (2.566)	B.BECKER	7.264
20	M.ZABALETA	0.105 (0.092)	J.COURIER	0.099 (0.087)	M.SAFIN	1.770* (1.225)	R.KRAJICEK	7.236
21	M.DAMM	0.097 (0.086)	M.ZABALETA	0.094 (0.091)	W.ARTHURS	1.683 (1.939)	G.CORIA	7.22
22	J.COURIER	0.083 (0.086)	T.ENQVIST	0.093 (0.073)	A.AGASSI	1.675** (0.000)	J.COURIER	7.192
23	D.NALBANDIAN	0.083 (0.065)	A.CLEMENT	0.089* (0.061)	R.FEDERER	1.479 (1.411)	C.PIOLINE	7.188
24	R.GINEPRI	0.082 (0.069)	T.HAAS	0.088** (0.055)	C.PIOLINE	1.438 (1.490)	T.ROBREDO	7.168
25	J.FERRERO	0.079 (0.067)	M.SAFIN	0.084* (0.056)	T.MARTIN	1.437 (1.193)	A.MEDVEDEV	7.16

*** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$; significance levels are based on two-sided tests. The estimates in each category are normalized so that the average is zero and the tests are performed against this average. The standard errors are reported in parentheses.

To represent the magnitude of heterogeneity in critical ability, Figure 2 depicts a kernel density of the estimated distribution in our sample of players. By construction, the mean ability is zero, but the density shows substantial variation in estimated abilities.

TABLE 4. Correlations between abilities

	Rating	Serving	Returning	Critical
Rating	1	-	-	-
Serving	0.55	1	-	-
Returning	0.39	0.26	1	-
Critical	0.37	0.34	0.20	1

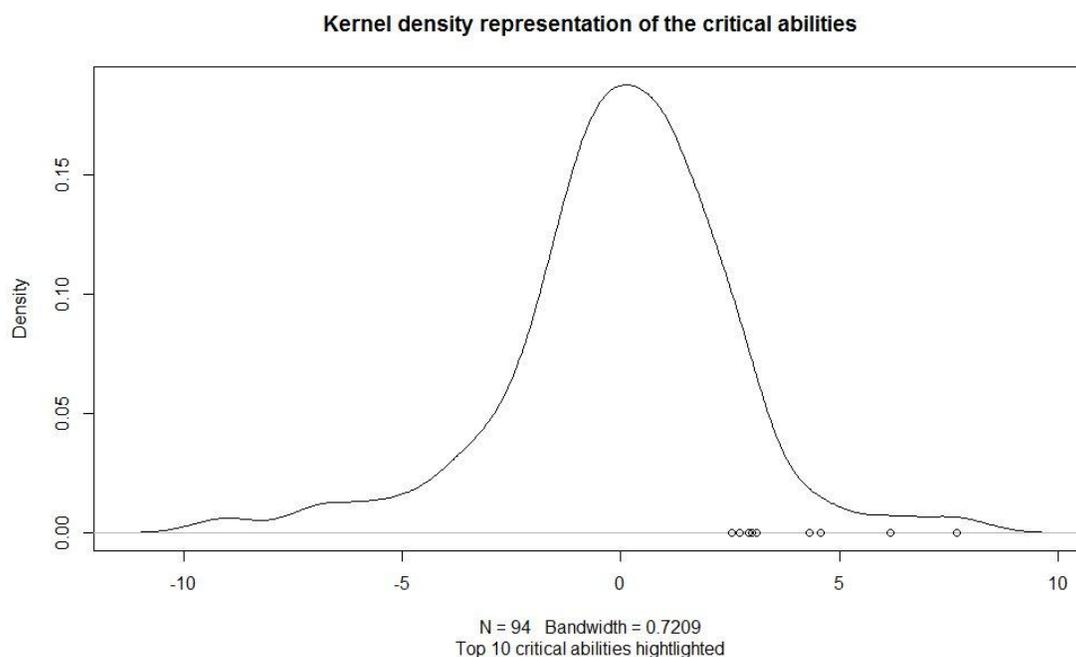


FIGURE 2. Kernel density representation of critical abilities among our set of 94 players.

In Table 5 we explore further the implied marginal effects of this variation in critical ability. This table represents players' probabilities of winning a point against the average player of our population when they are on serve. In the second column we represent these probabilities in points of zero importance. The third column represents the same probability but in a point with high importance (99th percentile); the players included in the table are the top 25 players according to this column. Columns 4 and 5 represent, respectively, this probability when serving and critical abilities are increased by one standard deviation. It can be

readily verified from the table that the importance of the point to be played has a large impact on the predicted winning probabilities. The average change in these probabilities across the 94 players in our sample is around 9% and for some players the effect is much higher. For instance, if we take T. Robredo, the player with highest critical ability, we find that he is more dangerous in these points than even the best server in our sample, who is A. Roddick (and who also happens to have a good critical ability). On the other hand, from Columns 4 and 5 it can be seen that, for these highly important points, one standard deviation increase in the critical ability of a player has a bigger impact than a one standard deviation increase in his serving ability.

TABLE 5. Probabilities of winning a point on serve.

Player	$PiM = 0$	99th % PiM		
		Estimated SA & CA	SA + 1 st.d.	CA + 1 st.d.
T. ROBREDO	0.633	0.826	0.847	0.869
A. RODDICK	0.739	0.802	0.825	0.850
A. CORRETJA	0.624	0.789	0.813	0.839
P. SAMPRAS	0.726	0.783	0.807	0.835
J. FERRERO	0.653	0.775	0.800	0.828
A. COSTA	0.634	0.753	0.780	0.810
M. ZABALETA	0.659	0.745	0.772	0.803
N. ESCUDE	0.679	0.740	0.768	0.799
R. FEDERER	0.698	0.737	0.765	0.797
L. HEWITT	0.674	0.737	0.765	0.797
P. RAFTER	0.684	0.733	0.762	0.794
A. AGASSI	0.684	0.730	0.758	0.791
M. ROSSET	0.631	0.721	0.750	0.783
M. WOODFORDE	0.647	0.718	0.747	0.781
M. STICH	0.688	0.716	0.746	0.779
M. SAFIN	0.666	0.716	0.746	0.779
W. ARTHURS	0.661	0.709	0.739	0.773
T. MARTIN	0.665	0.706	0.736	0.771
R. KRAJICEK	0.715	0.701	0.731	0.766
G. IVANISEVIC	0.624	0.699	0.730	0.765
G. RUSEDISKI	0.683	0.694	0.725	0.761
B. KARBACHER	0.620	0.692	0.723	0.759
R. GINEPRI	0.654	0.691	0.722	0.758
R. SCHUETTLER	0.603	0.691	0.722	0.758

We turn now to assessing the relationship between critical ability and career success.

4.3. Career effect of heterogeneity in ability. Our analysis so far demonstrates significant variation not only in players' serving and returning performance, but also in their critical abilities. The first two abilities clearly translate into making a player stronger, in the sense that the higher his serving or returning ability, the more points he will win, all else being equal. Critical ability also translates into making a player strong but in the (different) sense that, given a baseline probability of winning points, the more important the point is in determining the outcome of the match, the more likely the player is to win that point. This generally leads to a higher probability of winning a match, all else being equal, without changing the frequency of points won. We now seek to demonstrate that, like serving and returning ability, critical ability is strongly related to career success.

The main question we want to address is: To what extent are players' serving, returning, and critical abilities able to explain career success as measured by the ATP ratings? Table 4 already hints at the potential of critical ability to be a key factor in success, as its correlation with the ATP rating is 0.37, approximately as strong as the correlation between a player's returning ability and his rating. To explore this question more rigorously, we take the regression estimates for serving, returning, and critical abilities as estimated before, and use them as regressors to explain ATP ratings. By construction, high abilities translate into better performances at the US Open and, since this tournament is among those used to compute the ATP ratings, it would not be surprising to see that these abilities (estimated from the US Open tournaments) are significant in explaining ATP ratings. Therefore, in our regressions we use the average ratings of the players after having subtracted the points obtained at the US Open.

In Table 6 we report the results of regressing players' abilities on ratings.²⁰ Since some of the regressors are estimates coming from another regression, we run

²⁰We have replicated the analysis using instead ATP rankings as the dependent variable. The qualitative results are the same.

generalized least squares (GLS) instead of ordinary least squares (OLS).²¹ In the first column of Table 6 we regress ratings on serving and returning abilities. Both variables are significant at the 1 percent level, showing that the individual-level estimates we recovered from the first set of regressions are useful in explaining the professional success of tennis players.

TABLE 6. Player abilities and career success

Dep. Var.	Rating	Rating	Rating	Rating
Intercept	7.82*** (0.22)	7.80*** (0.21)	7.76*** (0.21)	3.41 (19.88)
Serve	1.57*** (0.27)	1.40*** (0.28)	1.34*** (0.29)	1.45*** (0.33)
Return	0.86*** (0.27)	0.80*** (0.27)	0.69*** (0.29)	0.74** (0.32)
Critical		0.035** (0.017)	0.028** (0.016)	0.033** (0.016)
Endurance			61.7 (52.6)	45.8 (54.1)
Birth				0.003 (0.01)
Height				-0.003 (0.007)
Lefty				0.091 (0.12)
GDP				$-7 \cdot 10^{-7}$ ($2 \cdot 10^{-6}$)
Bollettieri				0.13 (0.13)
USA				-0.23* (0.12)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; significance levels are based on one-sided tests of each ability having a positive relationship with success. The dependent variable, rating, is used in log scale. The controls are date of birth, height, left/right-handed, GDP of home country, if they attended the renowned Bollettieri tennis school, and if the person is American. Setting aside the controls, each specification uses as regressors abilities estimated in the first stage regressions reported in Table 2.

²¹In Appendix C we discuss this issue in some detail. It is worth noting that the qualitative results are the same using GLS and OLS. We have also obtained similar results using bootstrap techniques, although significance was slightly lower.

Our second main result comes from adding critical abilities to this regression. The second column in Table 6 shows that critical abilities are significant at the 5 percent level, indicating that the ability to perform better at more important points is significantly related to a player's overall success. The last two regressions show that the significance of critical ability is quite robust to the introduction of the endurance ability of the players and of other controls. One of these control variables, the player's year of birth, serves as a proxy for player experience, which might a priori be positively correlated with both critical ability and rating.

Both the correlation between the abilities and rating and the regression results validate our estimation of players' abilities. In particular, this is a clear indication that the heterogeneity of critical abilities is substantial, and that our measure of individual critical abilities is meaningful. The regression analysis also clearly shows that individual success is strongly related to each of the three abilities, including critical ability. We take this result as an indication that the ability to adjust performance according to the importance of the situation is one determinant of career success.

Figure 3 plots ATP rating (left panel) and ATP ranking (right panel) against critical ability in our sample of players. It is evident from both plots that higher critical ability is associated with greater career success. Further, all of the top performing players have better than average critical ability, suggesting that it is not possible to be a top professional without responding well to the most important points. For lower rated players, however, the distribution of critical ability is quite varied.

One way to express the effect of critical ability on career success is to consider standardized coefficients. Using the average standard error of the coefficients of each of the three abilities from the regression reported in Table 2 (first column) and the coefficients of these abilities in Table 6 (second column), we find that, relative to the average player, the effect on a player's ATP rating (in the log scale) of increasing his serving ability by one standard deviation in the population is worth 0.11 ATP points, while a comparable increase in his returning ability

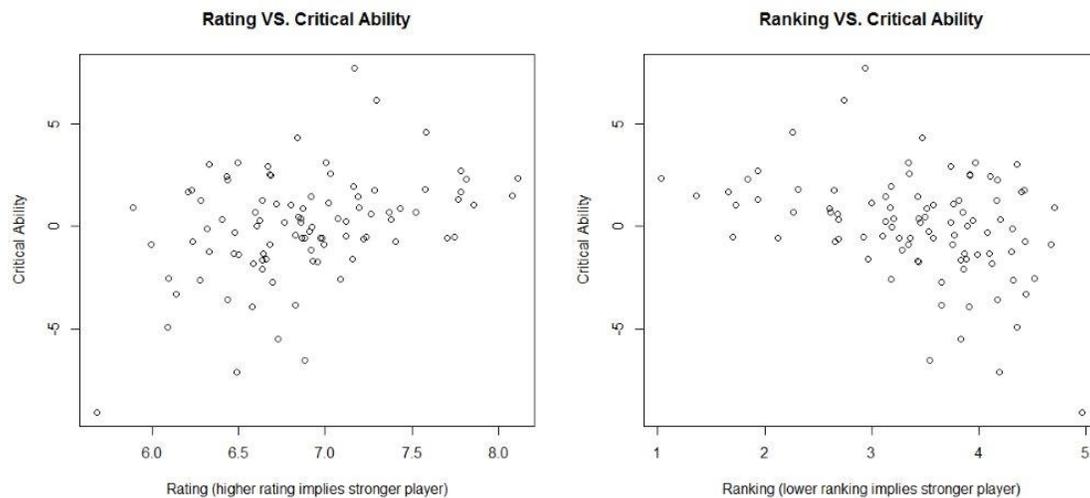


FIGURE 3. Scatter plots of ATP rating (left panel) and ATP ranking (right panel) against critical ability.

is worth 0.06 points; interestingly, a comparable increase in his critical ability is worth 0.07 points. These effects can be judged by looking at the last column in Table 3.

5. CONCLUSION

The main purpose motivating this study has been to understand the potential consequences of reactions to critical situations. In particular, we hypothesize that individuals will respond to critical situations heterogeneously, and that these reactions will translate into observable effects on long-term success. We assessed the empirical content of this argument in the specific context of professional tennis tournaments. We found strong evidence of persistent heterogeneity in players' responses to the importance of the situation, which we called their critical ability, and a significant relationship between individual critical ability and career success.

Only relatively recently have economic analyses taken account of the fact that individuals may not always perform as they intend and that, as a result, the way in which performance is adjusted to context matters. This paper contributes to

the body of research providing evidence that such effects can be important determinants of individual success. Using the setting of professional tennis provided particularly clean data with which to estimate critical abilities, but we believe the effects we find in tennis are likely to be important elsewhere as well.

To build on these results, an important goal of future work should be to understand the impact of psychological responses in other competitive environments, and especially in market settings. For instance, matching in labor markets is one particular context where we expect critical abilities may play an important role. Workers are likely to vary both in their general abilities and also in their critical ability, i.e., their capacity to adjust their performance to the importance of the task. Jobs are likely to vary in how important these two skills are. If these assumptions hold true, then an efficient matching of workers to firms requires accounting for the heterogeneity in workers' critical ability.

Finally, given the evidence that critical ability is related to success, it seems important to better understand where high critical ability comes from. It may be that some aspects of critical ability relate to psychological skills that are difficult to learn. On the other hand, it may be that critical ability can be improved by training, or simply through experience. Understanding the determinants of critical ability will allow one to better predict which individuals can be expected to perform at their best in important situations, a valuable asset to firms and to society as a whole.

6. ACKNOWLEDGEMENTS

We thank Caterina Calsamiglia, Sven Feldman, Ignacio Palacios-Huerta, Pedro Rey-Biel, Betsy Sinclair, and Scott Stern for helpful comments, as well as audiences at Málaga University, Northwestern University, and Paris School of Economics. The associate editor and two anonymous referees provided valuable suggestions. Julio González-Díaz acknowledges financial support from the European Commission through a Marie Curie fellowship and from the Spanish Ministry for Science and Innovation and FEDER through project ECO2008-03484-C02-02.

We also thank the United States Tennis Association (USTA) for kindly providing us with point-level data on twelve US Open tournaments.

REFERENCES

- ABRAMITZKY, R., L. EINAV, S. KOLKOWITZ, AND R. MILL (2011): “On the optimality of line call challenges in professional tennis,” *International Economic Review*, forthcoming.
- ALBERT, J. (2007): “Measuring Clutch Play,” in *Statistical thinking in sports*, Chapman and Hall/CRC, 111–133.
- ALBERT, J. AND J. BENNET (2003): “Measuring Clutch Play,” in *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*, Springer, 267–322.
- APESTEGUIA, J. AND I. PALACIOS-HUERTA (2010): “Psychological Pressure in Competitive Environments: Evidence from a Randomized Natural Experiment,” *American Economic Review*, forthcoming.
- ARIELY, D., U. GNEEZY, G. LOEWENSTEIN, AND N. MAZAR (2009): “Large Stakes and Big Mistakes,” *Review of Economic Studies*, 76, 451–469.
- BAUMEISTER R. F. AND A. STEINHILBER (1984): “Paradoxical effects of supportive audiences on performance under pressure: The home field disadvantage in sports championships” *Journal of Personality and Social Psychology*, 47, 85–93.
- CHIAPPORI, P.-A., S. LEVITT, AND T. GROSECLOSE (2002): “Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer,” *American Economic Review*, 92(4): 1138–1151.
- DOHMEN, T. (2008): “Do professionals choke under pressure?” *Journal of Economic Behavior and Organization*, 65, 636–653.
- FROST, C. AND S. THOMPSON (2000): “Correcting for regression dilution bias: comparison of methods for a single predictor variable,” *Journal of the Royal Statistical Society Series A*, 163, 173–190.
- HUME, D. (1739): *A Treatise of Human Nature*, North-Holland.

- KAHNEMAN, A. AND D. TVERSKY (1979): “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47, 263–291.
- KLAASEN, F. AND J. MAGNUS (2001): “Are Points in Tennis Independent and Identically Distributed? Evidence from a Dynamic Binary Panel Data Model,” *Journal of the American Statistical Association*, 96, 500–509.
- MORRIS, C. (1977): “The Most Important Points in Tennis,” in *Optimal Strategies in Sports*, ed. by S. P. Kadany and R. E. Machol, Amsterdam: North-Holland Publishing Company, 131–140.
- OTTEN, M. (2009): “Choking vs. clutch performance: A study of sport performance under pressure.” *Journal of Sport & Exercise Psychology*, 31, 583–601.
- PASERMAN, D. (2010): “Gender Differences in Performance in Competitive Environments: Evidence from Professional Tennis Players,” mimeo.
- WALKER, M. AND J. WOODERS (2001): “Minimax Play at Wimbledon,” *American Economic Review*, 91, 1521–1538.
- WALKER, M., J. WOODERS, AND R. AMIR (2010): “Equilibrium play in matches: Binary Markov games,” *Games and Economic Behavior*, in press.

APPENDIX A. PROOF OF PROPOSITION 1

The proof can be illustrated by showing that the importance of a point θ in determining the outcome of the current set (denoted $PiS(\theta)$), is equal to $PiG(\theta) * GiS(\theta)$. Define the following notation: $W_G(\theta)$ (or $L_G(\theta)$) is the event in which player 1 wins (or loses) the current game at point θ , while $W_S(\theta)$ (or $L_S(\theta)$) is the event in which player 1 wins (or loses) the current set at point θ . Suppressing the

dependence on θ , we have the following:

$$\begin{aligned}
PiS &= P(W_S | W) - P(W_S | L) \\
&= P(W_S | W_G)P(W_G | W) + P(W_S | L_G)P(L_G | W) - \\
&\quad [P(W_S | W_G)P(W_G | L) + P(W_S | L_G)P(L_G | L)] \\
&= P(W_S | W_G)[P(W_G | W) - P(W_G | L)] + P(W_S | L_G)[P(L_G | W) - P(L_G | L)] \\
&= P(W_S | W_G) * PiG - P(W_S | L_G) * PiG \\
&= PiG * GiS.
\end{aligned}$$

The first equality holds by definition, the second expresses Bayes' Law, the third rearranges terms, and the remainder apply our definitions. The argument to complete the proof proceeds in an analogous fashion.

APPENDIX B. QUALITATIVE FEATURES OF POINT IMPORTANCE

Unless the players are very unequally matched, there is substantial variation in the importance of points in a typical match.²² It is this natural variability that allows us to identify critical abilities.

As explained in the text, one component of a point's importance is its importance in determining the outcome of its game (the first term in the decomposition of Proposition 1). We begin by describing how this game importance (PiG) responds to the relative strength of the server at various scores in the game.

Suppose that player 1 is serving and that, given his ability to serve and his opponent's ability to return the serve, player 1 has a probability q of winning any given service point, on average. Serving is generally an advantage, and so typically $q > 1/2$. What are the importance values of points in the game at the different possible scores, and how do they depend on q ? We start with the tied score, $p_1 = p_2 \geq 2$, known as "deuce."²³ As in Appendix A, W_G denotes the

²²When one player is much stronger than the other, his probability of winning the match is close to one for all scores except those at which he is very close to already losing the match. Since such scores have very low probability of being realized, the importance of most points is close to zero in this case.

²³The score of $p_1 = p_2 = 2$ is strategically equivalent, but not known as "deuce".

event in which player 1 wins the game. Let D denote a deuce score with player 1 serving. We have $\Pr(W_G | D) = q^2 + 2q(1 - q) \Pr(W_G | D)$, so that

$$\Pr(W_G | D) = \frac{q^2}{1 - 2q(1 - q)}.$$

The probability of player 1 winning the game at all other points can be computed recursively, using this relationship:

$$\Pr(W_G | p_1, p_2) = q \Pr(W_G | p_1 + 1, p_2) + (1 - q) \Pr(W_G | p_1, p_2 + 1).$$

Then, importance (PiG) is computed by taking the relevant differences in outcome probabilities as before.

Figure 1 (right panel) in the text shows the PiG importance of each score in our data at the average value of q , which is approximately 0.63. Notice that the most game-important point is $(q_1, q_2) = (2, 3)$, a score known as break point. This coincides with the conventional wisdom that break points represent important situations in the match. This finding arises under our definition because typically $q > 1/2$, and so the server is expected to win service games. The most pivotal situations are those in which this outcome has one chance to be reversed, which is when the score is $(2, 3)$.

If we go back to Figure 1 in the text, we see that its right panel shows the importance of all possible situations within a game, but only for a particular value of q . On the other hand, the left panel of Figure 1 plots the importance of several prominent scores as a function of q , in order to show how relative importance changes depending on the relative strength of the server. Notice that the relative game importance of points depends very much on q . In other words, the points that are most important in determining the outcome of a game depend on the distribution of point outcomes; moreover, their rankings vary nontrivially in the range of q observed in our data.

For instance, of the points shown in Figure 1, we see that $(2, 2)$ is the most important point in the game whenever q is below about 0.66. However, for slightly larger values of q , $(0, 2)$ is the more pivotal point, and then for even larger values of q , $(0, 3)$ is the most important point. As the server becomes relatively stronger

and stronger, the only points with nontrivial importance are break points: those at which the point would award the game to the returner. The importance of all other points tends to zero as q grows, since the server will then win the game with arbitrarily high probability, independent of the outcome of the current point.

Thus, when one considers only the importance of a point in the game, a complex relationship already exists between importance and the characteristics of the players involved. However, one must also consider the importance of the game in the set, as well as the set in the match. Notice that when one considers games and sets, their importance will depend on each player's probability of winning a service point for both players. So define q_i as the probability that player i wins a typical service point, $i = 1, 2$. In Figure 4 the left and right panels present the importance values of GiS and SiM , respectively, when $q_1 = 0.75$ and $q_2 = 0.63$.

$g_2 \backslash g_1$	0	1	2	3	4	5				
0	0.34	0.13	0.12	0.02	0.01	0.00				
1	0.39	0.40	0.14	0.13	0.01	0.01	$s_2 \backslash s_1$	0	1	2
2	0.33	0.46	0.47	0.16	0.13	0.01	0	0.12	0.08	0.03
3	0.30	0.31	0.55	0.57	0.17	0.14	1	0.36	0.29	0.17
4	0.08	0.25	0.26	0.67	0.69	0.17	2	0.69	0.83	1.0
5	0.03	0.03	0.16	0.17	0.83	0.69				

FIGURE 4. The importance of a game in the current set, GIS (left panel), and the importance of each set in a match, SiM (right panel), for $q_1 = 0.75$ and $q_2 = 0.63$.

In the example given, the most important game is at the game score $(4, 5)$, whose outcome affects the outcome of the set with importance 0.83. Notice that this contrasts with the conventional wisdom that the seventh game in the set is the most important.²⁴ Of course, these importance values will vary with the particular values of q_1 and q_2 for any given match.

Figure 4 (right panel) shows, for instance, that the final set determines with probability one the outcome of the match whenever that set is reached. Also notice that the most important sets are when player 2, the weaker player, is winning.

²⁴The origin of this “wisdom” is disputed. See, for instance, Klaasen and Magnus (2001).

When player 1 is winning, the sets are less important because even if player 1 loses the set, he will still win the match with high probability.

To illustrate the heterogeneity in points' importance, consider the following scenario from the 1995 US Open final match between Andre Agassi and Pete Sampras. The difference in purse winnings between the winner and loser of the match was \$287,500. In that match, Agassi won 65 percent of his service points, and Sampras won 72 percent of his service points, so $q_A = 0.65$ and $q_S = 0.72$. At one point late in the match, the score was such that the point importance was 0.0008, implying that \$380 was riding on the point in expected terms. However, at an earlier point when Agassi, the weaker player in the match, had a break point, the importance was 0.13, so that \$64,000 was riding on the point in expected terms. Such variation is typical in later rounds of the tournaments.

APPENDIX C. TECHNICAL DIFFICULTIES FOR THE ANALYSIS IN SECTION 4.3

As we already mentioned in the text, there are several reasons why one has to be careful with the regressions of a player's ratings on a player's abilities. Most importantly, the regressors are themselves the estimates of a previous regression. Since they should be viewed as having been measured with noise, using them in a secondary regression presents some issues.

First, there is the fact that the noise inherent in our estimates may produce attenuation bias.²⁵ Although we do not have a good way to correct for this effect, it is clear that if the results suggest significant effects of players' abilities on ratings, then accounting for attenuation bias would only strengthen the significance of the results.

Second, we want to use the information from our set of estimated abilities optimally. In particular, we have 94*3 relevant estimates of player abilities, and the complete variance-covariance matrix. The generalized least-squares model we estimated is a natural way to use this information efficiently. We also performed

²⁵See, for example, Frost and Thompson (2000).

a bootstrap procedure as a robustness check, and the results were consistent with our findings from the GLS.

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, UNIVERSITY OF SANTIAGO DE COMPOSTELA

PARIS SCHOOL OF ECONOMICS AND LONDON SCHOOL OF ECONOMICS

MEDS, KELLOGG SCHOOL OF MANAGEMENT, NORTHWESTERN UNIVERSITY