The Predictive Power of Noisy Round-Robin Tournaments

Dmitry Ryvkin
Andreas Ortmann

CERGE-EI

Charles University
Center for Economic Research and Graduate Education
Academy of Sciences of the Czech Republic
Economics Institute

WORKING PAPER SERIES (ISSN 1211-3298)
Electronic Version

236

THE PREDICTIVE POWER OF NOISY ROUND-ROBIN TOURNAMENTS[1]

Dmitry Ryvkin[2] and Andreas Ortmann[3]
CERGE-EI
Charles University and Academy of Sciences of the Czech Republic
Polítických vězňů 7, 11121 Prague, Czech Republic

September 3, 2004

Abstract
The round-robin tournament format for N players is a scheme that matches players with one another in all possible $N(N-1)/2$ pairwise comparisons. A noisy round-robin tournament adds the possibility of upsets, or noise, and hence reduces the power of the tournament to reveal the true ranking of the players. In this article we study theoretically (analytically and by way of computational simulations) the predictive power of noisy round-robin tournaments for three prominent distributions of players' abilities, as a function of the level of noise and the number of players. At first sight, some of our results (e.g., non-monotonicity as a function of the number of players N, which makes some ranges of N non-optimal) are quite counterintuitive but should be of help to a tournament designer who tries to maximize, or maybe minimize, the probability of the best player winning.

Keywords: round-robin tournaments, noise, power distributions, design economics
JEL Classification: C73, C90, D21

Abstrakt
Turnajový formát "každý s každým" pro N hráčů je schéma, kde se odehraje všech $N(N-1)/2$ možných srovnání. Přítomnost šumu (která umožní i vítězství slabšího hráče nad silnějším) snižuje sílu, se kterou turnaj dokáže odhalit správný rank jednotlivých hráčů. V článku studujeme teoreticky (analyticky i výpočetními simulacemi) predikční sílu turnajů typu "každý s každým" s šumem pro tři nejvýznamnější rozdělení hráčských dovedností jako funkci úrovně šumu a počtu hráčů v turnaji. Některé naše výsledky (např. ztráta monotonicity v závislosti na počtu hráčů, která činí některé intervaly počtu hráčů neoptimální) jdou na první pohled proti běžné intuici, ale měly by pomoci tvůrcům turnajů přui jejich snaze maximalizovat (či jindy minimalizovat) pravděpodobnost vítězství toho nejlepšího.

[1]We thank Dirk Engelmann for his comments.
[2]dmitry.ryvkin@cerge-ei.cz
[3]andreas.ortmann@cerge-ei.cz or 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", round-robin tournaments, and elimination tournaments. Contests are essentially one-shot all-pay auctions whose properties have been widely discussed in the literature [e.g., theoretically by Lazear and Rosen (1981), Green and Stokey (1983), Taylor (1995), Hvide (2002); empirically by Knoeber and Thurman (1994); experimentally by Schotter and Weigelt (1992), Gneezy *et al* (2003); see also reviews by Lazear (1999) and Prendergast (1999)]. In contrast, round-robin tournaments and elimination tournaments are polar cases of schemes that compare agents pairwise and sequentially in various degrees of completeness. These schemes allow (most) agents to perform repeatedly, typically against a stream of ever changing opponents. A round-robin tournament is a complete pairwise matching scheme. An elimination tournament is an incomplete multi-stage pairwise matching scheme whose exact realization, and efficiency, depends on the initial seeding and the history of play [see Ryvkin (2004)].

Sport provides us with a simple and useful language to describe tournaments. Competitors are typically called *players*, and pairwise comparisons are called *matches*. In every match, there is a *winner* and a *loser*, or there is a tie. Below we will use this terminology.

In the presence of noise which affects the performance of players either positively or negatively, a tournament can be thought of as a probabilistic device whose output - the ranking - is a statistic of sorts of the "true" ordering of the set of players. Such an ordering identifies who is the best player ex ante.

In the present paper, we analyze theoretically (analytically and by way of computational simulations), the properties of round-robin tournaments as a function of noise level, number of players, and distribution of players' abilities. In Ryvkin (2004), employing essentially the technology laid out here, we analyze similarly elimination tournaments. In future work, we will compare the properties of these two polar matching schemes, as well as variants thereof.

In economics, round-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 ranking 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" ranking scheme in the sense that it satisfies certain natural axioms. Rubinstein (1980) also shows that it is *the only* such scheme. Below we will make use of this result.

Our paper contributes to both literatures by analyzing theoretically the probabilistic properties of a round-robin tournament and by studying these properties across what we consider the most prominent distributions of player abilities. The investigation of round-robin tournaments is an important step towards solving the problem of optimal design [see Moldovanu and Sela (2004) for a theoretical study of "contest architecture"; see more generally Roth (2002), for a discussion of how simulations with artificial agents and experiments with human subjects serve to extend simple theoretical models, which become too complex in engineering-like situations].

The present paper is organized as follows: In Section 2 we present the general setup. In Section 3 we discuss a model underlying the parametrization of winning probabilities. In Section 4 we calculate analytically the predictive power. In Section 5 we present and discuss our results. We conclude in Section 6.

2 General setup

Let $\mathcal{P} = \{1, \ldots, N\}$ be a set of N players. A *round-robin tournament* on the set $\mathcal{P}$ consists of all $M = N(N-1)/2$ possible pairwise matches (i, j), $1 \leq i < j \leq N$, of the players.

Every match (i, j) has one of the two allowed outcomes: either $i \to j$ (player i defeated player j) or $j \to i$ (player j defeated player i) [4]. We introduce a variable p_{ij} such that

$$p_{ij} = 1 \quad \text{if} \quad i \to j, \quad p_{ij} = 0 \quad \text{otherwise} \tag{1}$$

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 $m = 1, \ldots, M$:

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

which can be described by

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

The outcome of the tournament can then be represented by an M-bit binary

[4] We ignore the possibility of ties. We did 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: the effect of noise and distribution of players' abilities.

number $\mathbf{b} = \langle b_1 \dots b_M \rangle$, where $b_{m(i,j)} = p_{ij}$. There are 2^M possible tournament outcomes, ranging from $\langle 00 \dots 0 \rangle$ to $\langle 11 \dots 1 \rangle$.

The 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. Since every player plays $N-1$ matches and the total number of matches is M, for any score vector $\mathbf{s}$ the follwing constraints hold:

$$\sum\nolimits_{i=1}^{N} s_i = M, \quad 0 \leq s_i \leq N-1. \tag{3}$$

Note that properties (3) are necessary but not sufficient for a vector $\mathbf{s}$ to be a score vector. For example, not more than one component of $\mathbf{s}$ can equal 0 or $N-1$; many other constraints can be identified.

If the outcome of the tournament, $\mathbf{b}$, is known, then the score vector can be calculated directly, using the function $\mathbf{S}$, whose components are given by

$$S_i(\mathbf{b}) = \sum\nolimits_{j=1}^{i-1} (1 - b_{m(j,i)}) + \sum\nolimits_{j=i+1}^{N} b_{m(i,j)}, \quad i = 1, \dots, N. \tag{4}$$

The winners of the tournament are the players with maximal scores. There may be more than one such player, and then an additional rule (perhaps another tournament) has to be applied if one needs to determine the best among them. In the present paper, we do not adopt any additional rules and consider all players with a maximal score to be winners.

Assume now that the outcomes of the matches are random, i.e. in a match (i,j) the result is $i \to j$ with some probability w_{ij} and $j \to i$ with probability $w_{ji} = 1 - w_{ij}$. Then every p_{ij} becomes a Bernoulli random variable with the probability of success w_{ij}, and the tournament outcome $\mathbf{b}$ becomes a multivariate Bernoulli vector with independent components. The probability of outcome $\mathbf{b}$ is

$$P(\mathbf{b}) = \prod_{i=1}^{N-1} \prod_{j=i+1}^{N} w_{ij}^{b_{m(i,j)}} (1 - w_{ij})^{1-b_{m(i,j)}}. \tag{5}$$

Also, the score vector $\mathbf{s}$ becomes a random vector satisfying conditions (3) and other related constraints. The distribution of $\mathbf{s}$ is non-trivial since its components are not independent. The probability density function (pdf) of $\mathbf{s}$ can be written as

$$\pi(\mathbf{s}) = \sum_{\mathbf{b} \in \{0,1\}^M} P(\mathbf{b}) \boldsymbol{\delta}[\mathbf{s} - \mathbf{S}(\mathbf{b})]. \tag{6}$$

Here the $\boldsymbol{\delta}$-function[5] is M-dimensional; summation goes over all M-bit binary numbers; $\mathbf{S}(\mathbf{b})$ is determined by Eq. (4); $P(\mathbf{b})$ is given by Eq. (5).

We are interested in the probability ρ_1 for a specific player (player 1, for concreteness) to be among the winners of the tournament, i.e. to have a maximal score. We will calculate this probability both analytically and using computational simulations.

Winning probabilities w_{ij} can be given exogenously, through past statistics or rating data. Alternatively, we can calculate the winning probabilities using a simple model in which it is assumed that every player has a *power level* (or ability) which can be represented by a single real number. The power level is assumed to be distributed in the population of players with some known pdf $f(\cdot)$. The randomness is introduced through the assumption that in every match the power levels of players are distorted by additive noise, whose pdf $g(\cdot)$ is known. It is then possible to calculate the average winning probabilities w_{ij}, where the indices indicate the *ordering*

[5]The Dirac delta-function $\delta(x)$ can be defined in many equivalent ways, one of which, $\delta(x) = \lim_{\epsilon \to 0} (2\pi\epsilon^2)^{-1/2} \exp[-x^2/(2\epsilon^2)]$, 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 $\int_a^b dx \delta(x - x_0) g(x) = g(x_0)$ when $a < x_0 < b$ and 0 otherwise, i.e. the delta-function "cuts" a particular point out of any continuous function, with which it is integrated as a weight.

of players by their power level[6].

For actual calculations, we have chosen three prominent distributions of players' abilities, or power levels: 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 (2004)] suggests in addition that the Pareto distribution is a widespread and pervasive phenomenon. In the next Section we present the model that allows us to calculate w_{ij}.

3 Parametrization of the winning probabilities

Let $f(x)$ be the pdf of the power levels in the population of players. Suppose N players are drawn from that population and ordered by their power levels $x_1, ..., x_N$ so that $x_1 \geq x_2 \geq ... \geq x_N$.

Consider an arbitrary match (i, j). The performance levels of the players in this match will be random numbers $Y_i = x_i + \epsilon_i$ and $Y_j = x_j + \epsilon_j$, where ϵ's represent the noise, which is i.i.d. across players and across matches with pdf $g(\epsilon)$. Since (x_i, x_j) are fixed numbers at this point, Y's will be distributed with pdf's $h_i(y|x_i) = g(y-x_i)$ and $h_j(y|x_j) = g(y - x_j)$. Therefore, the probability for player i to be the winner in this match is

$$\tilde{w}(x_i, x_j) \equiv \Pr\{Y_i - Y_j \geq 0 | x_i, x_j\} = \int_0^\infty dz \int_{-\infty}^\infty dz' g(z + z' - x_i) g(z' - x_j). \quad (7)$$

[6] Conceptually, there are several ways to think about parametrization of the winning probabilities. One alternative may be to model the results of the tournament as stemming directly from the distribution of players' abilities and to avoid the intermediate stage of calculating the winning probabilities. Exploratory calculations show that this approach yields qualitatively similar results to the one used in this paper, the quantitative difference being a consequence of the difference in averaging procedures. We would like to use the winning probabilities because they are statistically well-defined quantities that, at least in principle, can be observed in the real world. All other quantities, such as abilities or noise level, are not more than convenient model parameters that cannot be measured or otherwise directly assessed empirically.

Particularly, if $g(\epsilon)$ is normal with zero mean and variance σ^2, one obtains

$$\tilde{w}(x_i, x_j) = \Phi\left(\frac{x_i - x_j}{\sigma\sqrt{2}}\right), \tag{8}$$

where $\Phi(\cdot)$ is the cumulative standard normal density.

Now we can average over all possible realizations of $x_1, ..., x_N$ such that $x_1 \geq x_2 \geq ... \geq x_N$, to get the unconditional[7] probability for player ranked i to win against player ranked j:

$$w_{ij} = N! \int_{-\infty}^{\infty} dx_1 f(x_1) \int_{-\infty}^{x_1} dx_2 f(x_2) ... \int_{-\infty}^{x_{N-1}} dx_N f(x_N) \tilde{w}(x_i, x_j). \tag{9}$$

Here $(N!)^{-1} = \Pr\{x_1 \geq x_2 \geq ... \geq x_N\} = \int_{-\infty}^{\infty} dx_1 f(x_1) \int_{-\infty}^{x_1} dx_2 f(x_2) ... \int_{-\infty}^{x_{N-1}} dx_N f(x_N)$ is the renormalization denominator, which arises because we fixed a specific permutation of x's.

The N-dimensional integral in Eq. (9) can be reduced (see Appendix A) to give

$$\begin{aligned} w_{ij} &= \frac{N!}{(i-1)!(j-i-1)!(N-j)!} \int_{-\infty}^{\infty} dx_1 f(x_1) \int_{-\infty}^{x_1} dx_2 f(x_2) \tilde{w}(x_1, x_2) \\ &\times [1 - F(x_1)]^{i-1} [F(x_1) - F(x_2)]^{j-i-1} [F(x_2)]^{N-j}. \end{aligned} \tag{10}$$

Alternatively, the winning probabilities can be simulated [see Appendix B for a description of the simulation procedure]. The results of the simulations perfectly coincide with those obtained by numerically integrating Eq. (10)[8].

[7] Of course, probabilities w_{ij} remain conditional on the *ex ante* ordering of players.

[8] We use simulations for two reasons. First, it takes less computational time than the calculation of the double integral in (10). Second, since the predictive power is later simulated, exact calculation of the winning probabilities would not add accuracy.

4 The predictive power

Our main objective here is to calculate the probability ρ_1 for a specific player (player 1) to be among the winners of the tournament, i.e. to have a maximal score. For a score vector $\mathbf{s}$ this can be expressed as

$$\rho_1 = \Pr\{(s_1 \geq s_2) \wedge \ldots \wedge (s_1 \geq s_N)\}. \tag{11}$$

By introducing variables $q_1 = s_1, q_2 = s_1 - s_2, ..., q_N = s_1 - s_N$, we need to require that they all be non-negative, i.e. $\rho_1 = \Pr\{q_1 \geq 0, ..., q_N \geq 0\}$. The transformation $\mathbf{s} \leftrightarrow \mathbf{q}$ has a unitary Jacobian, with the inverse transfromation being $s_1 = q_1, s_2 = q_1 - q_2, ..., s_N = q_1 - q_N$. Therefore the joint pdf of $\mathbf{q} = (q_1, ..., q_N)$ is $\pi(q_1, q_1 - q_2, ..., q_1 - q_N)$, and the probability that all q_i are non-negative is

$$\rho_1 = \int_0^\infty dq_1 ... \int_0^\infty dq_N \pi(q_1, q_1 - q_2, ..., q_1 - q_N), \tag{12}$$

where function π is given by Eq. (6).

Then we write

$$\rho_1 = \sum_{\mathbf{b} \in \{0,1\}^M} P(\mathbf{b}) H[S_1(\mathbf{b}) - S_2(\mathbf{b})] \ldots H[S_1(\mathbf{b}) - S_N(\mathbf{b})]. \tag{13}$$

Here $H(z)$ is the step function defined as 1 for $z \geq 0$ and 0 for $z < 0$. The result is very intuitive: we sum over all mutually exclusive tournament outcomes and add up the probabilities of those of them for which player 1's score is maximal.

In Appendix B we describe computational simulations of the round-robin tournament, whose results perfectly agree with the analytical result (13) for moderate N. For large N, Eq. (13) becomes inapplicable practically, since its computational time grows as $2^{N(N-1)/2}$.

In Appendix C we calculate analytically the probability A_k^i for player i to get k points in a round-robin tournament. This result is not of immediate relevance for predictive power exploration, but it allows one to infer, for example, how many points are sufficient (on average) to become a winner of the tournament.

In the limit of $\sigma^2 \to \infty$ (for a fixed number of players N) all the winning probabilities $w_{ij} \to \frac{1}{2}$ independent of the ability distribution $f(\cdot)$. The predictive power therefore has a limiting behavior $\rho_1 \to \rho_1^\infty(N)$, where

$$\rho_1^\infty(N) = \frac{1}{2^M} \sum_{\mathbf{b}\in\{0,1\}^M} H[S_1(\mathbf{b}) - S_2(\mathbf{b})] \dots H[S_1(\mathbf{b}) - S_N(\mathbf{b})]. \tag{14}$$

For example, $\rho_1^\infty(2) = \rho_1^\infty(3) = \frac{1}{2}$, $\rho_1^\infty(4) = \frac{13}{32}$. Recall that we allow for several winners, hence $\rho_1^\infty(N) > 1/N$ for any $N > 2$.

5 Results and discussion

Figure 1 below illustrates the predictive power of a round-robin tournament to identify the best player, ρ_1, as a function of noise level, σ^2, and number of players, N, for three distributions of players' abilities. Specifically, we analyzed by way of computational simulations uniform, normal, and Pareto distributions of players' abilities, with the variance of all normalized to 1. This normalization defines a natural benchmark for the noise level, σ^2. Specifically, it defines three noise regimes: (i) small noise, $\sigma^2 \ll 1$; (ii) intermediate noise, $\sigma^2 \sim 1$; and (iii) large noise, $\sigma^2 \gg 1$. The Figure illustrates predictive power as a function of noise level, for selected numbers of players (N = 2, 4, 8, 16, 32, 64, 128, 256), and as a function of number of players, for selected noise levels, for the three distributions. The computation of the Figure is detailed in Appendix B. [We restricted our computational explorations to those values of N that can be observed in "real life" (e.g., we are not aware of round-robin

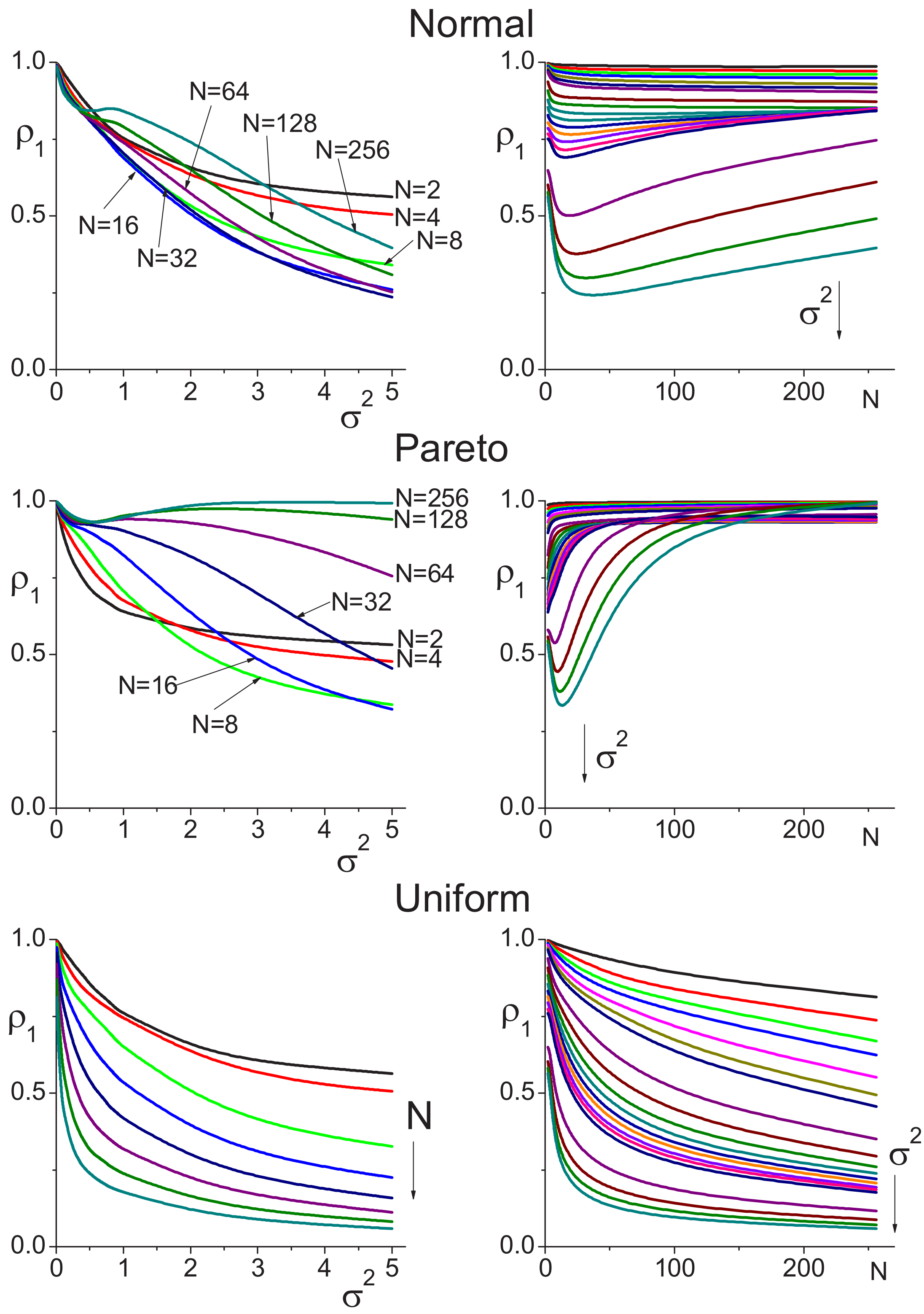

Figure 1: The predictive power ρ_1 as a function of noise level σ^2 for various N, and of the number of players N for various σ^2, for a normal, Pareto, and uniform distribution of players' abilities.

tournaments that involve more than 256 individuals or teams). Also, we varied noise level in a sufficiently broad range relative to the variance of the ability distributions.]

The following discussion of the key results draws on Figure 1. The predictive power ρ_1 displays qualitatively universal features for the small noise and large noise regimes. We discuss those features and their intuition first. Next, we discuss the behavior of ρ_1 in the intermediate noise regime (which turns out to be non-universal and displays non-trivial features) separately for the three distributions. Whenever possible, we present basic intuition why certain behavior is observed, acknowledging, however, that it is a consequence of a complicated interplay of various factors, such as shape of the distribution and combinatorics, which are hard to completely intuit.

- The predictive power for one player ($N = 1$) always equals 1. Therefore, the predictive power necessarily decreases initially when additional competitors join the tournament. For larger number of players, the behavior might become non-monotonic; if and when non-monotonicities occur depends on the distribution of players' abilities (and the noise level). More details below.

- Across all three distributions, the predictive power behaves qualitatively the same for the small noise regime. Small noise adds small probability of upsets. No discontinuity in the predictive power can be expected when going from zero to non-zero noise. Since the predictive power equals its maximal possible value 1 for zero noise, it must necessarily decline for small noise. Clearly, this result is universal, i.e. it holds for arbitrary number of players $N \geq 2$ and all distributions of players' abilities $f(x)$. How fast and how far ρ_1 will decline, of course, depends on N and specific shape of $f(x)$.

- Across all three distributions, the predictive power behaves qualitatively the same for the large noise regime. For any given N the predictive power converges to a constant which is a function of N only (and not the distributional specification). In effect, this result is a straightforward application of combinatorics, as we demonstrated in Section 4. Note that for both $N = 2$ and $N = 3$, the predictive power converges to $1/2$ as the noise goes to infinity. For $N = 4$, the predictive power converges to $13/32$. However, for noise levels that are of relevance for practical purposes, these results are of lesser importance. The basic intuition is worth re-iterating: As noise increases, ability becomes less and less important in determining the outcome of a match, while chance becomes increasingly important. This asymptotic behavior of ρ_1 is also universal.

- The behavior of the predictive power becomes qualitatively non-universal in the intermediate noise regime where the noise level is comparable with the variance of the distribution of abilities. For uniform distributions of abilities, predictive power is a monotonically decreasing function of noise and number of players. The intuition for this result is straightforward: As the performance fluctuations of equally distanced (in terms of ability) players increase, 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 beating a lower-ranked player decreases (except, of course, the lowest-ranked player), possibly neutralizing the net effect on the expected score. 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 less 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 less chances of being upset, and more chances of upsetting).

- The previous result illustrates a fundamental principle: Noise is a redistributor of scores in that it gives from the "haves" to the "havenots". Obviously this distributional process is affected by the shape of the distribution of players' abilities.

- For uniform distributions of abilities, competition at the top is tougher than for normal and Pareto distributions (as it will be for all distributions with falling upper tail) and therefore the predictive power is lower, the more so the more competitors there are. This is true for all noise levels. The intuition is clear: For uniform distributions interior players are symmetrically affected in that players indexed i and $N - i + 1$ have the same net effect in expected scores (albeit with opposite signs).

- For Pareto distributions, players are differentially affected by increasing noise. Specifically, the probabilities of upsets will be lower for the top players because lower-ranked players are bunched more tightly and hence have more chances to score upsets. Of course, the decreased probability of beating a lower-ranked player possibly neutralizes the net effect on the expected score. However, now the net effect of noise on the expected score is determined by the number of players in a more complicated fashion.

- For the two non-uniform distributions that we studied, and probably for all non-uniform distributions with falling upper tail, the predictive power displays a bifurcation-like behavior as a funcion of the noise level, σ^2. Specifically, for N below a critical value N_c (which depends on the distribution) the predictive power decreases monotonically, while for N above N_c, the predictive power displays both a local minimum and a local maximum. Take, for example, the

Pareto distribution. For $N = 8$, the predictive power falls monotonically; for $N = 64$, it has a minimum at about $\sigma^2 = .5$, and a maximum at about $\sigma^2 = 1$; with $N_c = 32$ being approximately the switching point.

- So, what intuition then drives the unintuitive behavior of the predictive power for large N (such as 64)?

  As N increases the long tail of the Pareto distribution implies that relative to the low-ability players the top-ranked player moves away from the second-ranked player ever more. Therefore, for sufficiently small noise, the predictive power drops less for larger N. This trend is countered by a countervailing trend as noise increases. Let us distinguish the cases of small N and large N. Obviously, as noise increases, so does the probability of upsets. However, this increase in probability is decreasing as we move up the ranking since the distance between players increases, on average. When N is small the increase in the probability of upsets is the only effect. When N is large then the top-ranked players score increases relatively to the lower-ranked players' because the lower-ranked a player is, the more ferocious competition that player faces. This result leads to the surprising and initially unintuitive upward swing in the predictive power as a function of noise for large N.

- For normal distributions, the intuition stemming from the Pareto distribution is a useful point of departure. Note that, ignoring for the time being the difference in tails, the Pareto distribution is in a sense the upper half of the normal distribution. Hence, we should expect qualitatively somewhat similar behavior. And indeed, for low noise (roughly up to $\sigma^2 = 1$) we see very similar behavior. (Of course, the quantitative behavior differs somehow.) Now look at the median player in the normal distribution. As the performance fluctuations of every player increase, the probability for upsets increases (but not uniformly) for all players (except, of course, the top-ranked player). But note that this result also implies that the probability of beating a lower-ranked player decreases (except, of course, the lowest-ranked player), possibly neutralizing the net effect on the expected score. 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). This, of course, is an argument similar to the one we made above for the case of the uniform distribution. Qualitatively, the net effect of this balancing act is negative for players to the right of the median player, and positive for players to the left of the median players, for the same rationale laid out for the uniform distribution above. The difference in behavior stems from the fundamental difference in the gestalt of the tails of the distribution.

6 Conclusion

In the present paper, we studied analytically and by way of computational simulations, the properties of round-robin tournaments as a function of noise level, number of players, and distribution of players' abilities. For moderate N we could use analytical results to verify our simulations, whereas for large N we had to rely on simulations. Perfect agreement of the analytical results with simulations for moderate N suggests that the simulation results are correct.

A planner who has decided to conduct a round-robin tournament might benefit from our insights in a number of ways: If he or she knows, or at least has some inkling, about the distribution of abilities, and if the distribution happens to be one of the three that we studied, then the planner can estimate from the probabilities of upsets the noise level σ^2 for our model. In fact, the planner could infer the properties of a distribution from the probabilities of upsets which, in principle, are observable.

We note that these assumptions are not quite as far fetched as they seem: most professional sports have ranking schemes. Some of these schemes are based on the statistics of upsets in various ways [e.g., in chess, table tennis, soccer, and American football)].

If the planner knows distribution and noise, then he can identify the optimal number of players. This decision, obviously, requires the definition of an objective function such as the maximization, or minimization, of the probability of the best team winning. But once that decision has been made, our results suggest interesting choices. For the maximization case, as regards the Pareto distribution at $\sigma^2 = 3$ and moderate N, it makes sense to decrease the number of participants. For the minimization case, with the same distribution but $\sigma^2 = 1$, the number of players again should be decreased. The planner, generally, wants to keep in mind that for non-uniform distributions of abilities the predictive power depends non-monotonically on N.

In Ryvkin (2004) we analyze similarly elimination tournaments, and in later papers we will compare the properties of these two polar matching schemes, as well as variants thereof. We will also analyze the robustness of the present results to different (asymmetric, heteroscedastic) specifications of the noise. We conjecture that such modification will have quantitative effects but will not change the results reported here in any fundamental manner.

References

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.

Esteban, J., Ray, D., 2001. Social desicion rules are not immune to conflict. Economics of Governance, 2, 59-67.

Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: gender differences. Quarterly Journal of Economics, 1049-1074.

Green, J., Stokey, N., 1983. A comparison of tournaments and contracts. The Journal of Political Economy, 91, 349-364.

Harary, E., Moser, L., 1966. The theory of round robin tournaments. American Mathematical Monthly, 73, 231-246.

Harrison, W., 2004. Field experiments and control. *Mimeo*.

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.

Knoeber, C., Thurman, W., 1994. Testing the theory of tournaments: An empirical analysis of broiler production. Journal of Labor Economics, 10, 357-379.

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.

Moldovanu, B., Sela, A., 2004. Contest Architecture. Discussion paper, University of Mannheim.

Moon, J., Pullman, N., 1970. On generalized tournament matrices. SIAM Review, 12, 384-399.

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.

Roth, A., 2002. The economist as engineer: game theory, experimentation, and computation as tools for design economics. Econometrica, 70, 1341-1378.

Rubinstein, A., 1980. Ranking the participants in a tournament. SIAM Journal of Applied Mathematics, 38, 108-111.

Ryvkin, D., 2004. The predictive power of noisy elimination tournaments. *Manuscript*.

Schotter, A., Weigelt, K., 1992. Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results The Quarterly Journal of Economics, 107, 511-539.

Taylor, C., 1995. Digging for golden Carrots: an analysis of research tournaments. The American Economic Review, 85, 872-890.

Appendix

A Winning probabilities

In this section of the Appendix we show how the expression (10) is obtained. The variables in the ordered integral in Eq. (9) are assumed to change in the following ranges:

$x_1 \in (-\infty, \infty)$, $x_2 \in (-\infty, x_1)$, ..., $x_N \in (-\infty, x_{N-1})$.

However, they can be rearranged so that

$x_i \in (-\infty, \infty)$, $x_j \in (-\infty, x_i)$,

$x_1 \in (x_i, \infty)$, $x_2 \in (x_i, x_1)$, ..., $x_{i-1} \in (x_i, x_{i-2})$,

$x_{i+1} \in (x_j, x_i)$, $x_{i+2} \in (x_j, x_{i+1})$, ..., $x_{j-2} \in (x_j, x_{j-1})$,

$x_{j+1} \in (-\infty, x_j)$, $x_{j+2} \in (-\infty, x_{x+1})$, ..., $x_N \in (-\infty, x_{N-1})$.

This reordering corresponds to the following integral:

$$
\begin{aligned}
w_{ij} &= N! \int_{-\infty}^{\infty} dx_i f(x_i) \int_{-\infty}^{x_i} dx_j f(x_j) \tilde{w}(x_i, x_j) \\
&\times \int_{x_i}^{\infty} dx_1 f(x_1) \int_{x_i}^{x_1} dx_2 f(x_2) \ldots \int_{x_i}^{x_{i-2}} dx_{i-1} f(x_{i-1}) \\
&\times \int_{x_j}^{x_i} dx_{i+1} f(x_{i+1}) \int_{x_j}^{x_{i+1}} dx_{i+2} f(x_{i+2}) \ldots \int_{x_j}^{x_{j-2}} dx_{j-1} f(x_{j-1}) \\
&\times \int_{-\infty}^{x_j} dx_{j+1} f(x_{j+1}) \int_{-\infty}^{x_{j+1}} dx_{j+2} f(x_{j+2}) \ldots \int_{-\infty}^{x_{N-1}} dx_N f(x_N).
\end{aligned}
$$

In the last three lines, the integrals are easily calculated, and we obtain Eq. (10).

B Simulations

In this section of the Appendix we describe how the simulations were done. The whole procedure consisted of two stages: (i) simulating the winning probabilities w_{ij}; (ii) simulating the predictive power ρ_1.

Winning probabilities. The algorithm for this simulation was the following:

1) Populate the matrix Z_{ij} with zeros;

2) Independently draw N numbers $x_1, \ldots, x_N$ from the distribution $f(\cdot)$ and order them so that $x_1 > x_2 > \ldots > x_N$;

3) Populate the matrix p_{ij} for all $1 \leq i < j \leq N$ according to the following rule: take x_i and x_j such that $i < j$, draw independently two noise terms ϵ_i and ϵ_j, let $y_i = x_i + \epsilon_i$ and $y_j = x_j + \epsilon_j$, and set $p_{ij} = 1$ if $y_i > y_j$ and $p_{ij} = 0$ if $y_i < y_j$;

4) Add the matrix p_{ij} from step 3 to Z_{ij};

5) Go to step 2.

The whole procedure is to be repeated a large number of times, T, and the average over realizations matrix p_{ij} will be the matrix of winning probabilies:

$$w_{ij} = \langle p_{ij} \rangle = \frac{Z_{ij}}{T}. \tag{15}$$

Predictive power. The algorithm for this simulation was the following:

1) Set a counter $c = 0$;

2) Populate the matrix p_{ij} for all $1 \leq i < j \leq N$ according to the following rule: take $i < j$, draw a uniform random number r from interval $(0, 1)$; set $p_{ij} = 0$ if $r > w_{ij}$ and $p_{ij} = 1$ if $r < w_{ij}$;

3) Calculate the scores of players as $s_i = \sum_{j \neq i} p_{ij}$;

4) If player 1 has a maximal score, i.e. $s_1 \geq s_i$ for all $1 < i \leq N$, then increment c by 1;

5) Go to step 2.

The whole procedure is to be repeated a large number of times, T, and then the share of times when the counter was incremented gives the predictive power:

$$\rho_1 = \frac{c}{T}. \tag{16}$$

The winning probabilities and the predictive power (the latter only for moderate N) were also calculated analytically using Eqs. (10) and (13). The results perfectly agree with those of the simulations.

C Individual distribution of scores

In this Section of the Appendix we obtain a result, which is not of immediate relevance for the predictive power analysis, but is interesting from the theoretical point of view. It answers the following question: what is the probability for player with *ex ante* rank i to get k points in a round-robin tournament.

Throughout the tournament, every player i plays $N-1$ matches with players $\mathcal{P} \setminus i$. Every match (i,j) is a Bernoulli trial with the probability of success w_{ij}. In case of success, player i gets 1 point. In a match (i,j), the number of points player i gets can be represented by a discrete random variable p_{ij} [cf. Eq. (1)] with pdf $\phi_{ij}(p) = w_{ij}\delta(p-1) + (1-w_{ij})\delta(p)$.

The total score player i will get is $s_i = \sum_{j\neq i} p_{ij}$. Let $\mathcal{B}_N^{-i}$ denote a set of $(N-1)$-bit binary vectors $\mathbf{b} = (b_1, ..., b_{i-1}, b_{i+1}, ..., b_N)$, whose every component is 0 or 1. The pdf of the total score for player i can then be written as

$$\pi_i(s) = \sum_{\mathbf{b}\in\mathcal{B}_N^{-i}} \prod\nolimits_{j\neq i} \left[b_j w_{ij} + (1-b_j)(1-w_{ij})\right] \delta(s - \sum\nolimits_{j\neq i} b_j). \tag{17}$$

It is clear that actually the pdf for s_i has the form

$$\pi_i(s) = \sum\nolimits_{k=0}^{N-1} A_k^i \delta(s-k), \tag{18}$$

where A_k^i is the probability for player i to get k points. Eq. (17) allows for the calculation of those probabilities.

Theorem. Let $\mathbf{w}^{-i} = (w_{i1}, ..., w_{i,i-1}, w_{i,i+1}, ..., w_{iN})$ be a $(N-1)$-dimensional vector of winning probabilities for player i. Then the probability for player i to get k points in a round-robin tournament of N players is

$$A_k^i = \sum\nolimits_{l=k}^{N-1} (-1)^{l-k} C_l^k Q_l(\mathbf{w}^{-i}). \tag{19}$$

Here C_l^k are the binomial coefficients; the polynomials Q_l are defined as follows. Let $\mathbf{u} = (u_1, ..., u_n)$ be a vector of n real numbers, and

$$\begin{aligned}
&Q_0^n(\mathbf{u}) = 1,\\
&Q_1^n(\mathbf{u}) = u_1 + ... + u_n,\\
&Q_2^n(\mathbf{u}) = u_1u_2 + u_1u_3 + ... + u_1u_n + u_2u_3 + ... + u_2u_n + ... + u_{n-1}u_n,\\
&...\\
&Q_k^n(\mathbf{u}) = \sum_{j_1<...<j_k} u_{j_1}u_{j_2}...u_{j_k},\\
&...\\
&Q_n^n(\mathbf{u}) = u_1u_2...u_n.
\end{aligned} \tag{20}$$

Proof. The only place i enters the right-hand side of Eq. (19) is through $Q_k^{N-1}(\mathbf{w}^{-i})$, but the polynomials Q_k^n are completely symmetric with respect to permutations of their arguments, therefore we only need to prove the theorem for any particular i, for example $i = 1$. We prove by induction over the number of players N. The induction base, $N = 2$, is obvious [it can be directly calculated using Eq. (17)]. Now assume that Eq. (19) holds for N players. Suppose we add one more player, $N + 1$ (again, using the symmetricity of Q_k^n, we can always assume that the added player occupies the last ranking position), with the winning probability for player 1 over her being $w_{1,N+1}$. Then $A^1_j(N+1)$ can be expressed as follows:

$$A^1_j(N+1) = A^1_j(N)(1 - w_{1,N+1}) + A^1_{j-1}(N)w_{1,N+1}. \tag{21}$$

Indeed, there are only two mutually exclusive ways to get j points: to get j points playing with the former $N-1$ players and to lose to the new player (first term), and to get $j-1$ points and win against the new player (second term), respectively. Note that the polynomials Q^n_k have the following property:

$$Q^{n+1}_k(u_1, ..., u_n, u_{n+1}) = Q^n_k(u_1, ..., u_n) + u_{n+1}Q^n_{k-1}(u_1, ..., u_n), \tag{22}$$

if we set $Q^n_{n+1} = Q^n_{-1} = 0$.

From Eq. (19) (that holds for N by the induction assumption) we have

$A^1_{j-1}(N) - A^1_j(N) = \sum_{k=j-1}^{N-1}(-1)^{k-j+1}C^{j-1}_k Q^{N-1}_k(\mathbf{w}^{-1}) - \sum_{k=j}^{N-1}(-1)^{k-j}C^j_k Q^{N-1}_k(\mathbf{w}^{-1}) =$
$= \sum_{k=j}^{N-1}(-1)^{k-j}[-C^{j-1}_k - C^j_k]Q^{N-1}_k(\mathbf{w}^{-1}) + Q^{N-1}_{j-1}(\mathbf{w}^{-1}) =$
$= \sum_{k=j}^{N-1}(-1)^{k-j+1}C^j_{k+1}Q^{N-1}_k(\mathbf{w}^{-1}) + Q^{N-1}_{j-1}(\mathbf{w}^{-1}) =$
$= \sum_{k=j-1}^{N-1}(-1)^{k-j+1}C^j_{k+1}Q^{N-1}_k(\mathbf{w}^{-1})$.

Therefore, using Eqs. (19), (21) and (22), one obtains

$A^1_j(N+1) = A^1_j(N) + w_{1,N+1}[A^1_{j-1}(N) - A^1_j(N)] =$
$= \sum_{k=j}^{N-1}(-1)^{k-j}C^j_k Q^{N-1}_k(\mathbf{w}^{-1}) + w_{1,N+1}\sum_{k=j-1}^{N-1}(-1)^{k-j+1}C^j_{k+1}Q^{N-1}_k(\mathbf{w}^{-1}) =$
$= \sum_{k=j}^{N}(-1)^{k-j}C^j_k Q^N_k(\mathbf{w}^{-1}, w_{1,N+1})$,

which completes the proof of the induction iteration. *Q.E.D.*

CERGE-EI
P.O.BOX 882
Politických vezòù 7
111 21 Prague 1
Czech Republic
http://www.cerge-ei.cz