Optimal prizes in tournaments with risk-averse agents

Mikhail Drugov* Dmitry Ryvkin†

This version: June 6, 2019

Abstract

We characterize the optimal allocation of prizes in rank-order tournaments with risk-averse agents. Risk-aversion as well as the presence of heavy tails in the distribution of noise can lead to the optimality of prize sharing. An increase in risk-aversion leads to more prize sharing being optimal, in the sense of the majorization order. The results for Tullock contests follow as a special case.

Keywords: tournament, optimal allocation of prizes, risk-averse agents, majorization.

JEL codes: C72, D82, J31.

*New Economic School and CEPR, mdrugov@nes.ru.
†Department of Economics, Florida State University, Tallahassee, FL 32306-2180, USA, dryvkin@fsu.edu.
1 Introduction

In this paper, we obtain a general solution to the canonical problem of optimal allocation of prizes in rank-order tournaments with risk-averse agents. The main question we address is whether, and under what conditions, a *winner-take-all* (WTA) prize allocation rule, whereby the entire prize budget is awarded to a single winner, is effort-maximizing.

Rank-order tournament incentives are ubiquitous in environments such as the design of creative products, innovation or job market competition. Think, for example, of teams of architects submitting their designs; or participants of a coding contest on a platform such as TopCoder,\(^1\) or an innovation contest such as XPRIZE,\(^2\) submitting their solutions; or job candidates competing for an attractive position. In these environments it is often hard to assign a precise cardinal measure to the agents’ output or quality, but it may be relatively easy for the principal to assign ordinal rankings. Agents are then awarded by fixed prizes based on these rankings.

We consider a simple static rank-order tournament model similar to the model of Lazear and Rosen (1981). Agents exert costly effort and produce random output that is equal to effort distorted by additive idiosyncratic noise. The agents are symmetric and risk-averse, with a well-behaved utility of money. The principal has a total prize budget she needs to allocate on the basis of the ranking of the agents’ output levels.

WTA incentives have always been at the center of a broader debate among economists and management practitioners about the role of high-powered incentives and reliance on strong meritocracy and, ultimately, strong *ex post* inequality in reward structures. Proponents of WTA argue it is efficient and motivates agents to work hard, while opponents appeal to humanistic values, such as employee morale and aversion to inequality, or perverse incentives generated by excessive competition. These effects are especially evident when agents are homogeneous or nearly homogeneous in ability, in which case tournaments incentivize effort but ultimately reward pure luck (Frank, 2016). Therefore, despite the preponderance of theoretical arguments in favor of the utilitarian optimality of WTA, of interest are results showing why *prize sharing*, in the form of a substantial departure from WTA, can also be optimal, especially if its optimality is based on standard, neoclassical preferences (as opposed to, say, other-regarding preferences or behavioral phenomena).

Drugov and Ryvkin (2018) studied how the optimal allocation of prizes in tournaments of risk-neutral agents is affected by the properties of noise. They showed that WTA

---

\(^{1}\)See [https://www.topcoder.com/](https://www.topcoder.com/).

\(^{2}\)See [https://xprize.org/](https://xprize.org/).
is optimal if the distribution of noise has an increasing failure rate (IFR), but prize sharing can become optimal when the noise has a decreasing failure rate (DFR) in its upper tail. Such heavy-tailed distributions, such as a power law, are common in many environments (Gabaix, 2016). The underlying mechanism is related to the behavior of spacings—the distances between the adjacent order statistics of noise. The fast growth of spacings with rank for heavy-tailed distributions reduces the marginal impact of effort at the top of the rankings; essentially, the luckiest agent’s shock is likely to be so large that it does not make sense to try to catch up. As a result, the equilibrium effort can be increased by sharing the prize among lower ranks until it reaches the area where spacings do not grow too fast.

Risk-aversion introduces another concern for the agents and the contest designer. Indeed, risk-averse agents view effort exertion in a tournament as risky because its cost is deterministic but benefit uncertain. Prize sharing then provides insurance to the agents and can generate a higher effort. However, it is not clear how the two effects—risk-aversion and noise—interact. Intuitively, for any degree of prize sharing that is optimal under risk-neutrality (due to heavy tails), risk aversion should favor even more prize sharing. Moreover, prize sharing can be optimal under risk-aversion even when WTA is optimal for risk-neutral agents (e.g., when the noise is IFR). In this paper, we provide a systematic characterization of these effects. We show that an increase in the agents’ risk-aversion, in the sense of the coefficient of absolute or relative risk-aversion (or, equivalently, the concave transform order for utility functions) leads to an optimal allocation of prizes that involves more prize sharing, in the sense of the majorization order.

The interaction between risk aversion and the distribution of noise runs deeper, however. One may think that as agents become more risk-averse, increasing the number of prizes is always optimal. Yet, this intuition is only partially correct. We show that for any unimodal distribution of noise and any number of agents there is a maximum number of prizes that is ever optimal to give, for any degree of risk-aversion. In other words, increasing risk-aversion leads to more prizes but only up to a point. A stronger and even more surprising result is that some noise distributions (for example, uniform) satisfy a universal optimality of WTA property such that awarding more than one prize is never optimal.

The rest of the paper is organized as follows. This section continues with a brief survey of the relevant literature. Section 2 introduces the model. Section 3 characterizes optimal prize allocations. The effect of risk aversion is studied in Section 4. Section 5 contains various examples. Section 6 discusses the results in connection to the optimality
of WTA prize schemes. Section 7 concludes. The proofs are contained in the Appendix.

**Relation to the existing literature.** This paper contributes to the literature on the optimal allocation of prizes in contests. Two types of models of (static) multi-prize contests have been used historically in the literature, differing mostly in their assumptions about the winner determination process. Noisy, or imperfectly discriminating, contest models are the rank-order tournament model of Lazear and Rosen (1981) and the nested contest model of Clark and Riis (1996), which is a multi-prize adaptation of the classic rent-seeking model of Tullock (1980). The main feature of these models is the presence of idiosyncratic noise, or luck, in the transformation of agents’ effort into output, and hence a higher effort does not guarantee a higher rank (and a higher prize). In contrast, perfectly discriminating contest models (e.g., Baye, Kovenock and De Vries, 1996; Moldovanu and Sela, 2001; Siegel, 2009) are essentially all-pay auctions with a one-to-one assortative mapping of effort rankings to prizes. The mechanisms underlying incentives in the two types of models are different, and hence they produce diverging results regarding the optimality of WTA.

Papers related most closely to ours are the ones studying the optimal allocation of prizes in noisy tournaments (Krishna and Morgan, 1998; Akerlof and Holden, 2012; Balafoutas et al., 2017; Ales, Cho and Körpeoğlu, 2017; Drugov and Ryvkin, 2018), as well as in nested Tullock contests (Clark and Riis, 1996; Schweinzer and Segev, 2012; Fu, Wang and Wu, 2019). Krishna and Morgan (1998) assume that the principal allocates a fixed budget and show that WTA is always optimal in tournaments of size $n \leq 3$ for risk-averse agents, and of size $n \leq 4$ for risk-neutral agents, if the distribution of noise is unimodal and symmetric (the result is referred to as the “winner-take-all principle” for small tournaments). Akerlof and Holden (2012) consider a tournament with symmetric risk-averse agents where the prize budget is not fixed but the agents’ participation constraint is binding. They show that various patterns of prize sharing are optimal depending on the agents’ risk-aversion and prudence, but the results are too convoluted to discern the effects of the properties of noise or general conditions for the optimality of WTA. Balafoutas et al. (2017) show that prize sharing can be optimal if agents are heterogeneity even when they are risk-neutral. Drugov and Ryvkin (2018) study the effects of the shape of the distribution of noise on optimal prize structures for risk-neutral agents and show that WTA

---

3In fact, the latter model can be written as a special case of the former (Fu and Lu, 2012).

4A separate strand of literature explores “large” contests where each player takes the distribution of actions of others as given, and the equilibrium is defined as a self-consistent distribution maximizing individual payoffs (Glazer and Hassin, 1988; Olszewski and Siegel, 2016).
is optimal when the noise is IFR but prize sharing becomes optimal in the presence of heavy tails. This result generalizes the findings of Clark and Riis (1996) and Schweinzer and Segev (2012) who showed the optimality of WTA for nested Tullock contests whose equilibrium is isomorphic to that of a tournament with additive noise with the Gumbel distribution (which is IFR). In a similar setup, Ales, Cho and Körpeoğlu (2017) show that awarding multiple prizes in innovation tournaments with risk-neutral agents is optimal when the distribution of noise is log-convex; however, unlike other papers, they do not assume that prizes are monotone in rank. Recently, Fu, Wang and Wu (2019) showed that WTA may be sub-optimal in a nested Tullock contest with risk-averse agents and a non-separable utility function combining in its argument the prizes and effort costs.

Our results directly generalize those of Krishna and Morgan (1998) and Drugov and Ryvkin (2018). Relative to the former paper, we generalize the WTA principle to tournaments with arbitrary sizes and (possibly skewed, multi-modal) noise distributions, and relate it directly to the properties of noise. Relative to the latter, we introduce agents’ risk-aversion and show how the properties of noise are combined with the curvature of the utility of money to generate optimal prize sharing. As a special case of our results, we obtain a sufficient condition for when WTA is no longer optimal in nested Tullock contests. The result is qualitatively similar to the one obtained by Fu, Wang and Wu (2019), although not exactly the same due to their assumption of non-separability of prizes and effort costs in the utility of money.

2 Model setup

Consider a tournament of \( n \geq 2 \) identical agents indexed by \( i = 1, \ldots, n \). The agents simultaneously and independently choose effort levels \( e_i \in \mathbb{R}_+ \). The cost of effort \( e_i \) to agent \( i \) is \( c(e_i) \), where \( c : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) is a strictly increasing, strictly convex and twice differentiable cost function, with \( c(0) = c'(0) = 0 \). The output of agent \( i \) is her effort perturbed by additive noise, \( y_i = e_i + X_i \).\(^5\) Shocks \( X_i \) are zero-mean, i.i.d. copies of random variable \( X \) that has a cumulative density function (cdf) \( F(\cdot) \) and probability density function (pdf) \( f(\cdot) \) defined on an interval support \( \mathcal{X} = [\underline{x}, \overline{x}] \) (finite or infinite). The pdf \( f(\cdot) \) is atomless, continuous, piece-wise differentiable and square-integrable.

A risk-neutral principal observes the ranking of outputs and allocates rank-dependent prizes \( v = (v_1, \ldots, v_n) \) to the \( n \) agents. Specifically, an agent whose output is ranked \( r \)

\(^5\) Via a change of variables, this model can also accommodate tournaments with multiplicative noise, with \( y_i = e_iX_i \) (see Jia, 2008; Jia, Skaperdas and Vaidya, 2013; Ryvkin and Drugov, 2019).
(where \( r = 1 \) corresponds to the highest output, \( r = 2 \) to the second highest, etc.) receives a prize \( v_r \).\(^6\) Prizes are nonnegative, decreasing\(^7\) in rank, \( v_1 \geq v_2 \geq \ldots \geq v_n \geq 0 \), and satisfy the budget constraint \( \sum_{r=1}^{n} v_r = 1 \).

The agents are risk-averse expected utility maximizers, with a bounded Bernoulli utility function of money \( u : [0,1] \to \mathbb{R}_+ \). Without loss of generality, we set \( u(0) = 0 \) and \( u(1) = 1 \). We assume that \( u(\cdot) \) is continuous, strictly increasing, concave, and twice differentiable except, possibly, at zero.

Consider a symmetric pure-strategy Nash equilibrium where all agents exert some effort \( e^* > 0 \). Following the symmetric opponents form approach (SOFA) (Hefti, 2017), assume that all agents except one indicative player choose effort \( e^* \). The expected utility of the indicative agent from some deviation effort \( e \) is

\[
U(e,e^*) = \sum_{r=1}^{n} \pi^{(r,n)}(e,e^*)u(v_r) - c(e_i),
\]

where \( \pi^{(r,n)}(e,e^*) \) is the probability that the indicative agent’s output is ranked \( r \). This probability is given by

\[
\pi^{(r,n)}(e,e^*) = \binom{n-1}{r-1} \int_{X} F(e - e^* + x)^{n-r}[1 - F(e - e^* + x)]^{r-1}dF(x).
\]

Indeed, in order to be ranked \( r \), the indicative agent’s output must be higher than the output of exactly \( n - r \) other agents, and there are \( \binom{n-1}{r-1} \) ways to choose those agents. The symmetric first-order condition, \( U_e(e^*,e^*) = 0 \), produces the equation

\[
\sum_{r=1}^{n} \beta_{r,n} u(v_r) = c'(e^*),
\]

where \( \beta_{r,n} \equiv \pi_{e}^{(r,n)}(e^*,e^*) \), the marginal probabilities of reaching rank \( r \), are given by

\[
\beta_{r,n} = \binom{n-1}{r-1} \int_{X} F(x)^{n-r-1}[1 - F(x)]^{r-2}[n - r - (n-1)F(x)]f(x)dF(x).
\]

In what follows, we assume that the \( e^* \) given by Eq. (2) is a symmetric equilibrium effort level.\(^8\) We also assume, similar to Krishna and Morgan (1998) and Moldovanu and Sela

---

6Ties in the ranking are broken randomly, but occur with zero probability for an atomless \( f(\cdot) \).

7Throughout this paper, “increasing” means nondecreasing and “decreasing” means nonincreasing. Whenever the distinction is important, we use the terms “strictly increasing” and “strictly decreasing.”

8For the same model with risk-neutral agents, Drugov and Ryvkin (2018) formulate sufficient con-
(2001), that the participation constraint, \( U(e^*, e^*) \geq u(0) = 0 \), is never binding.

Coefficients \( \beta_{r,n} \) are determined entirely by the distribution of noise. The following lemma summarizes some of their properties.

**Lemma 1**

(i) \( \sum_{r=1}^{n} \beta_{r,n} = 0 \), \( \beta_{1,n} > 0 \), \( \beta_{n,n} < 0 \); if \( f(\cdot) \) is symmetric, then \( \beta_{r,n} = -\beta_{n-r+1,n} \) for all \( r \).

(ii) If \( f(\cdot) \) is log-concave, then \( \beta_{r,n} \) is decreasing in \( r \).

(iii) If \( f(\cdot) \) is log-convex and \( f(\bar{x}) = 0 \), then \( \beta_{r,n} \) is increasing in \( r \) for \( r \leq n - 1 \).

(iv) If \( f(\cdot) \) is first log-concave, then log-convex with \( f(\bar{x}) = 0 \), then \( \beta_{r,n} \) is unimodal in \( r \) for \( r \leq n - 1 \).

(v) If \( f(\cdot) \) is unimodal, then \( \beta_{r,n} \) is single-crossing; that is, there exists an \( \hat{r} \leq n - 1 \) such that \( \beta_{r,n} > 0 \) for \( r \leq \hat{r} \) and \( \beta_{r,n} \leq 0 \) for \( r > \hat{r} \).

3 Optimal prize allocations

We consider a principal whose objective is to maximize total expected output, \( ne^* \). It follows immediately from (2), and the assumption that \( c(\cdot) \) is strictly convex, that the principal’s prize allocation problem has the form

\[
\max_{v_1, \ldots, v_n} \sum_{r=1}^{n} \beta_{r,n} u(v_r) \quad \text{s.t. } v_1 \geq \ldots \geq v_n \geq 0, \quad \sum_{r=1}^{n} v_r = 1. \tag{4}
\]

We start with three simple results.

**Noise scaling.** An important property of problem (4) is that the objective function is linear in coefficients \( \beta_{r,n} \). Therefore, any transformation of noise that leads to a linear transformation of coefficients \( \beta_{r,n} \) will not change the optimal allocation of prizes. One example of such transformation is scaling.
Lemma 2 Suppose random variable $Y$ is a scaling transformation of $X$ such that $Y = sX$ for some constant $s > 0$. Then the optimal allocation of prizes under noise $Y$ is the same as under noise $X$.

The case of $n = 3$, symmetric $f(\cdot)$. Problem (4) has a simple solution for small tournaments with $n = 3$ players and symmetric noise densities. Part (i) of Lemma 1 implies $\beta_{2,3} = 0$ and $\beta_{3,3} < 0$, producing the following result.

Corollary 1 Suppose $f(\cdot)$ is symmetric and $n = 3$. Then WTA is optimal.

Corollary 1 generalizes the “WTA principle for small tournaments” identified by Krishna and Morgan (1998) for unimodal symmetric distributions. In fact, unimodality is not needed for the result to hold.

Universally optimal WTA. If $\beta_{r,n} \leq 0$ for all $r > 1$, it is obvious that WTA is optimal. If this property holds for any $n \geq 2$ and any utility function $u(\cdot)$, we have a case of the universal optimality of WTA. One example is the uniform distribution of noise. The following lemma characterizes a class of noise distributions with this property.

Lemma 3 If $f(\cdot)$ is increasing, then WTA is optimal for any $n \geq 2$ and any utility function $u(\cdot)$.

From this point on, we consider the more interesting case when $f(\cdot)$ is not increasing.

3.1 Log-concave $f(\cdot)$

When $f(\cdot)$ is log-concave, it follows from Lemma 1(ii) that coefficients $\beta_{r,n}$ are decreasing in $r$. Then there exists an $\hat{r} \leq n - 1$ such that $\beta_{r,n} > 0$ for $r \leq \hat{r}$ and $\beta_{r,n} \leq 0$ for $r > \hat{r}$. It is clear that assigning positive prizes to any ranks $r > \hat{r}$ is not optimal. Thus, problem (4) reduces to

$$
\max_{v_1,\ldots,v_\hat{r}} \sum_{r=1}^{\hat{r}} \beta_{r,n} u(v_r) \quad \text{s.t. } v_1 \geq \ldots \geq v_{\hat{r}} \geq 0, \quad \sum_{r=1}^{\hat{r}} v_r = 1. 
$$

(5)

Since all coefficients $\beta_{r,n}$ in problem (5) are positive, it is a convex programming problem with a concave objective and linear constraints; therefore, Kuhn-Tucker conditions are
both necessary and sufficient for maximization. Consider a version of problem (5) without
the monotonicity constraint:

$$\max_{v_1,\ldots,v_{\hat{r}}} \beta_{r,n} u(v_r) \quad \text{s.t.} \quad v_1,\ldots,v_{\hat{r}} \geq 0, \quad \sum_{r=1}^{\hat{r}} v_r = 1. \quad (6)$$

As we show, a solution to problem (6) automatically satisfies the monotonicity constraint
$v_1 \geq \ldots \geq v_{\hat{r}}$, and hence it is also a solution to problem (5) and, augmented with the
assignment $v_{\hat{r}+1} = \ldots = v_n = 0$, to original problem (4). The Kuhn-Tucker conditions for
problem (6) are

$$\beta_{r,n} u'(v_r) \leq \lambda, \quad \text{with equality if } v_r > 0, \quad r = 1,\ldots,\hat{r}, \quad (7)$$

for some Lagrange multiplier $\lambda > 0$.

Define correspondence $\phi(\beta,\lambda) : \mathbb{R}_+ \times \mathbb{R}_+ \to [0,1]$,

$$\phi(\beta,\lambda) = \begin{cases} 0, & \text{if } \beta u'(0) \leq \lambda \\ 1, & \text{if } \beta u'(1) \geq \lambda \\ \{v : \beta u'(v) = \lambda\}, & \text{otherwise,} \end{cases} \quad (8)$$

and let $\Phi(\lambda) = \sum_{r=1}^{\hat{r}} \phi(\beta_{r,n},\lambda)$.

**Lemma 4** (i) $\phi(\beta,\lambda)$ is increasing in $\beta$ and decreasing in $\lambda$.

(ii) There exists a $\lambda^* > 0$ such that $\Phi(\lambda^*) = 1$.

The proof of Lemma 4 is based on the observation that correspondence $\phi(\beta,\lambda)$ is the
solution to maximization problem $\max_{v \in [0,1]} \beta u(v) - \lambda v$. The lemma gives our first major
result.

**Proposition 1** Suppose $f(\cdot)$ is log-concave, and take $\lambda^*$ from Lemma 4(ii). The following
allocation of prizes is optimal:

$$v^*_r = \phi(\beta_{r,n},\lambda^*), \quad r = 1,\ldots,\hat{r}; \quad v^*_{\hat{r}+1} = 0, \quad r = \hat{r} + 1,\ldots,n. \quad (9)$$

Indeed, it is straightforward to see that prize allocation (9) is a solution to problem (6). The
log-concavity of $f(\cdot)$ and Lemma 4(i) then imply that $v^*_r$ are decreasing in $r$ and
hence (9) is also a solution to problem (4).
As seen from Proposition 1, there are at most \( \hat{r} \) positive prizes. The exact number of positive prizes, \( s \), is such that \( v_s^* > 0 \) and \( v_{s+1}^* = 0 \), and depends on the utility function, the distribution of noise and the number of players, \( n \). Using the definition of \( \phi \), Eq. (8), we obtain \( s = \max\{r \leq \hat{r} : \beta_{r,n}u'(0) > \lambda^*\} \). Examples of optimal prize allocations for a log-concave \( f(\cdot) \) are shown in Figure 1.

### 3.2 General \( f(\cdot) \)

We now turn to the case when \( f(\cdot) \) is not necessarily log-concave and hence coefficients \( \beta_{r,n} \) can be nonmonotone. It is convenient to introduce nonnegative variables \( D_r = u(v_r) - u(v_{r+1}) \), which are differentials between the utilities of adjacent prizes. Since \( \beta_{n,n} < 0 \) (cf. Lemma 1(i)), \( v_n = 0 \) in any optimal prize schedule, and we can set \( D_{n-1} = u(v_{n-1}) \). This gives \( v_{n-1} = u^{-1}(D_{n-1}) \), and all other prizes can be restored from the differentials as \( v_r = u^{-1}(\sum_{k=r}^{n-1} D_k) \).

The budget constraint in the new variables takes the form \( \sum_{r=1}^{n-1} u^{-1}(\sum_{k=r}^{n-1} D_k) = 1 \). The objective function becomes \( \sum_{r=1}^{n-1} B_{r,n} D_r \), where coefficients \( B_{r,n} = \sum_{k=1}^{r} \beta_{k,n} \) are the
cumulative version of $\beta_{r,n}$. Problem (4) then transforms into

$$\max_{D_1,\ldots,D_{n-1}} \sum_{r=1}^{n-1} B_{r,n} D_r \quad \text{s.t.} \quad D_1,\ldots,D_{n-1} \geq 0; \quad \sum_{r=1}^{n-1} \left( \sum_{k=r}^{n-1} D_k \right) = 1. \quad (10)$$

It is straightforward to verify that

$$B_{r,n} = \frac{(n-1)!}{(n-r-1)!(r-1)!} \int_{x} F(x)^{n-r-1}[1-F(x)]^{r-1}f(x)dF(x). \quad (11)$$

Thus, $B_{r,n} > 0$ for $r \leq n-1$, and hence (10) is a convex programming problem with a linear objective and convex constraint in variables $D_1,\ldots,D_{n-1}$; therefore, Kuhn-Tucker conditions are both necessary and sufficient for maximization. In the original variables, these conditions take the form

$$B_{r,n} \leq \lambda \sum_{k=1}^{r} \frac{1}{w'(v_k)}, \quad \text{with equality if} \quad D_r > 0, \quad r = 1,\ldots,n-1, \quad (12)$$

for some Lagrange multiplier $\lambda > 0$, and subject to the constraint $\sum_{r=1}^{n-1} v_r = 1$.

Let $\tilde{\beta}_{r_1:r_2} = \frac{1}{r_2-r_1+1} \sum_{k=r_1}^{r_2} \beta_{k,n}$ denote the partial average of coefficients $\beta_{r,n}$ for $r_2 \geq r_1$.

Define a sequence of critical points $r_1^*, r_2^*, \ldots, r_K^*$, where the number of points, $K$, is at least 1 and at most $n-1$, as follows:

(i) Define $r_1^* = \max\{1 \leq r \leq n-1 : \tilde{\beta}_{1:r} \geq \tilde{\beta}_{1:k} \forall k = 1,\ldots,r-1\}$. In words, $r_1^*$ is the point where $\tilde{\beta}_{1:r}$ reaches its global maximum for $r = 1,\ldots,n-1$. If there is more than one maximum point, set $r_1^*$ to the largest such point. If $r_1^* = n-1$, we are done; otherwise, proceed to step (ii).

(ii) Define $r_2^* = \max\{r_1^*+1 \leq r \leq n-1 : \tilde{\beta}_{r_1+1:r} \geq \tilde{\beta}_{r_1+1:k} \forall k = r_1^*+1,\ldots,r-1\}$. In words, $r_2^*$ is the point where $\tilde{\beta}_{r_1+1:r}$ reaches its global maximum for $r = r_1^*+1,\ldots,n-1$. If there is more than one maximum point, set $r_2^*$ to the largest such point. If $r_2^* = n-1$ or $\tilde{\beta}_{r_1+1:r_2^*} \leq 0$, we are done; otherwise, proceed to step (iii).

(iii) Continue similarly until $r = n-1$ or $\tilde{\beta}_{r_{K-1}+1:r_K^*} \leq 0$ is reached; set $r_K^*$ to $n-1$ or the largest $k$ such that $\tilde{\beta}_{r_{K-1}+1:k} > 0$, whichever is reached first.

Using these critical points we can now characterize the optimal allocation of prizes.
Proposition 2 Let $r_1^*, r_2^*, \ldots, r_K^*$ denote the sequence of critical points defined above. Set $r_0^* = 0$ and let $\lambda^*$ denote a solution of the equation

$$\sum_{k=1}^{K} (r_k^* - r_{k-1}^*) \phi(\beta_{r_k^*+1:r_k^*}, \lambda) = 1. \tag{13}$$

The following allocation of prizes is optimal:

$$v_1^* = \ldots = v_1^* = \phi(\beta_{1:r_1^*}, \lambda^*),$$

$$v_{r_1^*+1}^* = \ldots = v_{r_1^*}^* = \phi(\beta_{r_1^*+1:r_2^*}, \lambda^*),$$

$$\ldots$$

$$v_{r_{K-1}^*+1}^* = \ldots = v_{r_{K-1}^*}^* = \phi(\beta_{r_{K-1}^*+1:r_K^*}, \lambda^*),$$

$$v_{r_K^*+1}^* = \ldots = v_n^* = 0. \tag{14}$$

As seen from (14), optimal prizes have a step-wise structure. There is some number $s \leq K$ of distinct strictly decreasing positive prizes at the critical points $r_1^*, \ldots, r_s^*$. Prizes between the critical points are flat, and the rest of the prizes are zero: $v_r = 0$ for all $r > r_s^*$. The number of distinct positive prizes is determined by conditions $v_{r_1^*} > 0$ and $v_{r_{s+1}^*} = 0$, i.e.,

$$s = \max\{k \leq K : \beta_{r_k^*+1:r_k^*} u'(0) > \lambda^*\}. \tag{15}$$

Figures 2 and 3 provide examples of optimal prize allocations in cases when $f(\cdot)$ is not log-concave.

Unimodal $f(\cdot)$. Part (v) of Lemma 1 implies that when $f(\cdot)$ is unimodal, there exists a rank $\hat{r} \leq n - 1$ such that it is always optimal to assign $v_r = 0$ for $r > \hat{r}$. This property is similar to the one identified for a log-concave $f(\cdot)$, except coefficients $\beta_{r,n}$ are not necessarily decreasing in $r$, and hence the optimal prize schedule may contain “steps.” However, importantly, there is a limit to the number of positive prizes regardless of the degree of risk-aversion.

Corollary 2 When $f(\cdot)$ is unimodal, the number of optimal positive prizes is restricted by the $\hat{r}$ defined in Lemma 1(v) for any $u(\cdot)$. 
Figure 2: Optimal prize allocation for Burr distribution with parameters 3 and 1 (its pdf is \( f(x) = \frac{3x^2}{(x^3 + 1)^2} \) for positive \( x \)) and \( n = 10 \) (which gives \( \hat{r} = 6 \)). Left: CARA utility function with \( \alpha = 1 \) (blue circles) and \( \alpha = 3 \) (red diamonds). Right: CRRA utility function with \( \rho = 0.25 \) (blue circles) and \( \rho = 0.75 \) (red diamonds).

Figure 3: Left: The pdf \( f(x) \) of a distribution with cdf \( F(x) = 0.2 \tan(2x) + 0.7 \) defined on \([-0.646, 0.491]\). Center: \( \beta_{r,n} \) for \( n = 10 \). Right: Optimal prizes for CRRA utility function with \( \rho = 0.25 \).
3.3 Heavy-tailed distributions and the role of \( r_1^* \)

The first critical point in the optimal allocation of prizes, \( r_1^* \), is defined as the largest \( r \) that maximizes \( \bar{\beta}_{1,r} = \frac{1}{r} \sum_{k=1}^{r} \beta_{r,n} \) – the running average of coefficients \( \beta_{r,n} \). As shown by Drugov and Ryvkin (2018), this running average can be written in the form

\[
\bar{\beta}_{1,r} = \frac{1}{n} E(h(X_{n-r:n})). \tag{16}
\]

Here, \( h(x) = \frac{f(x)}{1-F(x)} \) is the failure (or hazard) rate of noise, and \( X_{n-r:n} \) is its order statistic. The role of the failure rate can be understood intuitively from the following arguments. Eq. (16) can be rewritten as \( \bar{\beta}_{1,r} = \frac{1}{n} \int_X f(x \mid X \geq x) f_{(n-r:n)}(x) dx \), where the failure rate \( \frac{f(x)}{1-F(x)} \) is written as the density at \( x \) of variable \( X \) conditional on \( X \geq x \). Thus, \( \bar{\beta}_{1,r} \) is determined by the density at zero of the difference between \( X \) and \( X_{(n-r:n)} \) conditional on \( X \geq X_{(n-r:n)} \). Indeed, the probability of reaching a rank of at least \( r \) can be expressed as the probability of surpassing the \( r \)-th highest noise realizations out of \( n \) conditional on \( X \) being among the top \( r \) realizations, multiplied by the probability that \( X \) is in the top \( r \) (equal \( \frac{r}{n} \)).

Representation (16) together with Proposition 2 immediately imply \( r_1^* = n-1 \) if noise has a decreasing failure rate (DFR). This leads to maximum prize sharing.

**Corollary 3** If \( f(\cdot) \) is DFR, then the following allocation of prizes is optimal:

\[
v_r^* = \frac{1}{n-1}, \quad r = 1, \ldots, n-1; \quad v_n^* = 0. \tag{17}
\]

Allocation (17) can be characterized as the “extreme punishment” tournament. It is the polar opposite of WTA in that it punishes the worst-performing agent instead of rewarding the top performer. This allocation is optimal for DFR distributions even when agents are risk-neutral (Drugov and Ryvkin, 2018). Therefore, Corollary 3 is consistent with the expectation that, generally, there needs to be more prize sharing as agents become risk-averse.

Log-convex distributions with \( f(\bar{x}) = 0 \) are DFR (although the converse is not true), and hence allocation (17) is optimal for them as well. From Lemma 1(iii), \( \beta_{r,n} \) increases in \( r \) for \( r \leq n-1 \) in this case; therefore, the optimality of allocation (17) is rather intuitive. Indeed, any feasible prize schedule with unequal prizes can be improved upon by setting two adjacent unequal prizes to their average. However, \( \beta_{r,n} \) does not need to be increasing for \( \bar{\beta}_{1,r} \) to be increasing, and allocation (17) can be optimal even when \( \beta_{r,n} \)
is nonmonotone.

DFR distributions are often characterized as having heavy tails, the most prominent example being power laws (Gabaix, 2016). Heavy-tailed shocks can be encountered in many environments. Examples include fluctuations in financial markets, demand for creative products, or returns to innovation. The optimality of maximum prize sharing in these environments is rather striking given that the existing institutions of high-powered rewards and patent protection in these settings are, if anything, at the opposite end of the spectrum.

A more general class of distributions that can be characterized as heavy-tailed are those having a unimodal failure rate (first IFR, then DFR). In this case, Drugov and Ryvkin (2018) show that \( \bar{\beta}_{1,r} \) is unimodal and \( r_1^* > 1 \) for a sufficiently large \( n \). That is, at least some prize sharing at the top is optimal in sufficiently large tournaments.

Prize structure (14) simplifies somewhat when \( \beta_{r,n} \) is unimodal. However, as noted above, the modality of \( \bar{\beta}_{1,r} \) does not directly translate into the modality of \( \beta_{r,n} \). The latter is ensured by a stronger condition than a unimodal failure rate, namely, that \( f(\cdot) \) is first log-concave, then log-convex with \( f(\bar{x}) = 0 \) (call such a distribution FLTL), see Lemma 1(iv).

**Corollary 4** If \( f(\cdot) \) is FLTL, then the following allocation of prizes is optimal:

\[
\begin{align*}
v_1^* &= \ldots = v_{r_1^*}^* = \phi(\bar{\beta}_{1,r_1^*}, \lambda^*), \\
v_r^* &= \phi(\beta_{r,n}, \lambda^*), \quad r = r_1^* + 1, \ldots, \hat{r}, \\
v_r^* &= 0, \quad r = \hat{r} + 1, \ldots, n.
\end{align*}
\]

Indeed, denoting \( r_m \) the largest \( r \) such that \( \beta_{r,n} \geq \beta_{r-1,n} \), it is clear that \( r_1^* \geq r_m \) – the maximum of \( \bar{\beta}_{1,r} \) is to the right of the mode of \( \beta_{r,n} \) – implying that \( \beta_{r,n} \) is decreasing for \( r > r_1^* \). Therefore, a prize schedule that is flat for \( r \leq r_1^* \) and decreasing for \( r > r_1^* \) is optimal. Examples of FLTL distributions include the log-normal distribution, the Burr distribution, and F- and Beta-distributions for some parameters (see, e.g., Bagnoli and Bergstrom, 2005). Figure 2 shows an example with the Burr distribution.

**Unimodal** \( f(\cdot) \). When \( f(\cdot) \) is unimodal, Lemma 1(v) implies that there is a cutoff rank \( \hat{r} \leq n - 1 \) such that \( v_r^* = 0 \) is optimal for \( r > \hat{r} \). In this case, the FLTL condition in Corollary 4 can be relaxed because \( \beta_{r,n} \) only needs to be unimodal for \( r \leq \hat{r} \). For example, some symmetric heavy-tailed distributions, such as the \( t \)-distribution, are log-concave in
an interval around zero and log-convex otherwise. It can then be shown based on Lemma 1 that the upper half of $\beta_{r,n}$ are positive and their sequence is unimodal. In this case, prize structure (18) is still optimal.

4 The effect of risk-aversion

As discussed previously, it is broadly expected that the more risk-averse the agents are the more prize sharing is going to be optimal. In this section, we provide a precise formulation of this statement. We use the standard definition of “more risk-averse” through the ranking of the absolute (or relative) risk aversion coefficient, which is equivalent to the concave transform order for utility functions (Pratt, 1964). Thus, for two utility functions $u, \tilde{u} : [0, 1] \to \mathbb{R}$, $\tilde{u}(\cdot)$ is more risk-averse than $u(\cdot)$ if $\tilde{u} \circ u^{-1}$ is concave. An equivalent definition of this order is that $u'(v) \tilde{u}'(v)$ is increasing in $v$.

When agents are risk-neutral, optimal prize schedules assign equal prizes to the top $r^*_1$ ranks and zero prizes to ranks $r = r^*_1 + 1, \ldots, n$, where $r^*_1 \in \arg\max_{r=1,\ldots,n-1} \tilde{\beta}_{1,r}$ is determined by the properties of noise, cf. Section 3.2 (see also Drugov and Ryvkin, 2018). Such prize schedules are naturally ordered in prize sharing by the value of $r^*_1$. When agents are risk-averse, both the number and the relative values of the prizes may change. The more general concept of prize sharing can then be captured by the majorization order (Marshall, Olkin and Arnold, 2011).

Definition 1 For two vectors $v, \tilde{v} \in \mathbb{R}^n_+$ whose components are arranged in descending order, $v$ majorizes $\tilde{v}$ if $\sum_{k=1}^r v_r \geq \sum_{k=1}^r \tilde{v}_r$ for all $r = 1, \ldots, n$.

Components of a prize schedule $v$ such that the budget constraint holds can be interpreted as a probability mass function (pmf) of a discrete random variable taking values $1, \ldots, n$. Definition 1 then produces the inequality between the corresponding cumulative mass functions (cmfs) stating that $\tilde{v}$ is larger than $v$ in the FOSD sense, i.e., the mass in $\tilde{v}$ is shifted to the right relative to $v$. It is, therefore, natural to interpret $\tilde{v}$ as involving more prize sharing than $v$ if $v$ majorizes $\tilde{v}$. For two-prize schedules with $r^*_1$ equal prizes, $v$ majorizing $\tilde{v}$ implies $r^*_1$ is larger under $\tilde{v}$. The following proposition relates ordering by risk-aversion to prize sharing more generally.

9When $u(\cdot)$ and $\tilde{u}(\cdot)$ have a common range $[0, 1]$, they can be interpreted as cdfs of random variables. The “more risk-averse” order then can be interpreted as the reverse likelihood ratio order (Shaked and Shanthikumar, 2007).
Proposition 3  Consider two utility functions $u, \tilde{u} : [0, 1] \rightarrow [0, 1]$ and let $v$ and $\tilde{v}$ denote the corresponding optimal prize schedules. If $\tilde{u}(\cdot)$ is more risk-averse than $u(\cdot)$, then (i) $\tilde{v}$ has more positive prizes than $v$; (ii) $v$ majorizes $\tilde{v}$ (i.e, $\tilde{v}$ involves more prize sharing than $v$).

Proposition 3 is illustrated in Figures 1 and 2, where it is shown how optimal prizes change with risk-aversion for different distributions of noise and utility functions.

5 Examples

In this section, we provide several examples illustrating the optimal allocation of prizes and how it is affected by changes in risk-aversion. Sections 5.1 and 5.2 provide optimal prize allocations in closed form for constant absolute risk-aversion (CARA) and constant relative risk-aversion (CRRA) utility functions, respectively, when the distribution of noise is log-concave. Examples with other distributions follow.

5.1 CARA utility, log-concave $f(\cdot)$.

Suppose $f(\cdot)$ is log-concave in which case $\beta_{r,n}$ is decreasing, see Lemma 1(ii). Consider a CARA utility function $u(v) = \frac{1 - \exp(-\alpha v)}{1 - \exp(-\alpha)}$, where $\alpha > 0$ is the constant absolute risk aversion parameter. This gives $u'(v) = \frac{\beta \alpha \exp(-\alpha v)}{1 - \exp(-\alpha)}$ and the correspondence $\phi$, Eq. (8),

$$
\phi(\beta, \lambda) = \begin{cases} 
0, & \text{if } \frac{\beta \alpha}{1 - \exp(-\alpha)} \leq \lambda \\
1, & \text{if } \frac{\beta \alpha \exp(-\alpha)}{1 - \exp(-\alpha)} \geq \lambda \\
\frac{1}{\alpha} \ln \frac{\beta \alpha}{\lambda \exp(-\alpha)}, & \text{otherwise}
\end{cases}
$$

The optimal allocation of prizes is given by Proposition 1. The equation for $\lambda^*$ is

$$
\Phi(\lambda) = \frac{1}{\alpha} \sum_{r=1}^{s} \ln \frac{\beta_{r,n}\alpha}{\lambda(1 - \exp(-\alpha))} = 1,
$$

where $s \leq \hat{s}$ is the optimal number of positive prizes (to be determined below). Then,

$$
\lambda^* = \frac{\alpha}{1 - \exp(-\alpha)} \left[ \prod_{r=1}^{\hat{s}} \frac{\beta_{r,n}}{\exp(\alpha)} \right]^{\frac{1}{\hat{s}}}.
$$
and finally
\[ v_r^* = \frac{1}{s} + \frac{1}{\alpha} \ln \frac{\beta_{r,n}}{\left(\prod_{k=1}^{s} \beta_{k,n}\right)^{\frac{1}{s}}}, \quad r = 1, \ldots, s; \quad v_r^* = 0, \quad r = s + 1, \ldots, n. \]

The expression in parentheses is the geometric mean of coefficients \( \beta_{k,n} \). Thus, \( v_r \) is above (below) \( \frac{1}{s} \) if \( \beta_{r,n} \) is above (below) this geometric mean. The number of positive prizes, \( s \), is defined as \( s = \max \left\{ s' : \frac{1}{s'} + \frac{1}{\alpha} \ln \frac{\beta_{s',n}}{\left(\prod_{k=1}^{s'} \beta_{k,n}\right)^{\frac{1}{s'}}} > 0 \right\} \).

By definition, a higher \( \alpha \) means more risk aversion, and Proposition 3 applies. See Figure 1(left) for an example with Gumbel distribution (equivalent to the Tullock contest, see Section 6), which is log-concave.

### 5.2 CRRA utility, log-concave \( f(\cdot) \).

Consider now utility function \( u(v) = v^{1-\rho} \), where \( \rho \in (0, 1) \) is the agents’ constant relative risk aversion parameter. Correspondence \( \phi \) is given by

\[ \phi(\beta, \lambda) = \begin{cases} 
1, & \text{if } \beta(1-\rho) \geq \lambda \\
\left(\frac{\beta(1-\rho)}{\lambda}\right)^\frac{1}{\rho}, & \text{otherwise}
\end{cases} \]

The equation for \( \lambda^* \),

\[ \Phi(\lambda) = \sum_{r=1}^{\hat{r}} \phi(\beta_{r,n}, \lambda) = \left(\frac{1-\rho}{\lambda}\right)^\frac{1}{\rho} \sum_{r=1}^{\hat{r}} \beta_{r,n}^{\frac{1}{\rho}} = 1, \]

produces

\[ \lambda^* = (1-\rho) \left( \sum_{r=1}^{\hat{r}} \beta_{r,n}^{\frac{1}{\rho}} \right)^{\rho}. \]

The optimal allocation of prizes is, therefore,

\[ v_r^* = \phi(\beta_{r,n}, \lambda^*) = \frac{\beta_{r,n}^{\frac{1}{\rho}}}{\sum_{k=1}^{\hat{r}} \beta_{k,n}^{\frac{1}{\rho}}}, \quad r = 1, \ldots, \hat{r}; \quad v_r^* = 0, \quad r = \hat{r} + 1, \ldots, n. \]
The maximum number of positive prizes, \( s = \hat{r} \), is optimal in this case because \( u'(0) = +\infty \). The resulting prizes can also be rewritten as

\[
v^*_r = \frac{1}{\hat{r}} \left[ \frac{\beta_{r,n}}{\left( \frac{1}{\hat{r}} \sum_{k=1}^{\hat{r}} \beta_{k,n}^{\frac{1}{\rho}} \right)^{\frac{1}{\rho}}} \right]^{\frac{1}{\rho}},
\]

where the expression in parentheses is the generalized mean with exponent \( \frac{1}{\rho} \). Thus, prizes are above (below) \( \frac{1}{\hat{r}} \) if \( \beta_{r,n} \) is above (below) the generalized mean of coefficients \( \beta_{k,n} \).

The coefficient of absolute risk aversion is equal to \( \xi \). Hence, a higher \( \rho \) means a higher absolute risk aversion, and Proposition 3 applies. See Figure 1(right) for an example with Gumbel distribution (which is equivalent to the Tullock contest, see Section 6) which is log-concave.

### 5.3 Other distributions

Figure 2 presents an example with Burr distribution, which is FLTL, and CARA and CRRA utility functions with different degrees of risk aversion.

Finally, Figure 3 illustrates the optimal prize allocation for a log-convex distribution with \( f(\bar{x}) > 0 \) in which case Lemma 1(iii) does not apply.

### 6 The Winner-Take-All principle revisited

Krishna and Morgan (1998) obtained the “WTA principle” for small tournaments, whereby the WTA prize schedule is optimal for tournaments with \( n \leq 3 \) risk-averse agents when the distribution of noise is unimodal and symmetric. For risk-neutral agents, Krishna and Morgan (1998) proved the WTA principle under similar conditions for \( n \leq 4 \), and Drugov and Ryvkin (2018) generalized it to arbitrary \( n \) for IFR distributions.

We can use Proposition 2 to obtain a sufficient condition for the optimality of WTA for an arbitrary number of risk-averse agents beyond the case of increasing \( f(\cdot) \) in Lemma 3. In order for WTA to be optimal, it must be that \( v^*_1 = 1 \) and \( v^*_2 = 0 \) in the optimal prize schedule (14).

**Corollary 5** Suppose (i) \( r^*_1 = 1 \) and (ii) \( \beta_{1,n}u'(1) \geq \bar{\beta}_{2,r^*_2}u'(0) \). Then the WTA tournament is optimal.
First of all, an increasing \( f(\cdot) \) leads to \( \beta_{1,n} > 0 > \beta_{r,n} \) for \( r \geq 2 \) (see the proof of Lemma 3) which immediately implies both conditions (i) and (ii) of Corollary 5. In general, when \( r^*_1 = 1 \) we have \( v^*_1 = v^*_2 \) and \( \beta_{1,n} \geq \beta_{1:r} \) for all \( r = 1, \ldots, \hat{r} \). It remains to show that \( v^*_2 = v^*_{r^*_2} = 0 \) is optimal. It will hold if \( \beta_{2,r^*_2} u'(0) \leq \lambda^* \), where \( \lambda^* = \beta_{1,n} u'(1) \), which gives the result.

Note that if \( r^*_1 = 1 \) then \( \beta_{1,n} \geq \frac{1}{r} \sum_{k=1}^{r} \beta_{k,n} \) for all \( r \), implying \( (r-1)\beta_{1,n} \geq \sum_{k=2}^{r} \beta_{k,n} \) and hence \( \beta_{1,n} \geq \beta_{2:r} \) for all \( r \). In particular, \( \beta_{1,n} \geq \beta_{2:r^*_2} \). Therefore, for risk-neutral agents condition (ii) is implied by condition (i), and the result of Drugov and Ryvkin (2018) follows as a special case. For risk-averse agents, condition (i) alone is not sufficient; for the optimality of WTA, coefficient \( \beta_{1,n} \) must also be large enough to overcome the decreasing marginal utility of money. A sufficient condition for (i) is that \( f(\cdot) \) is IFR.

Condition (ii), however, combines the properties of noise and utility function \( u(\cdot) \).

**Tullock contests.** Consider the implications of our results for the standard Tullock contest. As shown by Fu and Lu (2012), the equilibrium in nested, multi-prize Tullock contests with discriminatory power \( \lambda \) (Clark and Riis, 1996) is equivalent to that of a rank-order tournament with additive noise with the Gumbel distribution with parameter \( \lambda \). Its pdf \( f(x) = \lambda \exp[-\lambda x - \exp(-\lambda x)] \) is log-concave. The corresponding coefficients \( \beta_{r,n} \) are, therefore, decreasing in \( r \) and hence \( r^*_1 = 1 \) and \( r^*_2 = 2 \) (provided \( n \geq 3 \) so that \( \hat{r} \geq 2 \); otherwise, WTA is optimal). This implies \( \beta_{2,r^*_2} = \beta_{2,n} \) and condition (ii) of Corollary 5 becomes simply \( \beta_{1,n} u'(1) \geq \beta_{2,n} u'(0) \).

Coefficients \( \beta_{r,n} \) can be written in a closed form as

\[
\beta_{r,n} = \frac{\lambda}{n} \left[ 1 + \frac{r}{n(n-r)} + \sum_{k=0}^{r-1} \frac{1}{n-r+k} \right],
\]

which implies \( \beta_{1,n} = \frac{\lambda(n-1)}{n^2} \) and \( \beta_{2,n} = \frac{\lambda(n^2-3n+1)}{n^2(n-1)} \). This produces a sufficient condition for the optimality of WTA in Tullock contests:

\[
\frac{u'(1)}{u'(0)} \geq 1 - \frac{n}{(n-1)^2}.
\]

The left-hand side of this condition is strictly below one for strictly risk-averse agents. Thus, for such agents a departure from WTA necessarily becomes optimal for a sufficiently large \( n \). This result (and more generally, the optimal prize allocation in the Tullock contest) is independent of \( \lambda \), which follows from the invariance of optimal prize allocations
to scaling transformations of noise (Lemma 2) because $\lambda$ is the scaling parameter of the Gumbel distribution.

7 Conclusions

In this paper we characterize the optimal allocation of prizes in rank-order tournaments with symmetric, risk-averse agents. One of the key questions in the literature on competitive incentives is whether, and under what conditions, the winner-take-all (WTA) prize schedule, i.e., awarding the entire prize budget to a single winner, is optimal. Advocates of WTA argue that it creates the strongest incentives at the top of the rankings and motivates workers the most. In the opposite camp are the proponents of prize sharing, often citing concerns about inequality, worker morale and perverse incentives high-powered pay schemes may create. The debate typically revolves around nonstandard preferences, such as relative payoffs or status concerns.

There are, however, at least two reasons for why WTA may not always be optimal even under standard preferences. First, as shown by Drugov and Ryvkin (2018), the optimal allocation of prizes depends critically on the details of the distribution of noise. Specifically, the presence of heavy tails in the distribution of noise makes prize sharing optimal. This result holds for risk-neutral, selfish, payoff-maximizing agents. The second, independent factor is agents’ risk-aversion, which may lead to the optimality of prize sharing even when WTA is optimal for risk-neutral agents. Effort provision in tournaments is risky because it has a certain cost and an uncertain benefit, and prize sharing can encourage effort because it reduces the risk.

We show how the optimal structure of prizes is determined by a combination of noise and risk-aversion. We derive a sufficient condition for the optimality of WTA that generalized the existing winner-take-all principles of Krishna and Morgan (1998) for small tournaments and of Schweinzer and Segev (2012) for Tullock contests. When the distribution of noise is Pareto, the extreme punishment tournament, i.e., a scheme awarding the same prize to all agents except the one ranked last, is optimal regardless of risk aversion. Moreover, we show that an increase in risk-aversion, in the sense of Pratt, leads to an increase in prize sharing in the optimal reward scheme in the sense of the majorization order.
References


Appendix

Proof of Lemma 1 Part (i) follows directly from Eq. (3).

Let $F^{-1}(z) = \inf \{ x \in \mathcal{X} : F(x) \geq z \}$ denote the quantile function of noise, and let $m(z) = f(F^{-1}(z)) : [0,1] \rightarrow \mathbb{R}_+$ denote the inverse quantile density (Parzen, 1979), which is continuous, piece-wise differentiable and integrable due to the properties of $f(\cdot)$. After the probability integral change of variable, $z = F(x)$, Eq. (3) becomes

$$
\beta_{r,n} = \binom{n-1}{r-1} \int_0^1 z^{n-r-1}(1-z)^{r-2}[n-r-(n-1)z]m(z)dz.
$$

Integrating Eq. (19) by parts, obtain

$$
\beta_{r,n} = \binom{n-1}{r-1} \int_0^1 m(z)z^{n-r-1}(1-z)^{r-2}[n-r-(n-1)z]dz
$$

$$
= \binom{n-1}{r-1} \int_0^1 m(z)d[z^{n-r}(1-z)^{r-1}]
$$

$$
= \binom{n-1}{r-1} \left[ m(z)z^{n-r}(1-z)^{r-1} \right]_0^1 - \int_0^1 z^{n-r}(1-z)^{r-1}m'(z)dz
$$

$$
= \binom{n-1}{r-1} \left[ m(1)\mathbb{I}_{r=1} - m(0)\mathbb{I}_{r=n} - \frac{1}{n(n-r)!}(r-1)! \int_0^1 z^{n-r}(1-z)^{r-1}m'(z)dz \right]
$$

$$
= m(1)\mathbb{I}_{r=1} - m(0)\mathbb{I}_{r=n} - \frac{1}{n}E(m'(Z_{n+1-r,n})).
$$

Here, $Z_{n+1-r,n}$ are order statistics of the uniform distribution on $[0,1]$. These order statistics are FOSD-decreasing in $r$.

For part (ii), note that if $f(\cdot)$ is log-concave then $m(\cdot)$ is concave, and hence $m'(z)$ is decreasing and the expectation is increasing in $r$. The first two terms in the expression above give $m(1)$ for $r = 1$, $-m(0)$ for $r = n$ and 0 otherwise; hence, combined we have a sequence that is decreasing in $r$.

For part (iii), note that if $f(\cdot)$ is log-convex then $m(\cdot)$ is convex, and $f(\overline{x}) = 0$ implies $m(1) = 0$. Eq. (20) then gives a sequence that is increasing in $r$ for $r \leq n - 1$.

We first prove part (v), and then go back to part (iv). For part (v), note that if $f(\cdot)$ is unimodal, then $m(\cdot)$ is also unimodal and hence $m'(\cdot)$ is single-crossing; that is, there exists a $\hat{z} \in [0,1]$ such that $m'(z) \leq 0$ for $z \leq \hat{z}$ and $m'(z) \geq 0$ for $z \geq \hat{z}$. The cases of monotone $m'(\cdot)$ are covered in parts (ii) and (iii). Suppose $m'(\cdot)$ is nonmonotone.

We know from part (i) that $\beta_{1,n} > 0$ and $\beta_{n,n} < 0$; therefore, for $n \leq 3$ the result is trivial (and does not require unimodality). Suppose $n \geq 4$, and consider some $r$ such that
3 \leq r \leq n - 1 \text{ and } \beta_{r,n} > 0. \text{ It is sufficient to show that } \beta_{r-1,n} > 0. \text{ From (20), }

\begin{align*}
\beta_{r-1,n} &= -\frac{(n - 1)!}{(n - r)!(r - 1)!} \int_0^1 z^{n-1-r}(1-z)^{r-2}m'(z)dz \\
&= -\frac{r-1}{n+1-r} \frac{(n - 1)!}{(n - r+1)!(r - 1)!} \int_0^1 \frac{z^{n-1-r}(1-z)^{r-1}m'(z)}{1-z}dz \\
&= -\frac{r-1}{n+1-r} \frac{(n - 1)!}{(n - r+1)!(r - 1)!} \left[ \int_0^1 \frac{z^{n-1-r}(1-z)^{r-1}m'(z)}{1-z}dz + \hat{r} \int_0^1 \frac{z^{n-1-r}(1-z)^{r-1}m'(z)}{1-z}dz \right] \\
&= -\frac{r-1}{n+1-r} \frac{(n - 1)!}{(n - r+1)!(r - 1)!} \frac{\hat{r}}{1-\hat{r}} \int_0^1 \frac{z^{n-1-r}(1-z)^{r-1}m'(z)}{1-z}dz = \frac{r-1}{n+1-r} \frac{\hat{r}}{1-\hat{r}} \beta_{r,n} > 0.
\end{align*}

The inequality on the fourth line follows because \( \frac{\hat{r}}{1-\hat{r}} \) is positive and increasing and \( m'(z) \) is positive (negative) in the first (second) integral.

For part (iv), it follows that \( m(\cdot) \) is first concave, then convex, which implies \( m'(\cdot) \) is U-shaped and hence \( m''(\cdot) \) is single-crossing. In order to show that \( \beta_{r,n} \) is unimodal, we will show that \( \beta_{r,n} - \beta_{r-1,n} \) is single-crossing. Assume \( n-1 \geq r \geq 2 \) and recall that \( m(1) = 0 \); using (20),

\begin{align*}
\beta_{r,n} - \beta_{r-1,n} &= -\frac{(n - 1)!}{(n - r)!(r - 1)!} \int_0^1 z^{n-1-r}(1-z)^{r-1}m'(z)dz \\
&+ \frac{(n - 1)!}{(n - r+1)!(r - 2)!} \int_0^1 z^{n-1-r}(1-z)^{r-2}m'(z)dz \\
&= -\frac{(n - 1)!}{(n - r+1)!(r - 1)!} \int_0^1 z^{n-1-r}(1-z)^{r-2}[(n - r+1)(1-z) - (r - 1)z]m'(z)dz \\
&= -\frac{(n - 1)!}{(n - r+1)!(r - 1)!} \int_0^1 m'(z)d[z^{n-1-r+1}(1-z)^{r-1}] \\
&= \frac{(n - 1)!}{(n - r+1)!(r - 1)!} \int_0^1 z^{n-1-r+1}(1-z)^{r-1}m''(z)dz,
\end{align*}

where the last line follows from integration by parts. The result then follows from the steps similar to the ones in the proof of part (v).

For part (vi), it is sufficient to show that \( \beta_{r,n} > 0 \) implies \( \beta_{r,n+1} > 0 \). From (20),

\[ ... \]
assuming $\beta_{r,n} > 0$ and following the steps similar to ones in the proof of part (v), obtain
\[
\beta_{r,n+1} = -\left(\frac{n}{r-1}\right) \int_0^1 z^{n+1-r}(1-z)^{r-1}m'(z)dz
\]
\[
= -\frac{n}{n+1-r} \left(\frac{n-1}{r-1}\right) \int_0^1 z \cdot z^{n-r}(1-z)^{r-1}m'(z)dz \geq \frac{n}{n+1-r} \hat{z} \beta_{r,n} > 0.
\]

**Proof of Lemma 2** The pdfs and cdfs of random variables $X$ and $Y$ such that $Y = sX$ for some constant $s > 0$ are related by
\[
f_Y(x) = \frac{1}{s} f_X \left(\frac{x}{s}\right), \quad F_Y(x) = F_X \left(\frac{x}{s}\right).
\]
Changing the variable of integration in (3), obtain $\beta_{r,n,Y} = \frac{1}{s} \beta_{r,n,X}$. ■

**Proof of Lemma 3** We need to show that $\beta_{r,n} \leq 0$ for $r \geq 2$. From (3), coefficients $\beta_{r,n}$ can be written as
\[
\beta_{r,n} = \frac{(n-1)!}{(n-1-r)!(r-1)!} \int_X F(x)^{n-r-1}[1-F(x)]^{r-1}f(x)dF(x)
\]
\[
- \frac{(n-1)!}{(n-r)!(r-2)!} \int_X F(x)^{n-r}[1-F(x)]^{r-2}f(x)dF(x)
\]
\[
= E(f(X_{n-r:n-1})) - E(f(X_{n-r+1:n-1})),
\]
Order statistics $X_{n-r:n-1}$ are FOSD-decreasing in $r$, and the result follows immediately because $f(\cdot)$ is an increasing function. ■

**Proof of Lemma 4** Notice that $\phi(\beta, \lambda)$ is the solution to maximization problem $\max_{v \in [0,1]} \beta u(v) - \lambda v$. For part (a), note that the objective function $\beta u(v) - \lambda v$ has strictly increasing differences in $(\beta, v)$ and in $(-\lambda, v)$, and the result follows from the monotone comparative statics (Milgrom and Shannon, 1994).

For part (b), note that correspondence $\phi(\beta, \lambda)$ is upper hemi-continuous, and, therefore, so is $\Phi(\lambda)$. Let $\beta_{\min} = \min_{r=1,\ldots,\hat{r}} \beta_{r,n}$ and $\beta_{\max} = \max_{r=1,\ldots,\hat{r}} \beta_{r,n}$. Then
\[
\Phi(\beta_{\max} u'(0)) = \sum_{r=1}^{\hat{r}} \phi(\beta_{r,n}, \beta_{\max} u'(0)) = 0,
\]
\[
\Phi(\beta_{\min} u'(1)) = \sum_{r=1}^{\hat{r}} \phi(\beta_{r,n}, \beta_{\min} u'(1)) = \hat{r} \geq 1.
\]
It then follows that there exists a $\lambda^* > \beta_{\min} u'(1) \geq 0$ such that $\Phi(\lambda^*) = 1$. ■

27
Proof of Proposition 2 The proof is based on showing that the set of prizes given by (14) satisfies the system of Kuhn-Tucker conditions (12) with \( \lambda = \lambda^* \), where \( \lambda^* \) can be found from Eq. (13). We, therefore, start by showing that Eq. (13) has a solution. Let 

\[ \Phi(\lambda) = \sum_{k=1}^{K} (\bar{r}_{k} - r_{k}^*) \phi(\bar{r}_{k-1} + 1; \bar{r}_k^*, \lambda). \]

Then

\[ \Phi(\bar{r}_1^* u'(0)) = \sum_{k=1}^{K} (\bar{r}_{k} - r_{k}^*) \phi(\bar{r}_{k-1} + 1; \bar{r}_k^* u'(0)) = 0 \]

because, by construction, \( \bar{r}_1 \geq \bar{r}_{k-1} + 1; \bar{r}_k^* \) for all \( k \). Similarly,

\[ \Phi(\bar{r}_{K}^* u'(1)) = \sum_{k=1}^{K} (\bar{r}_{k} - r_{k}^*) \phi(\bar{r}_{k-1} + 1; \bar{r}_k^* u'(1)) = \sum_{k=1}^{K} (\bar{r}_{k} - r_{k}^*) = r_K^* \geq 1, \]

and hence a \( \lambda^* > 0 \) exists such that \( \Phi(\lambda^*) = 1 \).

From the construction of critical points it follows that prizes \( v_{r_k}^* \) in (14) are all distinct, except possibly some number of zero prizes at the end. Indeed, numbers \( \bar{r}_{k-1} + 1; \bar{r}_k^* \) are all distinct and strictly decreasing in \( k \); hence \( v_{r_k}^* \) are also distinct and strictly decreasing in \( k \) as long as they are positive. Let \( s \leq K \) denote the number of distinct positive prizes; that is, either \( s = K \) and \( v_{r_k}^* > 0 \), or \( s < K \) and \( v_{r_s}^* > 0 \) and \( v_{r_{s+1}^*} = 0 \). This implies \( s = \max\{k \leq K : \bar{r}_{k-1} + 1; \bar{r}_k^* u'(0) > \lambda^* \} \).

It follows that the Kuhn-Tucker conditions (12) hold as equalities at the first \( s \) critical points. Indeed, for \( k = 1 \) the Kuhn-Tucker condition,

\[ B_{r_1^*, n} = \lambda^* \frac{r_1^*}{u'(v_{r_1}^*)}, \]

holds for \( v_{r_1}^* \) defined in (14). For \( k = 2 \), the Kuhn-Tucker condition,

\[ B_{r_2^*, n} = \lambda^* \left[ \frac{r_1^*}{u'(v_{r_1}^*)} + \frac{r_2^* - r_1^*}{u'(v_{r_2}^*)} \right], \]

holds for \( v_{r_1}^* \) and \( v_{r_2}^* \) defined in (14), etc.

This implies all Kuhn-Tucker conditions with \( r \leq r_s^* \) hold. Indeed, consider some \( k \leq s \) and \( r \in \{r_{k-1}^* + 1, \ldots, r_k^* - 1\} \). The corresponding Kuhn-Tucker condition,

\[ B_{r, n} = \lambda^* \left[ \frac{r_1^*}{u'(v_{r_1}^*)} + \frac{r_2^* - r_1^*}{u'(v_{r_2}^*)} + \ldots + \frac{r - r_{k-1}^*}{u'(v_{r_k}^*)} \right], \]
is satisfied because \( \beta_{r_{k-1}+1,n} + \ldots + \beta_{r,n} \leq (r - r_{k-1}^*) \beta_{r_{k-1}+1,r_k^*} \) from the construction of \( r_k^* \).

Finally, consider Kuhn-Tucker conditions with \( r \in \{r_s^* + 1, n - 1\} \) which must be satisfied as inequalities with zero prizes \( v_s^* = 0 \). A generic condition like this takes the form

\[
B_{r,n} \leq \lambda^* \left[ \frac{r_1^*}{u'(v_1^*)} + \frac{r_2^* - r_1^*}{u'(v_2^*)} + \ldots + \frac{r_s^* - r_{s-1}^*}{u'(v_s^*)} + \frac{r - r_s^*}{u'(0)} \right].
\]

It holds because

\[
\beta_{r_{s+1},n} + \ldots + \beta_{r,n} = (r - r_s^*) \beta_{r_{s+1},r_n} \leq (r - r_s^*) \beta_{r_{s+1},r_{s+1}}
\]

from the construction of critical points, and \( \beta_{r_{s+1},r_{s+1}} u'(0) \leq \lambda^* \) due to the definition of \( s \).

We conclude that prizes (14) satisfy the Kuhn-Tucker conditions and hence they are optimal. \( \blacksquare \)

**Proof of Proposition 3** The sequence of critical points \( r_1^*, \ldots, r_K^* \) is determined only by the distribution of noise; therefore, it is the same for both utility functions. Let \( \lambda^* \) and \( \tilde{\lambda}^* \) denote the corresponding optimal Lagrange multipliers for prize schedules \( v \) and \( \tilde{v} \).

For part (i), suppose \( s < K \) distinct positive prizes are optimal under \( \tilde{u}(\cdot) \) (if \( K \) distinct prizes are optimal, we are done since at most \( K \) prizes can be optimal under \( u(\cdot) \)), i.e., \( \tilde{v}_{r_s^*} > 0 \) and \( \tilde{v}_{r_{s+1}^*} = 0 \) (or, equivalently, \( v_{r_{s+1}^*} = 0 \)). It is sufficient to show that \( v_{r_{s+1}^*} = 0 \) (or, equivalently, \( v_{r_{s+1}^*} = 0 \)), i.e., it is impossible to have \( s + 1 \) or more distinct positive prizes under \( u(\cdot) \). Suppose this is not true and \( v_{r_{s+1}^*} > 0 \). Then \( v_{r_s^*} > 0 \) as well, and the corresponding Kuhn-Tucker conditions imply

\[
\tilde{\beta}_{r_{s-1}+1,r_s^*} u'(v_{r_s^*}) = \beta_{r_{s-1}+1,r_s^*} u'(v_{r_{s+1}^*}) = \lambda^*, \quad \tilde{\beta}_{r_{s-1}+1,r_s^*} \tilde{u}'(\tilde{v}_{r_s^*}) = \tilde{\lambda}^* \geq \beta_{r_{s-1}+1,r_{s+1}^*} \tilde{u}'(0).
\]

This gives \( \frac{u'(v_{r_{s+1}^*})}{u'(v_{r_{s+1}^*})} \leq \frac{\tilde{u}'(\tilde{v}_{r_s^*})}{\tilde{u}'(0)} \). Because \( u(\cdot) \) is a convex transform of \( \tilde{u}(\cdot) \) and \( v_{r_s^*} > v_{r_{s+1}^*} \), we have \( \frac{u'(v_{r_{s+1}^*})}{u'(v_{r_{s+1}^*})} \leq \frac{\tilde{u}'(v_{r_s^*})}{\tilde{u}'(0)} \). Combining with the previous inequality, obtain \( \frac{\tilde{u}'(v_{r_s^*})}{\tilde{u}'(v_{r_{s+1}^*})} \leq \frac{\tilde{u}'(\tilde{v}_{r_s^*})}{\tilde{u}'(0)} \). But \( v_{r_{s+1}^*} > 0 \); therefore, it must be that \( \tilde{v}_{r_s^*} \leq v_{r_s^*} \).

Applying the same arguments to prizes for ranks \( r_s^* \) and \( r_{s-1}^* \), obtain \( \frac{\tilde{u}'(v_{r_{s-1}^*})}{\tilde{u}'(v_{r_s^*})} \leq \frac{\tilde{u}'(v_{r_{s-1}^*})}{\tilde{u}'(v_{r_s^*})}, \) and hence \( \tilde{v}_{r_{s-1}^*} \leq v_{r_{s-1}^*} \). Proceeding similarly, we obtain that \( \tilde{v}_{r_k^*} \leq v_{r_k^*} \) for
all \( k = 1, \ldots, s \). This implies
\[
\begin{align*}
& r_1^* \bar{v}_{r_1^*} + \ldots + (r_s^* - r_s^*) \bar{v}_{r_s^*} \leq r_1^* v_{r_1^*} + \ldots + (r_s^* - r_s^*) v_{r_s^*} \\
& < r_1^* v_{r_1^*} + \ldots + (r_s^* - r_s^*) v_{r_s^*} + (r_{s+1}^* - r_s^*) v_{r_{s+1}^*} \leq 1,
\end{align*}
\]
giving \( \sum_{r=1}^s \bar{v}_r < 1 \), which is impossible.

For part (ii), let \( s \) and \( \tilde{s} \) denote the number of distinct positive prizes in \( v \) and \( \tilde{v} \), respectively (we showed in part (i) that \( \tilde{s} \geq s \)). We need to show that \( \sum_{k=1}^s v_r \geq \sum_{k=1}^s \bar{v}_r \) for all \( r \leq r_{s}^* \). Suppose this inequality is not satisfied for some \( r \), and let \( q \) denote the lowest \( r \) such that it does not hold. Then it must be that \( v_q < \bar{v}_q \), or, equivalently, \( v_{r_k^*} < \bar{v}_{r_k^*} \) for some \( k \leq \tilde{s} \). There are two possible cases: (a) \( k < s \) and (b) \( k \geq s \).

(a) Suppose \( k < s \) and consider prizes at ranks \( r_k^* \) and \( r_{k+1}^* \), which are all positive in \( v \) and in \( \tilde{v} \). From the Kuhn-Tucker conditions,
\[
\begin{align*}
\bar{\beta}_{r_{k-1}^*+1:r_k^*} u'(v_{r_k^*}) &= \beta_{r_{k-1}^*+1:r_k^*} u'(v_{r_k^*}) = \lambda^*, \\
\beta_{r_{k-1}^*+1:r_k^*} \tilde{u}'(\tilde{v}_{r_k^*}) &= \tilde{\beta}_{r_{k-1}^*+1:r_k^*} \tilde{u}'(\tilde{v}_{r_k^*}) = \tilde{\lambda}^*.
\end{align*}
\]
This gives
\[
\frac{u'(v_{r_k^*})}{\tilde{u}'(\tilde{v}_{r_k^*})} = \frac{\tilde{u}'(\tilde{v}_{r_k^*})}{\tilde{u}'(\tilde{v}_{r_k^*})}.
\]
Because \( u(\cdot) \) is a convex transform of \( \tilde{u}(\cdot) \) and \( v_{r_k^*} > v_{r_{k+1}^*} \), we have
\[
\frac{u'(v_{r_k^*})}{\tilde{u}'(\tilde{v}_{r_k^*})} \leq \frac{u'(v_{r_{k+1}^*})}{\tilde{u}'(\tilde{v}_{r_{k+1}^*})}.
\]
Combining with the previous inequality, obtain
\[
\frac{\tilde{u}'(\tilde{v}_{r_k^*})}{\tilde{u}'(\tilde{v}_{r_{k+1}^*})} \leq \frac{\tilde{u}'(\tilde{v}_{r_{k+1}^*})}{\tilde{u}'(\tilde{v}_{r_{k+1}^*})}.
\]
But \( v_{r_k^*} < \bar{v}_{r_k^*} \); therefore, it must be that \( v_{r_{k+1}^*} \leq \bar{v}_{r_{k+1}^*} \).

(b) Suppose \( k \geq s \). If \( \tilde{s} = s \), this is impossible (if the majorization inequality is violated for the last positive prize, then the budget constraint cannot hold); therefore, \( \tilde{s} > s \). Then \( v_{r_{k+1}^*} = 0 \) and \( v_{r_{k+1}^*} \leq \bar{v}_{r_{k+1}^*} \) holds automatically.

Thus, we obtained that a violation of the majorization inequality for some \( r \) implies that \( v_{r_k^*} < \bar{v}_{r_k^*} \) for some \( k \), which in turn implies \( v_{r_{k+1}^*} \leq \bar{v}_{r_{k+1}^*} \). Continuing the same argument, it follows that \( v_{r_l^*} \leq \bar{v}_{r_l^*} \) for all \( l \in \{k+1, \ldots, \tilde{s}\} \), and hence the majorization inequality is violated for the last positive prize, which is impossible. ■