

Three prominent tournament formats: predictive power and costs*

Dmitry Ryvkin[†] and Andreas Ortmann[‡]

This version: June 4, 2006

Abstract

We analyze tournaments of heterogeneous players from an organizer's perspective. Using a simple model of a noisy tournament, we demonstrate how the likelihood of selecting the best player, here termed the "predictive power" of a tournament, depends on the tournament format, the distribution of players' types, and the overall noise level. We formalize the organizer's decision problem for varying time and measurement costs and compare the predictive power of three widely used tournament formats – contests, binary elimination tournaments, and round-robin tournaments. We show which formats are preferred in the various scenarios and find that for certain parameter constellations, certain formats are not viable.

Keywords: tournaments, design, predictive power

JEL Classification: C73, C90, D21

*We appreciate comments of Kai Konrad, Stergios Skaperdas and Stefan Szymanski on earlier versions of this paper. Special thanks to Dirk Engelmann.

[†]Florida State University, Tallahassee, FL, USA; dmitry.ryvkin@cerge-ei.cz; d.ryvkin@gmail.com

[‡]Center for Economic Research and Graduate Education – Economics Institute, Prague, Czech Republic; andreas.ortmann@cerge-ei.cz; aortmann@yahoo.com

1 Introduction

Agents (individuals or teams) are usually rewarded based on their performance. Often it is the relative performance that matters. As a means of assessing the relative performance of agents, principals extensively use tournaments.

A *tournament* is a procedure that ranks a set of agents. Such a ranking can be constructed in various ways. Prominent examples are contests, binary elimination tournaments, and round-robin tournaments. Contests are essentially one-shot all-pay auctions where submission of the highest contribution (effort) does not necessarily lead to a win. Properties of contests have been widely discussed in the literature.¹ In contrast, binary elimination tournaments and round-robin tournaments are polar cases of tournament formats that compare agents pairwise and sequentially in various degrees of completeness. These formats allow (most) agents to perform repeatedly, typically against a stream of ever-changing opponents. A binary elimination tournament is an incomplete multi-stage pairwise matching format whose exact realization, and efficiency, depends on the initial seeding and the history of play.² A round-robin tournament is a complete pairwise matching format where the winner is determined by point counting.³

Sport events provide us with a simple and useful language to describe tournaments.

¹For a theoretical discussion see e.g. Tullock (1980), Lazear and Rosen (1981), Green and Stokey (1983), Taylor (1995), Moldovanu and Sela (2001), Hvide (2002), Baye and Hoppe (2003), Cornes and Hartley (2005); for empirical results see e.g. Knoeber and Thurman (1994), Eriksson (1999); for experimental results see e.g. Schotter and Weigelt (1992), Gneezy *et al.* (2003), Orrison *et al.* (2004). See also reviews by Lazear (1999), Prendergast (1999), Frick (2003), Szymanski (2003), and papers in Lockard and Tullock (2001).

²Binary elimination tournaments model the competition of workers in hierarchical organizations, see e.g. Rosen (1986), O’Flaherty and Siow (1995). For empirical results on elimination in sports see Ehrenberg and Bognanno (1990), Bognanno (2001). For various other aspects see also Horen and Reizman (1985), Knuth (1987), Gradstein (1998), Gradstein and Konrad (1999), Groh *et al.* (2004).

³Round-robin tournaments have been discussed in the context of public choice models such as voting schemes and decision rules in committees (see, e.g., Levin and Nalebuff 1995, Ben-Yashir and Nitzan 1997, Esteban and Ray 2001). In mathematics, round-robin tournaments have been studied as complete directed graphs (see Harary and Moser 1966 for a review; Moon and Pullman 1970 for a discussion of tournament matrices). Importantly, Rubinstein (1980) shows that the point counting scheme that assigns 1 point to the winner and 0 to the loser of a match, and then sums up every player’s points across all matches he or she played, is a very “good” point counting scheme in the sense that it satisfies certain natural axioms. Rubinstein (1980) also shows that it is *the only* such scheme.

Competing agents are typically called *players*; possible pairwise comparisons of players are called *matches*. In every match, there is a *winner* and a *loser*, or there is a tie. We use this terminology throughout the paper.

Noise affects the performance of players either positively or negatively. In the presence of noise a tournament can be thought of as a probabilistic device whose output, the *ranking*, is a statistic of the “true” ordering of the set of players. Such an ordering identifies, in particular, who is the best player *ex ante*.

In terms of organizers’ objectives, we focus on the informational utility of tournaments. We leave the incentive provision problem aside and essentially view tournaments as estimators of the unobserved ranking of players. We characterize such estimators quantitatively by their predictive power – a measure of how reliably a tournament identifies better players. This view of tournaments is relatively new in the literature, although organizers’ objectives stemming from it are important in many applications. In situations such as recruitment and promotion in firms or selection of public finance projects, a reasonable organizer’s objective would be maximization of the predictive power. In sports, where uncertainty and upsets increase the “entertainment value” of tournaments for spectators, organizers may tend to minimize the predictive power.

In the present paper, we explore theoretically (analytically and by way of computational simulations), the predictive power of three prominent tournament formats: contests, binary elimination tournaments, and round-robin tournaments.⁴ We explore predictive power as a function of noise level, number of players, and distribution of players’ abilities. Using a simple model of organizational costs, similar to the one in Bolton and Dewatripont (1994), Gradstein (1998), we analyze the decision problem of an organizer who selects an optimal tournament format.

⁴These issues have been partly addressed earlier by in the operations research and statistics community, see, for example, Searls (1963), Appleton (1995), McGarry and Schutz (1997). Our study has a different focus. See also Harbaugh and Klumpp (2005), Klumpp and Polborn (2005) for some results regarding the probability of the best players advancing in tournaments in the contexts of, respectively, sports and presidential elections.

Our paper contributes to the literature by analyzing and comparing the probabilistic properties of these tournament formats and by studying their properties across what we consider the most prominent distributions of players' abilities. The investigation of the three formats is an important step towards solving the problem of optimal tournament design.⁵

The paper is organized as follows. In Section 2 we specify a model of a noisy tournament of heterogeneous players and formulate the organizer's decision problem. In Section 3 we calculate the predictive power of the three tournament formats. In Section 4 we present and discuss the dependence of the predictive power of the three formats on the number of players and the noise level for different distributions of players' abilities. In Section 5 we illustrate the solution to the organizer's decision problem and show how the organizer will switch between the three tournament formats as costs change. Section 6 concludes.

2 The model

In this Section we present the model of a noisy tournament of heterogeneous players that we use for our subsequent analysis. The model includes the following ingredients: (i) a tournament format; (ii) players; (iii) noise; (iv) costs; (v) organizer's decision problem.

2.1 Tournament formats

A tournament format is an institutionally defined set of rules by which players are matched and the winner is determined. We consider three prominent tournament formats:⁶

⁵See Ryvkin (2005) for a general approach to the tournament design problem; see more generally Roth (2002), for a discussion of how simulations with artificial agents and experiments with human subjects can extend theoretical models, which become too complex in engineering-like situations.

⁶These three tournament formats are the benchmark formats in sports. More complex formats usually include these three as building blocks (see Ryvkin 2005 for details).

Contests, where all players perform simultaneously only once, and the player with the best performance is the winner.

Binary elimination tournaments, where players are matched pairwise, and the losers of each pair are eliminated, while the winners are again matched pairwise, and so on, until only one player, the winner, remains.

Round-robin tournaments, where players are matched pairwise in all possible matchings, the winner of every match gets one point, and the player with the highest total score is the winner.

2.2 Players

We assume that there are N players who are participants in the tournament. The players are characterized by abilities $\mathbf{x} = (x_1, \dots, x_N)$ taken from a known distribution with a probability density function (pdf) $f(\cdot)$. For our illustrative calculations and computational simulations, we have chosen three prominent distributions of players' abilities: the uniform, the normal, and the Pareto distribution. The uniform and normal distributions are useful, and frequently used, benchmarks and need no further justification as such. Empirical evidence (e.g., Reed 2001; see also Hertwig *et al.* 1999, or Harrison 2005) suggests in addition that the Pareto distribution is a widespread and pervasive phenomenon.

We further assume that the incentives are set in such a way that the players do not strategically choose their efforts but always perform at the highest possible effort level, proxied here by ability. We abstract from players' effort choice problem in order to isolate the influence of tournament formats, noise, and distributions of players' abilities from the effect of the prize. This is a simplification that we believe to be inconsequential: sufficiently high prizes and players' non-zero probability of winning a tournament are likely to extract maximal effort. For example, at high-level sports tournaments, such as the Olympics, the value of winning even a single match or stage, let alone the first prize,

is so high that the assumption of maximal effort seems plausible. For the case of firms one can think of final rounds of hiring or promotion tournaments, where the remaining candidates all have already invested a significant effort to reach the current stage, and hence are likely to do their best.

2.3 Noise

We introduce uncertainty into the model by assuming that at every stage s player i 's output is given by

$$y_{is} = x_i + \epsilon_{is}, \quad i = 1, \dots, N; s = 1, \dots, S.$$

Here ϵ_{is} are zero-mean idiosyncratic random shocks, i.i.d. across players and across stages; S is the number of stages. A similar specification was used, for example, by Lazear and Rosen (1981), Nalebuff and Stiglitz (1983), Green and Stokey (1983). We denote the pdf and the cumulative density function (cdf) of ϵ_{is} as $\phi(\cdot)$ and $\Phi(\cdot)$, respectively. For illustrative calculations and computational simulations we use the zero-mean normal distribution with a variance σ^2 , which determines the noise intensity. Thus, the level of noise is characterized by one parameter. We choose the normal distribution because of its theoretically and empirically established property of being the asymptotic distribution of noise in large samples under fairly general assumptions. The rationalization stems, essentially, from the central limit theorem.

2.4 Costs

We associate two types of costs with a given tournament format: (i) time costs, and (ii) measurement costs. Time costs E_t , introduced here similarly to Gradstein (1998) and Bolton and Dewatripont (1994), are proportional to the number of stages S , $E_t = c_t S$, where c_t denotes the unit of time costs. Contests always have one stage, therefore their time costs are independent of the number of players N unless $N = 1$, $E_t = c_t(1 - \delta_{N1})$.

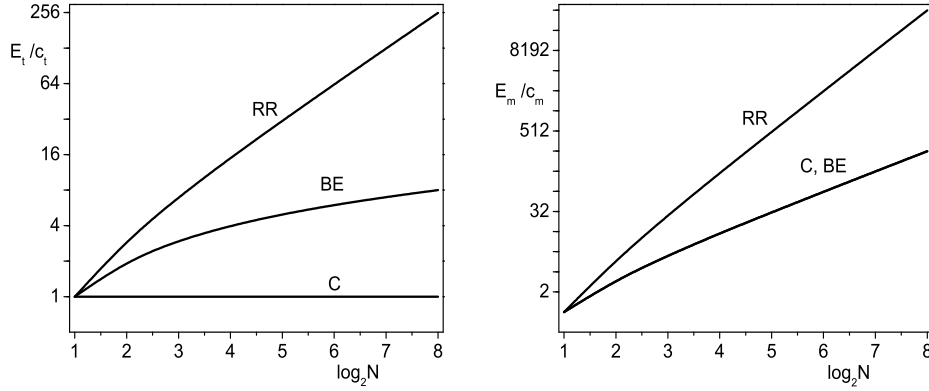


Figure 1: Time (left) and measurement (right) costs of contests (C), binary elimination (BE) tournaments and round-robin (RR) tournaments as functions of the number of players N (logarithmic base 2 scale).

Binary elimination tournaments of $N = 2^R$ players have $S = R$ stages, i.e. $E_t = c_t R = c_t \log_2 N$. Round-robin tournaments of N players are conducted in N or $N - 1$ stages if N is, respectively, even or odd. In what follows we consider only even values of N ,⁷ therefore $E_t = c_t(N - 1)$.

Measurement costs E_m are related to the number of elementary measurements (binary comparisons) needed to determine the best player. In a contest one needs $N - 1$ such comparisons, $E_m = c_m(N - 1)$. In a binary elimination tournament, it is the same, $E_m = c_m(N - 1)$. Finally, in a round-robin tournament $E_m = c_m N(N - 1)/2$. Here c_m is the unit of measurement costs. Figure 1 illustrates the dependence of the time and measurement costs on the number of players N for the three tournament formats.

2.5 Organizer’s decision problem

We assume that the tournament organizer’s objective is to select the best player. Let ρ denote the probability that the *ex ante* best player wins the tournament. We will call ρ

⁷Although contests and round-robin tournaments permit an arbitrary number of players, and the theory presented below for these two formats has no limitations on N , the binary elimination format, at least in its “pure” form, only admits a number of participants that is a positive integer power of 2. Therefore, for further comparisons across the three formats we restrict the analysis to the case of $N = 2^R$, where R is a positive integer. In a more general setting, any initial number of participants can be reduced to $N = 2^R$ through qualifying rounds (see Ryvkin 2005).

the predictive power of the tournament (for a more formal definition, see below). The higher the predictive power, the more certain the organizer can be that the tournament will reveal the best player as the winner.

We further assume that the organizer gets a payoff of V if the best player is selected, and zero payoff otherwise. This is a serious restriction, because the organizer could also benefit (although less) from selecting players ranked 2 and lower.⁸ There are, however, situations when this step-like profit function is a useful approximation of reality. In general, this happens when the best player is significantly better than all other players, which is a manifestation of the distribution of players' abilities having a long tail (such as the Pareto distribution). Other examples are situations in which finishing first is crucial, such as innovation races or formation of a national Olympic team.

Thus, we write the organizer's expected payoff as

$$\pi = V\rho - E_t - E_m, \tag{1}$$

where E_t and E_m are the time and measurement costs introduced above. Without loss of generality we set $V = 1$. The units of the organizational costs E_t and E_m , therefore, will be the same as the units of probability ρ .

3 Predictive power

In this Section we calculate the predictive power of contests, binary elimination tournaments and round-robin tournaments of N players. For contests the resulting expression is tractable by numerical integration software. For binary elimination and round-robin

⁸The organizer might also benefit from tournament formats *not* selecting the best players if it increases the "entertainment value" of the tournament for spectators. Such objectives seem of some importance in sports. For example, in the Four Hills ski jump competition, an annual winter event taking place in Austria and Germany, the best 30 jumpers out of 50 are selected in the following manner: 50 jumpers are split randomly into 25 pairs, and the winner of each pair advances. Additionally, 5 out of 25 losers with the best results advance. This modification of the rules (as compared to the previous grand contest format, where simply jumpers with the 30 best results advanced) led to the "spectacular" elimination of the 4 top ranked jumpers in 2005.

tournaments, we will use Monte Carlo simulations to evaluate the resulting expressions for large N . The simulations agree with the analytical results for small N .

Conditional predictive power. We define the conditional predictive power, $\rho(\mathbf{x})$, as the probability for the best player to be the winner of the corresponding tournament given all players' abilities $\mathbf{x} = (x_1, \dots, x_N)$.

Unconditional predictive power. We define the unconditional predictive power, or simply predictive power, as the conditional predictive power averaged over a given distribution $f(\cdot)$ of players' abilities in the population. In the averaging procedure it is still assumed that player 1 has the highest ability, $x_1 > x_k, k = 2, \dots, N$. Such averaging of an arbitrary function $A(\mathbf{x})$ takes the form

$$\langle A \rangle_f = N \int_{-\infty}^{\infty} dx_1 f(x_1) \int_{-\infty}^{x_1} dx_2 f(x_2) \dots \int_{-\infty}^{x_1} dx_N f(x_N) A(\mathbf{x}). \quad (2)$$

In the remainder of this Section, the unconditional predictive power for contests, binary elimination tournaments, and round-robin tournaments is given by three theorems.

3.1 Contests

In a contest all performance levels y_i of participating players are compared, and the player with the largest y_i is the winner.

Theorem 3.1. The predictive power of a contest is

$$\rho_c = N \int_{-\infty}^{\infty} dz_1 f(z_1) \int_{-\infty}^{\infty} dz \phi(z - z_1) \left[\int_{-\infty}^{z_1} dz_2 f(z_2) \Phi(z - z_2) \right]^{N-1}. \quad (3)$$

Proof. For a given vector of abilities \mathbf{x} , the joint pdf of players' performance levels $y_i, i = 1, \dots, N$, is $\prod_{i=1}^N \phi(y - x_i)$. Introduce new variables $y'_1 = y_1, y'_2 = y_1 - y_2, \dots, y'_N = y_1 - y_N$. The inverse transformation is given by $y_1 = y'_1, y_2 = y'_1 - y'_2, \dots, y_N = y'_1 - y'_N$

and has a unitary Jacobian, therefore the conditional predictive power is

$$\begin{aligned}\rho_c(\mathbf{x}) &= \Pr\{y'_2 > 0, \dots, y'_N > 0 | \mathbf{x}\} = \int_{-\infty}^{\infty} dz_1 \phi(z_1 - x_1) \prod_{i=2}^N \int_0^{\infty} dz_i \phi(z_1 - z_i - x_i), \\ &= \int_{-\infty}^{\infty} dz \phi(z - x_1) \prod_{i=2}^N \Phi(z - x_i).\end{aligned}\quad (4)$$

After averaging $\rho_c(\mathbf{x})$ according to Eq. (2), we obtain Eq. (3). *Q.E.D.*

3.2 Binary elimination tournaments

In a binary elimination tournament, $N = 2^R$ players compete sequentially in R stages. At each stage, the players are matched pairwise, and the winners are promoted to the next stage. This tournament format can be conveniently represented in the form of a binary tree. Figure 2 shows the tree representation for an elimination tournament of 8 players. Consider a binary tree with $N = 2^R$ terminal nodes. Enumerate the terminal nodes by R -bit binary numbers $b^{(k)} = \langle b_1^{(k)} \dots b_R^{(k)} \rangle$, where $b_r^{(i)} \in \mathcal{B} = \{0, 1\}$ are digits, $i = 1, \dots, N$, $r = 1, \dots, R$.

For convenience, we enumerate stages backwards in time.⁹ At stage R , matches $(b^{(1)}, b^{(2)}), (b^{(3)}, b^{(4)}), \dots, (b^{(N-1)}, b^{(N)})$ are played. Note that within every match the players' binary numbers differ only in the R -th bit (cf. Fig. 2).

Without loss of generality, we assume that at the beginning of the tournament, at stage R , player i is assigned to terminal node $b^{(i)}$, $i = 1, \dots, N$.¹⁰ The winners of stage R 's matches are promoted to stage $R - 1$, where they are matched according to the flow along the branches of the tree towards the root. Formally, the matching occurs as follows: two players at stage $R - 1$ are matched if and only if the first $R - 2$ bits of their binary numbers coincide, and the $(R - 1)$ -th bits are different (see Fig. 2).

⁹It is common in sports to call the stages close to the root of the tree by the number of matches ("brackets") in them, e.g. the final ($r = 1$, $2^{1-1} = 1$ match), semifinals ($r = 2$, $2^{2-1} = 2$ matches), quarterfinals ($r = 3$, $2^{3-1} = 4$ matches), etc.

¹⁰We assume here that the organizer does not have any *a priori* knowledge of players' abilities, and ignore the possibility of *seeding*, which plays an important role in tournament design (see e.g. Groh *et al.* 2004, and references therein) and, of course, affects the predictive power.

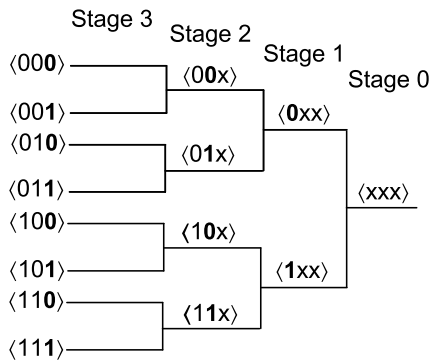


Figure 2: The binary tree representation for an elimination tournament of $N = 8 = 2^3$ players and $R = 3$ stages. The players are enumerated by binary numbers. Symbol “x” denotes unknown bits of advancing players. In each match the key differing bits are shown in bold.

The procedure is repeated recursively. At stage $R - r$ the winners from stage $R - r + 1$ are matched. Within each match the first $R - r - 1$ bits of players’ binary numbers coincide, and the $(R - r)$ -th bits are different. The tail bits are not important for matching, since $R - r$ bits suffice to enumerate the remaining players at stage $(R - r)$. Those bits, however, have to be dragged behind in order to preserve the initial “identities” of all surviving players. At stage 1 there is just one (final) match. The players of this match necessarily have different first bits. The winner of the final match (the only participant of “stage 0”) is the winner of the whole tournament.

Suppose the players’ abilities \mathbf{x} are given. Then every match of players i, j is a contest of two players, and the probability for player i to win is given by Eq. (4) with $N = 2$:

$$w(x_i, x_j) = \int_{-\infty}^{\infty} dz \phi(z - x_i) \Phi(z - x_j). \quad (5)$$

Let $P_r(b^{(i)}|\mathbf{x})$ denote the probability for the player seeded as $b^{(i)}$ to advance to stage r . At stage r player $b^{(i)}$ can only be matched with a player whose binary number has the following properties: (i) the first $r - 1$ bits are the same as in $b^{(i)}$; (ii) the r -th bit is different from that of $b^{(i)}$. Thus, the players who can potentially be matched with player

$b^{(i)}$ at stage r have a binary number of the form $\langle b_1^{(i)} \dots b_{r-1}^{(i)} \overline{b_r^{(i)}} c_{r+1} \dots c_R \rangle$. Here the overline denotes a flipped bit: $\overline{0} = 1$ and $\overline{1} = 0$; $c_s \in \mathcal{B}$ are arbitrary bits. Also, the player to be matched with $b^{(i)}$ needs herself to survive until stage r . This reasoning leads to the following result.

Theorem 3.2. The predictive power of a binary elimination tournament is $\rho_e = \langle P_0(b^{(1)}|\mathbf{x}) \rangle_f$, where $P_0(b^{(1)}|\mathbf{x})$ is a solution of the recurrence relation

$$\frac{P_{r-1}(b^{(i)}|\mathbf{x})}{P_r(b^{(i)}|\mathbf{x})} = \sum_{c_{r+1}, \dots, c_R \in \mathcal{B}} P_r(\langle b_1^{(i)} \dots b_{r-1}^{(i)} \overline{b_r^{(i)}} c_{r+1} \dots c_R \rangle|\mathbf{x}) w \left(x_{b^{(i)}}, x_{\langle b_1^{(i)} \dots b_{r-1}^{(i)} \overline{b_r^{(i)}} c_{r+1} \dots c_R \rangle} \right) \quad (6)$$

with $i = 1$ and the initial condition $P_R(b^{(i)}|\mathbf{x}) = 1$, $i = 1, \dots, N$. Here $w(\cdot, \cdot)$ is given by Eq. (5).

Proof. Recurrence relation (6) is very intuitive. The probability of advancing to the next stage is the sum of probabilities of beating all possible candidates for the match multiplied by their corresponding probabilities of reaching this point in the game. The initial conditions simply state that all N players with probability 1 start the tournament at stage R . The $R \times N$ numbers $P_r(b^{(i)}|\mathbf{x})$ completely define the probabilistic properties of the tournament conditional on abilities \mathbf{x} . The predictive power is then calculated by averaging [cf. Eq. (2)]. *Q.E.D.*

3.3 Round-robin tournaments

In a round-robin tournament, players are matched in all possible $M = N(N - 1)/2$ pairwise matchings (i, j) , $1 \leq i < j \leq N$. Every match (i, j) has one of the two allowed outcomes: either $i \rightarrow j$ (player i defeated player j) or $j \rightarrow i$ (player j defeated player i).¹¹

¹¹We ignore the possibility of ties. We do so because the literature (e.g., Rubinstein 1980) does not provide us with a consistent point counting scheme. Introducing ties would have required us to build a different model. We believe that such a model would not materially affect the results that we are interested in.

We introduce a variable p_{ij} such that

$$p_{ij} = 1 \quad \text{if } i \rightarrow j, \quad p_{ij} = 0 \quad \text{otherwise} \quad (7)$$

Since the ordering of matches is not important (all matches are assumed to be statistically independent) we will, for convenience, adopt the following (lexicographic) ordering $m = 1, \dots, M$:

$$\begin{array}{cccccccc} (i, j) : & (1, 2) & \dots & (1, N) & (2, 3) & \dots & (2, N) & \dots & (N-1, N) \\ & \downarrow & & \downarrow & \downarrow & & \downarrow & & \downarrow \\ m(i, j) : & 1 & \dots & N-1 & N & \dots & 2N-3 & \dots & M, \end{array}$$

which can be described by

$$m(i, j) = N(i-1) - \frac{i(i-1)}{2} + j. \quad (8)$$

Any outcome of the tournament can then be represented by an M -bit binary number $b = \langle b_1 \dots b_M \rangle$, where $b_{m(i,j)} = p_{ij}$. There are 2^M possible tournament outcomes, ranging from $\langle 00 \dots 0 \rangle$ to $\langle 11 \dots 1 \rangle$.

Any result of the tournament is an N -dimensional vector of scores $\mathbf{s} = (s_1, \dots, s_N)$ where every player's score is the number of wins she has, i.e. players start with a score of 0 and then add 1 point for each win. Given an outcome b , the score vector is $\mathbf{s} = \mathbf{S}(b)$, with

$$S_i(b) = \sum_{j=1}^{i-1} (1 - b_{m(j,i)}) + \sum_{j=i+1}^N b_{m(i,j)}, \quad i = 1, \dots, N. \quad (9)$$

The winner of the tournament is the player with a maximal score. There may be more than one such player, and then an additional rule has to be applied to determine the best among them. In the present paper, we assume that the winner is chosen randomly among the players with maximal scores.

For a fixed vector of abilities \mathbf{x} , player i wins a match against player j with probability $w(x_i, x_j)$ given by Eq. (5). Then every p_{ij} becomes a Bernoulli random variable with probability of success $w(x_i, x_j)$ and a tournament outcome b becomes a multivariate Bernoulli vector with independent components. The probability of outcome b is

$$\alpha(b|\mathbf{x}) = \prod_{i=1}^{N-1} \prod_{j=i+1}^N w(x_i, x_j)^{b_{m(i,j)}} [1 - w(x_i, x_j)]^{1-b_{m(i,j)}}. \quad (10)$$

Also, the score vector \mathbf{s} becomes a random vector. The distribution of \mathbf{s} is non-trivial since its components are not independent. For example, $\sum_{i=1}^N s_i = M$. The pdf of \mathbf{s} can be written as

$$\pi(\mathbf{s}|\mathbf{x}) = \sum_{b \in \{0,1\}^M} \alpha(b|\mathbf{x}) \delta[\mathbf{s} - \mathbf{S}(b)]. \quad (11)$$

Here the δ -function¹² is M -dimensional; summation goes over all M -bit binary numbers; $\mathbf{S}(b)$ is determined by Eq. (9); $\alpha(b|\mathbf{x})$ is given by Eq. (10). We are now in a position to calculate the predictive power.

Theorem 3.3. The predictive power of a round-robin tournament is $\rho_r = \langle \rho_r(\mathbf{x}) \rangle_f$, where the conditional predictive power is given by

$$\rho_r(\mathbf{x}) = \sum_{b \in \{0,1\}^M} \frac{\alpha(b|\mathbf{x}) \prod_{i=2}^N H[S_1(b) - S_i(b)]}{1 + \sum_{k=2}^N \delta_{S_1(b)S_k(b)}}. \quad (12)$$

Here $H(z)$ is the step function defined as 1 for $z \geq 0$ and 0 for $z < 0$; δ_{ij} is the Kronecker δ -symbol defined as $\delta_{ij} = 1$ if $i = j$ and 0 otherwise; $\alpha(b|\mathbf{x})$ is given by Eq. (10); $\mathbf{S}(b)$ is given by Eq. (9).

Proof. The conditional predictive power is defined as the probability that player 1 has

¹²The one-dimensional Dirac delta-function $\delta(x)$ can be defined in many equivalent ways, one of which, $\delta(x) = \lim_{\epsilon \rightarrow 0} (2\pi\epsilon^2)^{-1/2} \exp[-x^2/(2\epsilon^2)]$, is the zero-variance limit of the normal probability density. The delta-function $\delta(x)$ can be visualized as a very high but narrow peak of area 1 centered at $x = 0$. Its key property is that for any continuous function $g(x)$ the integral $\int_u^v dx \delta(x-x_0)g(x) = g(x_0)$ when $u < x_0 < v$ and 0 otherwise, i.e. the delta-function "cuts" a particular point out of any continuous function, with which it is convoluted. The M -dimensional delta-function is a product of M one-dimensional delta-functions.

a maximal score. Additionally, if several players have a maximal score, the probability should be divided by the number of such players, because the winner is chosen randomly among them. In Eq. (12) we sum over all mutually exclusive tournament outcomes. For each tournament outcome, we divide the probability that player 1's score is *greater than or equal to* all other players' scores by the number of players who have the same score as player 1. The former probability, similarly to Eq. (4), is given by

$$\Pr\{s_1 \geq s_2, \dots, s_1 \geq s_N | b, \mathbf{x}\} = \int_{-\infty}^{\infty} dz_1 \int_0^{\infty} dz_2 \dots \int_0^{\infty} dz_N \alpha(b|\mathbf{x}) \boldsymbol{\delta}[\mathbf{z} - \mathbf{S}(b)],$$

which yields, after integration, the numerator in Eq. (12). The number of players with the same score as player 1 is in the denominator of Eq. (12). *Q.E.D.*

4 Results for predictive power

We use the results of Section 3 for illustrative calculations and computational simulations of the predictive power of contests, binary elimination tournaments and round-robin tournaments. For comparisons across the three formats we have chosen the number of players to take on the values $N = 2, 4, 8, 16, 32, 64, 128, 256$, i.e., powers of 2 ranging from 1 to 8. These values of N are admissible for all three formats. The restriction to powers of 2, obviously, comes from the binary elimination format. The upper bound of $N = 256$ is set by computational resources. We also believe that, for practical purposes, it covers the most interesting situations.

We have explored the normal, Pareto, and uniform distributions of players' abilities. Parameters of the distributions have been selected so that all three distributions have unit variance, $\text{Var}(x) = 1$:

$$\begin{aligned} \text{Normal} \quad & f(x) = (2\pi\Sigma^2)^{-1/2} \exp[-x^2/(2\Sigma^2)], \quad \Sigma = 1, \\ \text{Pareto} \quad & f(x) = px^{-(p+1)} I_{[1,\infty]}(x), \quad p \approx 2.8393, \end{aligned} \tag{13}$$

$$\text{Uniform} \quad f(x) = I_{[0,a]}(x), \quad a = 2\sqrt{3}.$$

Here $I_{[a,b]}(x)$ denotes the indicator function equal to 1 within the specified interval and 0 outside of it.

Given our choice of distribution parameters, it makes sense to distinguish three noise regimes: (i) small noise, where noise intensity is much smaller than the variance in players' ability, $\sigma \ll 1$; (ii) intermediate noise, where noise intensity is comparable to the variance of players' ability, $\sigma \sim 1$; (iii) large noise, where noise intensity is much larger than the variance of players' ability, $\sigma \gg 1$. To cover all the regimes, we let σ vary with step size 0.1 from 0.1 to 2.0. In addition, we considered values of $\sigma = 3, 4, 5, 10$.

For contests, we calculated the predictive power using Eq. (3). For the binary elimination format, we used the recurrence relation (6) to find the conditional predictive power for given players' abilities \mathbf{x} , and averaged over the realizations of \mathbf{x} using simulations. For the round-robin format, we used Eq. (12) with the subsequent simulation-based averaging for $N = 2, 4, 8$, and relied on the simulations entirely (i.e. also simulated the individual matches) for larger N . Here simulations become very efficient because Eq. (12) has exponentially many terms in the sum, but most of the terms are negligible.

Figures 3, 4 and 5 show the predictive power of, respectively, contests, binary elimination tournaments and round-robin tournaments as a function of the number of players N and noise level σ for the three distributions of players' abilities. The discussion of our results in the remainder of this Section draws on these figures. As seen from the figures, the behavior of the predictive power of the three tournament formats is, with some notable exceptions, qualitatively very similar. This leads us to conjecture that there are certain "universal" features of predictive power. We discuss these common features first, and then proceed with a discussion of format-specific and distribution-specific features the predictive power displays.

4.1 Universal properties of the predictive power

4.1.1 Small- N behavior

The predictive power for one player ($N = 1$) always equals 1. Therefore, across all tournament formats and ability distributions, predictive power should decrease initially when additional competitors join the tournament except for the trivial case of no noise. This is indeed what we see. For larger N , the behavior might become non-monotonic; if and when non-monotonicities occur depends on the details. For example, for the Pareto distribution and small noise, the predictive power starts rising with N immediately after $N = 2$.

4.1.2 Small-noise behavior

For all tournament formats and ability distributions, predictive power behaves qualitatively the same in the small-noise regime. Small noise adds a small probability of upsets (i.e. situations in which a stronger player loses to a weaker player). No discontinuity in predictive power can be expected when going from zero to non-zero noise: since predictive power equals its maximal possible value of 1 for zero noise, it must necessarily decline for small noise. Clearly, this result is universal, i.e. it holds for arbitrary numbers of players $N \geq 2$ and all distributions of players' abilities $f(x)$. How fast and how far predictive power will decline depends on N and the specific shape of $f(x)$. Recall that noise is defined as small relative to the fixed variance of the ability distribution. An alternative strategy would be to fix the noise level: the small-noise regime then corresponds to large variance in ability relative to noise, i.e. the players' abilities are very different and therefore easily distinguishable, hence the predictive power is close to 1.

4.1.3 Large-noise behavior

Across all tournament formats and ability distributions, predictive power behaves also qualitatively the same for the large-noise regime. For any given N the predictive power

converges from above to $1/N$ as $\sigma \rightarrow \infty$. This is intuitive: as noise increases, ability becomes less important in determining the outcome of the tournament while chance becomes increasingly important. As in the small-noise regime, we could fix the noise level and interpret the large-noise regime as the one in which variance in ability is small, i.e. the players are hard to distinguish and hence in fierce competition. As a result, predictive power tends to its “indifference” limit $1/N$.

4.1.4 Behavior for uniformly distributed abilities

The predictive power of all three tournament formats behaves similarly for the case of the uniform distribution of players’ abilities. As seen from the third rows of Figures 3, 4 and 5, predictive power is a monotonically decreasing function of both N and σ . Indeed, for a fixed σ with increasing N the “density” of players’ abilities increases uniformly everywhere, and competition between the top-ranked players becomes tighter, as it becomes more likely that players appear whose abilities are close to that of the best player. Also, for a fixed N with increasing σ the probability for upsets increases uniformly for all players (except, of course, the top-ranked player). This result also implies, however, that the probability of outperforming a lower-ranked player decreases (except, of course, for the lowest-ranked player), possibly neutralizing the net effect. This neutralization will be complete only for the median player, and will happen partially for other “interior” players (i.e., players other than the top-ranked and lowest-ranked players). In fact, interior players to the right of the median player will experience a negative effect because relative to the median player there is more mass of the distribution to their left (and hence more chances of being upset, and fewer chances of upsetting). Likewise, interior players to the left of the median player will experience a positive effect because relative to the median player there is less mass of the distribution to their left (and hence fewer chances of being upset, and more chances of upsetting).

4.2 Non-universal properties of predictive power

Maybe unsurprisingly, the small-noise and large-noise properties of predictive power are universal. These extreme regimes, however, are not of much interest from the institutional design point of view because there is not much a designer can do to influence the situation. Indeed, when noise is small (or, in other words, when players have very diverse abilities), design efforts are not necessary because the *ex ante* top-ranked player will almost surely win the tournament anyway. On the other hand, when noise is large (when players are almost identical in ability), there is not much a designer can do against pure chance, and there is not much a designer would be *willing* to do because she is nearly indifferent about who the winner is. Of most interest is the regime of intermediate noise, when the players' abilities vary on the same scale as the noise level. In this regime, tournament design can substantially affect predictive power and hence the organizer's payoff.

Predictive power has very intuitive and universal behavior for the uniform distribution of abilities because the players are affected uniformly by an increase in N and σ . In most real-life cases, however, the distribution of abilities is non-uniform. We explored two empirically prominent non-uniform distributions, the normal and Pareto ones. In both cases, players are differentially affected by an increase in N and σ , leading to non-trivial and counter-intuitive behavior of the predictive power, such as non-monotonicity.

4.2.1 Non-monotonicity of ρ as a function of N

As seen from Figures 3, 4 and 5, predictive power can be a non-monotonic function of the number of players for a given non-zero noise level σ . As illustrated by the relevant right panels, contests and binary elimination tournaments display non-monotonicity in N for the Pareto distribution of abilities, while round-robin tournaments also display it for the normal distribution. Thus, one question is where non-monotonicity comes from at all, while another question is why the behavior for the normal and Pareto distributions can be different for different formats.

The decaying upper tail of the Pareto distribution implies that relative to the low-ability players the average “distance” between the top-ranked players increases in N , which affects the predictive power positively. The countervailing effect comes from an increase in the probability of upsets with N . Which of the two effects prevails depends on N and σ . For small noise, the positive effect should prevail already for small N ; for larger noise, the negative effect prevails for smaller N while for larger N the positive effect eventually wins. As a result, for larger noise there appears a minimum in predictive power as a function of N . The value of N at the minimum corresponds to the turning point where noise becomes small relative to the average distance between the top-ranked players.

For the normal distribution, the intuition stemming from the Pareto distribution seems a useful point of departure. Note that, ignoring the difference in tails of the two distributions, the Pareto distribution is in a sense the upper half of the normal distribution. Hence, we should expect qualitatively somewhat similar behavior. This, however, is not the case for contests and binary elimination tournaments.

Let us first discuss why the predictive power of contests displays no non-monotonicity for the normal distribution, while it does for the Pareto distribution. The Pareto distribution has a long tail, meaning that the probability of finding a value significantly larger than the mean is very high compared to an exponentially decaying distribution with the same variance. For example, for the distributions with unit variance that we use, $\Pr\{x - E(x) > 3\}$ (i.e. the probability of exceeding the mean by more than 3 standard deviations) is 0.00135 for the normal distribution and 0.0136 – nearly ten times larger – for the Pareto distribution. As a result, when the number of players increases, the top-ranked player’s ability also increases, on average, but much slower for the normal distribution than for the Pareto distribution. Therefore the concurrent increase in the probability of upsets from less able players dominates for all values of N for the normal distribution, as opposed to the Pareto distribution where it only dominates for smaller

N .

The difference between the normal and Pareto distributions is also observed for binary elimination tournaments but not observed for round-robin tournaments. The round-robin tournament format is very different from the other two formats in that it allows for losses. Indeed, winning a contest or a binary elimination tournament of N players is similar, in a very simplified picture that ignores inter-player correlations, to getting a maximal score (i.e. winning all matches) in a round-robin tournament. We know, however, that a player can sometimes win a round-robin tournament with a smaller score. Accommodation of losses is the feature of the round-robin format that makes it qualitatively different from the other formats where only “flawless” performance leads to a win. Besides, in a round-robin tournament players are matched a comparatively large number of times without any elimination, i.e. they have many opportunities to upset one another. Interestingly, an increase in the probability of upsets, for whatever reason, helps the top-ranked player in that the players ranked 2 and lower upset one another more intensively, on average, than they upset the top-ranked player, and lose more points as a result. This effect, in turn, allows the top-ranked player to lose more points and still remain the winner.

The above arguments are very qualitative and by no means explain in full what happens. The exact quantitative behavior of predictive power, as described in Section 3, results from a complex entanglement of combinatorial and statistical factors that are hard to completely intuit. For example, although intuitively it seems that the round-robin format is more “robust” than the binary elimination format where a single loss leads to elimination, Table 1 shows that for $N = 4$ and $\sigma = 1$ the binary elimination format predicts the best players at least as well as the round-robin format, and even better for the uniform distribution of abilities. For larger N , however, the “intuitive picture” is restored.

4.2.2 Non-monotonicity of ρ as a function of σ

This even more counter-intuitive feature can be seen in the left panels of Fig. 5. It is only observed for round-robin tournaments and for the normal and Pareto distributions of players' abilities. As seen from Fig. 5, predictive power as a function of σ undergoes a bifurcation for some critical value of $N = N_c$: for N below N_c predictive power is a monotonically decreasing function of σ , while for $N > N_c$ it acquires a local minimum and a local maximum. For the normal distribution $128 < N_c < 256$, while for the Pareto distribution $64 < N_c < 128$.

Obviously, as noise increases, so does the probability of upsets. However, as explained above, a feature of the round-robin tournament is that upsets do not necessarily lead to the overall loss of the best player, especially when N is large. This is precisely what we observe: for $N > N_c$ there is a range of σ when the predictive power *increases* in σ . In this range of σ the lower-ranked players immediately following the best player (who are most likely to upset her) compete among themselves and lose points more intensively than they upset the best player. This leads to the surprising upward swing in predictive power as a function of noise level σ for large N .

5 Optimal choice of the tournament format

Which format then would an organizer choose? For the most interesting case of intermediate levels of noise the organizer's choice of tournament format depends significantly on her objectives. We assume that the organizer maximizes her expected payoff given by Eq. (1) with $V = 1$:

$$\begin{aligned}
 \text{Contest} & \quad \pi_c = \rho_c(N, \sigma) - c_t(1 - \delta_{N1}) - c_m(N - 1), \\
 \text{Binary elimination} & \quad \pi_e = \rho_e(N, \sigma) - c_t \log_2 N - c_m(N - 1), \\
 \text{Round - robin} & \quad \pi_r = \rho_r(N, \sigma) - c_t(N - 1) - c_m N(N - 1)/2.
 \end{aligned} \tag{14}$$

N	contest					binary elimination					round-robin				
	$\frac{E_t}{c_t}$	$\frac{E_m}{c_m}$	$\rho_c^{(n)}$	$\rho_c^{(p)}$	$\rho_c^{(u)}$	$\frac{E_t}{c_t}$	$\frac{E_m}{c_m}$	$\rho_e^{(n)}$	$\rho_e^{(p)}$	$\rho_e^{(u)}$	$\frac{E_t}{c_t}$	$\frac{E_m}{c_m}$	$\rho_r^{(n)}$	$\rho_r^{(p)}$	$\rho_r^{(u)}$
4	1	3	.58	.50	.58	2	3	.60	.51	.60	3	6	.60	.51	.59
8	1	7	.46	.46	.42	3	7	.50	.49	.45	7	28	.55	.56	.49
16	1	15	.37	.48	.28	4	15	.42	.54	.31	15	120	.55	.68	.40
128	1	127	.22	.69	.06	7	127	.30	.73	.08	127	8128	.68	.85	.19

Table 1: Predictive power ($\rho^{(n)}$ for the normal, $\rho^{(p)}$ for the Pareto, and $\rho^{(u)}$ for the uniform distribution of players' abilities) and costs of the three tournament formats for $N = 4, 8, 16, 128$ players and noise level $\sigma = 1$.

Consider the organizer's choice of tournament format for the case of intermediate noise level $\sigma = 1$. We let N take on the values $N = 4, 8, 16$, and 128 . Table 1 gives the values of the predictive power and costs of the three formats for these parameter values. Depending on the unit time and measurement costs c_t and c_m , the organizer will switch between the three formats. Figure 6 shows the regions in the (c_t, c_m) plane where a particular format is preferred to the other two.

As can be seen from Figure 6, the boundary between the contest and binary elimination formats for $N > 2$ has the shape of a vertical line parallel to the c_m -axis. The reason is, as seen from Eq. (14), that the two formats have the same measurement costs. The switching value of c_t between the two formats is given by the equation $\pi_c = \pi_e$, which yields

$$c_t^{c-e} = \frac{\rho_e(N, \sigma) - \rho_c(N, \sigma)}{\log_2(N/2)}. \quad (15)$$

The predictive power of binary elimination tournaments dominates the predictive power of contests. The difference between the two, which appears in the numerator of Eq. (15), is the predictive power *gain* from using the more complex (and costly) binary elimination format as compared to the basic contest. The critical value of c_t at which the organizer will switch between the two formats is proportional to this gain, and decreasing logarithmically in the number of players. For $N = 2$ there is no difference between the two formats, hence the critical value of c_t tends to infinity.

The boundary between the binary elimination and round-robin formats is determined

by the equation $\pi_e = \pi_r$, which gives

$$c_m^{e-r} = \frac{\rho_r(N, \sigma) - \rho_e(N, \sigma)}{(N-1)(N/2-1)} - \frac{N-1-\log_2 N}{(N-1)(N/2-1)} c_t. \quad (16)$$

Equation (16) describes a downward sloping linear boundary for $N > 2$ and $\rho_r > \rho_e$. Note that $\rho_r(4, 1) \leq \rho_e(4, 1)$, as seen from Table 1, therefore the round-robin format never arises as the optimal choice for $N = 4$ and $\sigma = 1$.

Contests are the least costly in terms of time and have the same measurement costs as binary elimination tournaments. Round-robin tournaments are the most costly in all respects. Unsurprisingly, the same is true, albeit with some exceptions, for predictive power, which makes the optimization meaningful. For sufficiently small marginal costs of time and measurement, the organizer always prefers the round-robin tournament as the most accurate. As the marginal costs increase, the organizer will switch to the binary elimination format. Eventually, when the marginal cost of time reaches a threshold value, the contest format becomes the optimal choice. Interestingly, for $N = 4$ players and $\sigma = 1$ the round-robin format is dominated by the binary elimination format and never arises as the optimal choice.

For other organizer's objectives, such as the maximization of the "entertainment value" of a tournament, the results could be totally different. As Terry Bradshaw¹³ put it, "I am always for the underdog," and so seem to be most sport spectators. Organizers, therefore, may think about minimizing rather than maximizing predictive power. More generally, there is evidence that some sort of "competitive balance" needs to be sustained in sports (see Szymanski 2003).

¹³Terry Paxton Bradshaw is a former American football quarterback with the Pittsburgh Steelers in the American National Football League (NFL), a current football analyst and co-host of FOX NFL Sunday (Wikipedia).

6 Conclusions

We have investigated three prominent tournament formats – contests, binary elimination tournaments and round-robin tournaments – in terms of their predictive power to select the best player as the winner. We formulated and solved the decision problem of an organizer whose objective is to choose among the three tournament formats one that reveals the best player most accurately. In our model of a noisy tournament of heterogeneous players, the predictive power, defined as the probability of the player with the highest ability to win the tournament, is a function of the number of players N , overall noise level σ , the distribution of players' abilities $f(x)$, and the tournament format.

Noise determines the probability of upsets and acts as an equalizer of players' abilities. It is clear intuitively, and supported by our calculations and computational simulations, that in the small-noise regime (when players have very diverse abilities) the predictive power approaches 1, while in the large-noise regime (when players' abilities are difficult to distinguish) it tends to the “indifference” limit $1/N$. In the intermediate-noise regime, when noise varies on the same scale as the players' abilities, the behavior of the predictive power is non-universal, and it is in the organizer's interest to extract information about players' ranking from their noisy performance. We have found that predictive power can be non-monotonic as a function of the number of players N . This counter-intuitive feature can be observed for non-uniform distributions of players' abilities. We observe it for all three tournament formats when $f(x)$ is a power law, and for round-robin tournaments also when $f(x)$ is normal. Remarkably, for round-robin tournaments predictive power can also be non-monotonic as a function of σ .

The organizer's optimal choice of tournament format is determined by balancing predictive power and organizational costs. We introduced two types of such costs, those related to time and measurements. Time costs are associated with the number of stages a tournament has, while measurement costs are proportional to the number of “elementary ranking measurements” (binary comparisons) an organizer has to implement. With

increasing marginal costs of time and measurements, the organizer will switch from the round-robin format to the binary elimination format, and eventually to the contest format. Interestingly, for some parameter values some formats never arise as the optimal choice.

The major contribution of our paper is the formulation of a unified framework that allows us to analyze and compare the predictive power of arguably the most important tournament formats and their variants (e.g. involving seeding and qualifying rounds). Important findings are the non-desirability of some tournament formats for some parameterizations and the non-monotonicity of predictive power as a function of N and σ .

Our results can conceptually inform the decision-makers of recruiting committees, research and public project funding agencies and other principals whose objective is to select the best agent. In principle, it should be possible to calibrate our model with historic or experimental data. This endeavor is part of our research agenda.

References

- Appleton D., 1995. May the best man win? *The Statistician*, 4, 529-538.
- Baye M., Hoppe H., 2003. The strategic equivalence of rent-seeking, innovation, and patent-race games. *Games and Economic Behavior*, 44, 217-226.
- Ben-Yashar R., Nitzan S., 1997. The optimal decision rule for fixed-size committees in dichotomous choice situations: the general result. *International Economic Review*, 38, 175-186.
- Bognanno M., 2001. Corporate tournaments. *Journal of Labor Economics*, 19, 290-315.
- Bolton P., Dewatripont M., 1994. The firm as a communication network. *The Quarterly Journal of Economics*, 109, 809-839.

- Cornes R., Hartley R., 2005. Asymmetric contests with general technologies. *Economic Theory*, 26, 923-946.
- Frick B., 2003. Contest theory and sport. *Oxford Review of Economic Policy*, 19, 512-529.
- Ehrenberg R., Bognanno M., 1990. (a) The incentive effects of tournaments revisited: evidence from the European PGA Tour. *Industrial and Labor Relations Review*, 43, 74S-88S. (b) Do tournaments have incentive effects? *The Journal of Political Economy*, 98, 1307-1324.
- Eriksson T., 1999. Executive compensation and tournament theory: Empirical test on Danish data. *Journal of Labor Economics*, 17, 262-280.
- Esteban J., Ray D., 2001. Social decision rules are not immune to conflict. *Economics of Governance*, 2, 59-67.
- Fullerton R., McAfee R., 1999. Auctioning entry into tournaments. *The Journal of Political Economy*, 107, 573-605.
- Gneezy U., Niederle M., Rustichini A., 2003. Performance in competitive environments: gender differences. *Quarterly Journal of Economics*, 1049-1074.
- Gradstein M., 1998. Optimal contest design: volume and timing of rent seeking in contests. *The European Journal of Political Economy*, 14, 575-585.
- Gradstein M., Konrad K., 1999. Orchestrating rent-seeking contests. *The Economic Journal*, 109, 536-545.
- Green J., Stokey N., 1983. A comparison of tournaments and contracts. *The Journal of Political Economy*, 91, 349-364.
- Groh C., Moldovanu B., Sela A., Sunde U., 2004. Optimal seedings in elimination tournaments. Working paper.

- Harary E., Moser L., 1966. The theory of round robin tournaments. *American Mathematical Monthly*, 73, 231-246.
- Harbaugh, R., Klumpp, T., 2005. Early round upsets and championship blowouts. *Economic Inquiry*, 43, 316-329.
- Harrison G., 2005. Field experiments and control, in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics*. Greenwich, CT: JAI Press, *Research in Experimental Economics*, Volume 10.
- Hertwig R., Hoffrage U., Martignon L., 1999. Quick estimation: letting the environment do the work, in Gigerenzer G. et al, *Simple heuristics that make us smart*, NY Oxford, Oxford University Press, pp. 209-234.
- Hvide H., 2002. Tournament rewards and risk taking. *Journal of Labor Economics*, 20, 877-898.
- Klumpp, T., Polborn, M., 2005. Primaries and the New-Hampshire effect. Forthcoming in *Journal of Public Economics*.
- Knoeber C., Thurman W., 1994. Testing the theory of tournaments: An empirical analysis of broiler production. *Journal of Labor Economics*, 10, 357-379.
- Knuth D., 1987. Random knockout tournament: Problem 86-2. *SIAM Review*, 29, 127-129.
- Lazear E., 1999. Personnel economics: past lessons and future directions. Presidential address to the society of labor economists, San Francisco, May 1, 1998. *Journal of Labor Economics*, 17, iv+199-236.
- Lazear E., Rosen S., 1981. Rank-order tournaments as optimal labor contracts. *The Journal of Political Economy*, 89, 841-864.

- Levin J., Nalebuff B., 1995. An introduction to vote-counting schemes. *The Journal of Economic Perspectives*, 9, 3-26.
- Lockard A., Tullock G. (Eds.), 2001. *Efficient rent-seeking: chronicle of an intellectual quagmire*. Kluwer Academic Publishers, Boston.
- McGarry T., Schutz R., 1997. Efficacy of traditional sport tournament structures. *Journal of Operational Research Society*, 48, 65-74.
- Moldovanu B., Sela A., 2001. The optimal allocation of prizes in contests. *American Economic Review*, 91, 542-558.
- Moon J., Pullman N., 1970. On generalized tournament matrices. *SIAM Review*, 12, 384-399.
- Nalebuff B., Stiglitz J., 1983. Prizes and incentives: Towards a general theory of compensation and competition. *The Bell Journal of Economics*, 14, 21-43.
- O'Flaherty B., Siow A., 1995. Up-or-out rules in the market for lawyers. *Journal of Labor Economics*, 13, 709-735.
- O'Keefe M., Viscusi W., Zeckhauser R., 1984. Economic contests: Comparative reward schemes. *Journal of Labor Economics*, 2, 27-56.
- Orrison A., Schotter A., Weigelt K., 2004. Multiperson tournaments: an experimental examination. *Management Science*, 50, 268-279.
- Prendergast C., 1999. The provision of incentives in firms. *Journal of Economic Literature*, 37, 7-63.
- Reed W., 2001. The Pareto, Zipf, and other power laws. *Economics Letters*, 74, 15-19.
- Rosen S., 1986. Prizes and incentives in elimination tournaments. *The American Economic Review*, 76, 701-715.

- Roth A., 2002. The economist as engineer: game theory, experimentation, and computation as tools for design economics. *Econometrica*, 70, 1341-1378.
- Ryvkin D., 2005. The predictive power of composite noisy tournaments. Unpublished manuscript at <http://www.msu.edu/~ryvkindm>.
- Searls D., 1963. On the probability of winning with different tournament procedures, *Journal of the American Statistical Association*, 58, 1064-1081.
- Szymanski S., 2003. The economic design of sporting contests: a Review. *Journal of Economic Literature*, 41, 1137-1187.
- Taylor C., 1995. Digging for golden carrots: an analysis of research tournaments. *The American Economic Review*, 85, 872-890.
- Tullock G., 1980. Efficient rent seeking. In *Buchanan, J., Toltison, R., Tullock, G., eds. Towards a Theory of Rent Seeking Society*, College Station: Texas A&M University Press, 97-112.

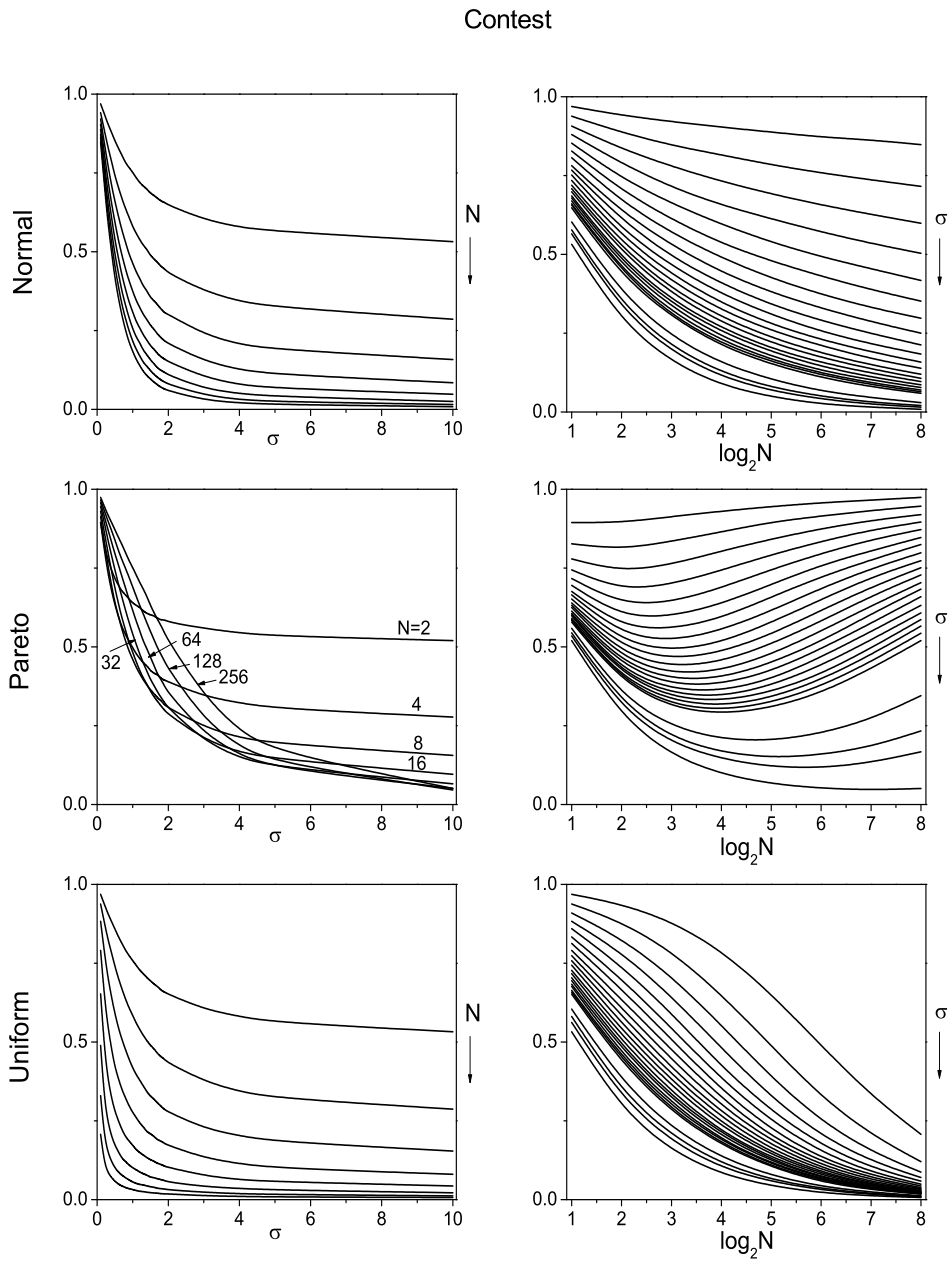


Figure 3: Predictive power ρ_c of the contest format as a function of noise level σ and number of players N .

Binary elimination

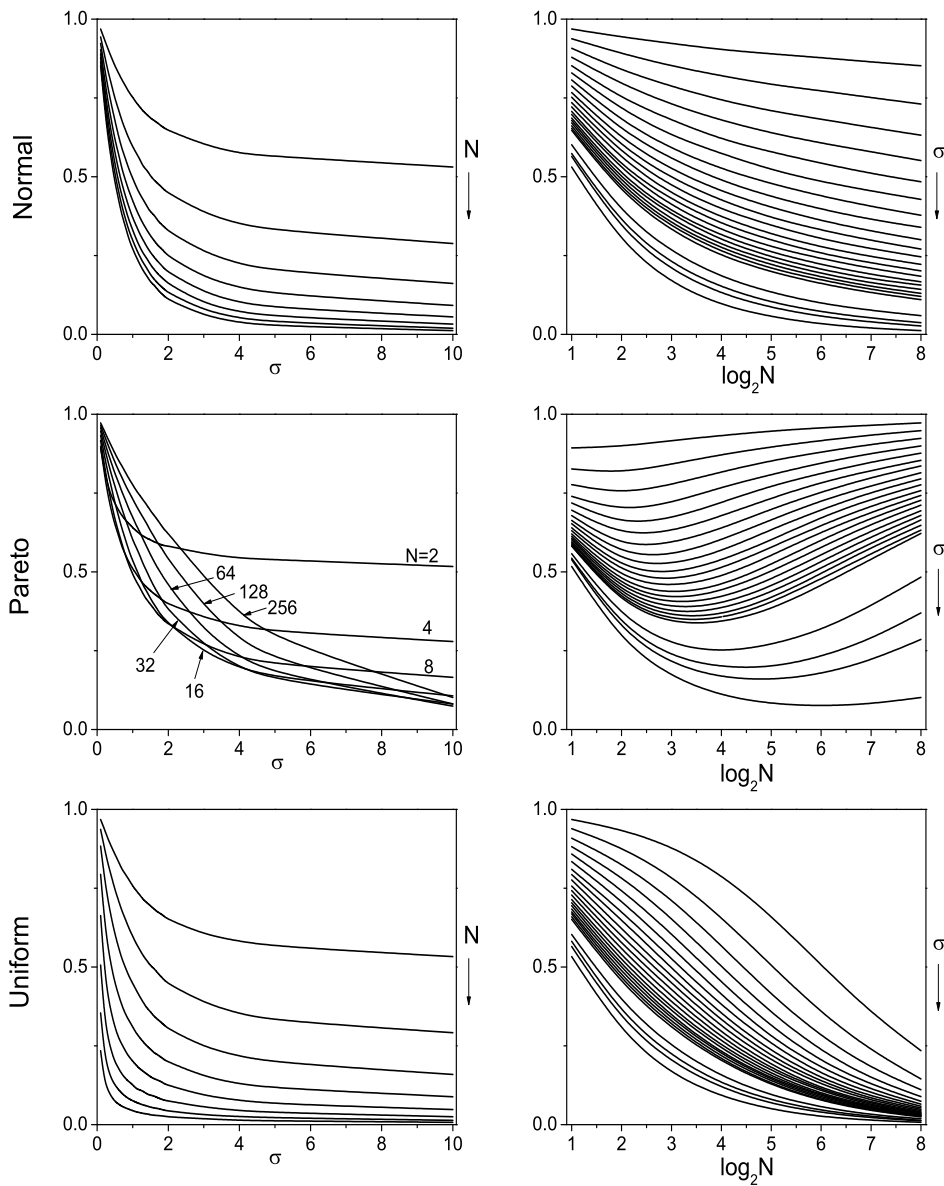


Figure 4: Predictive power ρ_e of the binary elimination format as a function of noise level σ and number of players N .

Round-robin

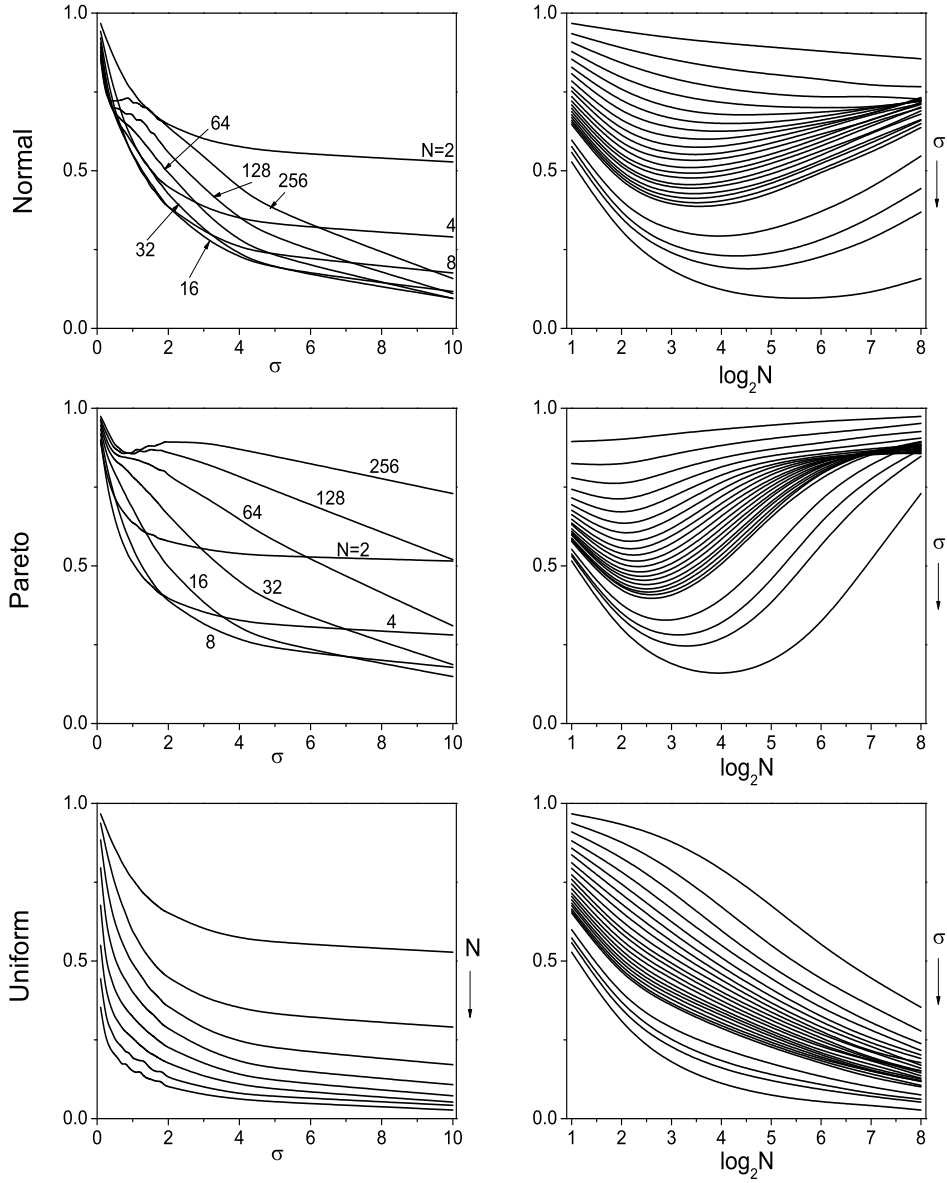


Figure 5: Predictive power ρ_r of the round-robin format as a function of noise level σ and number of players N .

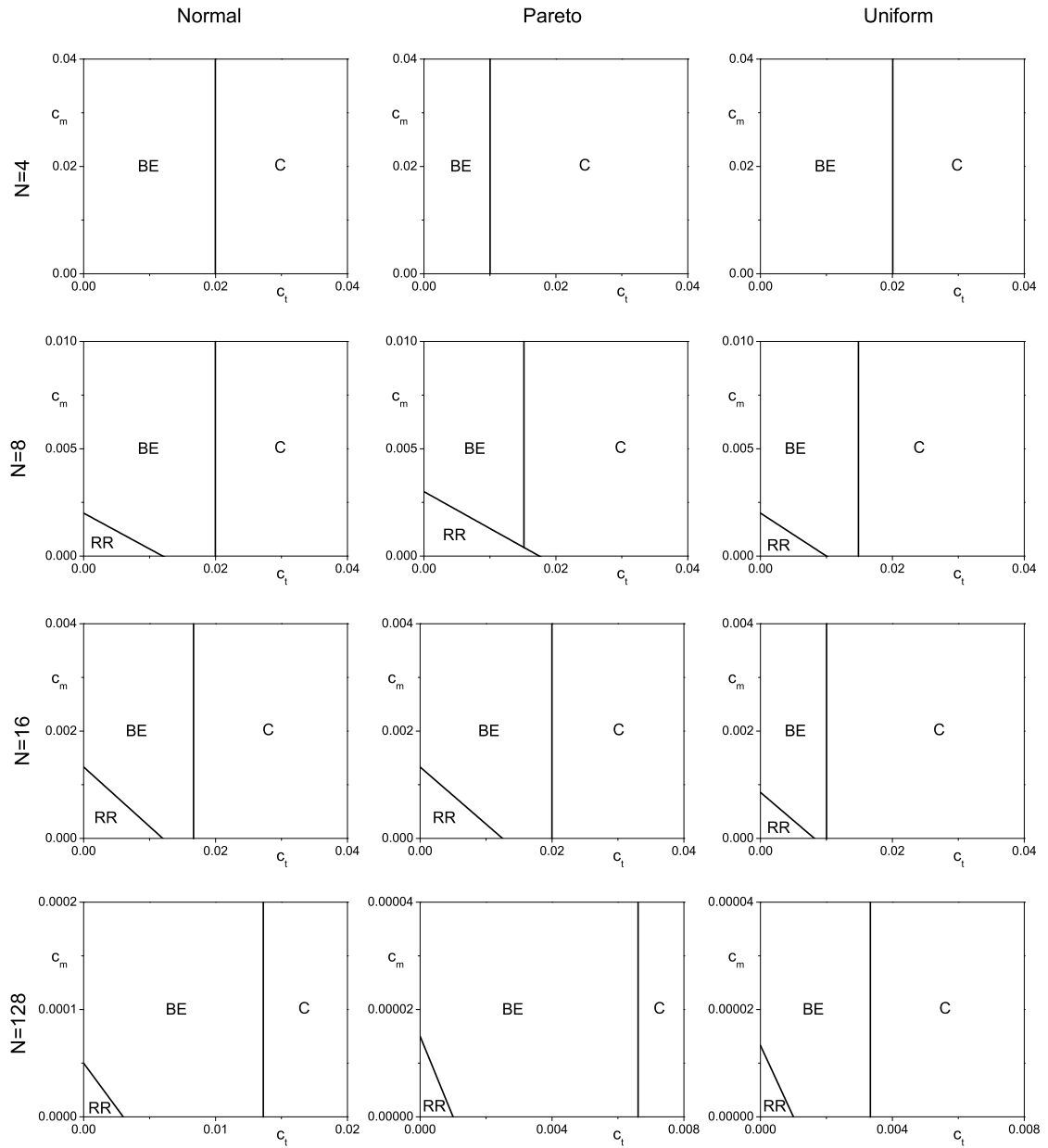


Figure 6: Regions in the (c_t, c_m) plane where each of the three tournament formats (contests – C, binary elimination tournaments – BE, round-robin tournaments – RR) is chosen. The results refer to the case of $\sigma = 1$.