Performance-maximizing large contests

Wojciech Olszewski
Department of Economics, Northwestern University

Ron Siegel
Department of Economics, The Pennsylvania State University

Many sales, sports, and research contests are put in place to maximize contestants' performance. We investigate and provide a complete characterization of the prize structures that achieve this objective in settings with many contestants. The contestants may be ex ante asymmetric in their abilities and prize valuations, and there may be complete or incomplete information about these parameters. The prize valuations and performance costs may be linear, concave, or convex. A main novel takeaway is that awarding numerous different prizes whose values gradually decline with contestants' ranking is optimal in the typical case of contestants with convex performance costs and concave prize valuations. This suggests that many real-world contests can be improved by increasing the number of prizes and making them more heterogeneous. The techniques we develop can also be used to formulate and solve other contest design questions that have so far proven intractable.

Keywords. Large contests, contest design, optimal contests, mechanism design.

JEL classification. D44, D47, D82.

1. Introduction

Contests are used in a variety of settings to motivate people and increase their performance. Cisco Systems, one of the largest technology companies, regularly runs contests among its thousands of partners to boost sales.[1] HubSpot, a three-billion dollar software company, and Clayton Homes, the largest builder of manufactured housing and modular homes in the United States, are two other examples.[2] In the context of entertainment, sporting contests play an important role. In the context of academia, funding

[1]A recent example with multiple prizes is the 2017 "Cisco Commercial Champs Sales Competition—Win a trip to Taipei, Taiwan." See https://www.cisco-commercialxcelerate.com/AppFiles/pdf/tnc/sc/CCX_SC_FY17Q3_TnC.pdf

[2]"The Right Way to Use Compensation," Harvard Business Review, April 2015 and "The Sales Director Who Turned Work into a Fantasy Sports Competition," Harvard Business Review, March 2015.

agencies administer large contests that motivate researchers to generate high-quality research proposals. The widespread use of contests makes contest design relevant and important for many real-world domains, but at this time many aspects of contest design are not well understood.

This paper improves our understanding of contest design by characterizing the prize structures that maximize contestants' aggregate performance in contests with many contestants. More precisely, given a prize budget, we investigate how many prizes should be awarded and how the budget should be allocated among the prizes to induce maximal performance. Should a small number of high-value prizes be awarded, or a larger number of lower-value prizes? Or perhaps awarding prizes of different values is optimal? And if so, how should the prize values change with their rank order?

Our main qualitative finding characterizes the optimal prize structures in the most relevant case of players with convex performance costs and concave prize valuations. In this case, the performance-maximizing contest awards many prizes, all of them different, with gradually decreasing values. This is because both convex costs and concave valuations push toward gradual prize structures. But we also find that the qualitative effects of convex costs and concave valuations on the optimal prize structure are not identical. In particular, the optimal number of prizes depends more heavily on the curvature of the costs. With linear costs, the optimal number of prizes, while still large, is only a fraction of the number of contestants, even if valuations are very concave. In contrast, if costs are convex and the marginal cost at zero performance is zero, then even with linear valuations almost every contestant is optimally awarded a prize.

We obtain our results in a relatively general environment with many contestants. Each contestant chooses a performance and pays the associated cost, and the prizes are awarded according to the rank order of the performances.[3] The contestants may be ex ante asymmetric in their abilities and prize valuations, and there may be complete or incomplete information about these parameters. Contestants' prize valuations and performance costs may be linear, concave, or convex. We consider all possible prize structures, including identical prizes, heterogeneous prizes, and a combination of identical and heterogeneous prizes. This is important, because restricting the prize structures a priori may rule out the optimal ones.

Solving for equilibrium in the contest environment we consider is an intractable problem. We deal with this difficulty by using the approximation approach for large contests developed by Olszewski and Siegel (2016). In that paper, we considered large contests with a fixed prize structure, and showed that players' equilibrium behavior is approximated by the unique single-agent mechanism that assortatively allocates a continuum of prizes to a continuum of agent types and gives the lowest type a utility of 0. This result allows us to formulate the contest design problem in this paper as a tractable single-agent mechanism design problem, which we then solve.

The rest of the paper is organized as follows. Section 2 describes the contest environment and the contest design problem. Section 3 solves the problem when players

---

[3]This deterministic prize allocation given contestants' choices is similar to the allocation in an all-pay auction and different from the random allocation in a tournament (Lazear and Rosen 1981) or a Tullock (1980) contest.

have linear costs, and Section 4 extends the analysis to more general costs. Section 5 presents a comparative statics result and discusses how to endogenize the prize budget. Section 6 discusses several related papers. Section 7 concludes. Appendix A.1 contains the approximation results. Appendices A.2 and A.3 contain some material omitted from the main text. All proofs are in Appendix A.4.

2. Model

A contest is a game in which players compete for prizes. Each player is characterized by her privately-known type $x \in [0, 1]$, which is drawn from a player-specific distribution (independently across players), and each prize is characterized by its cost $y \in [0, \bar{y}]$, where $y = 0$ corresponds to "no prize" and $\bar{y}$ is the highest possible prize. For concreteness, we will assume that prizes are monetary. Each player chooses her performance $t \geq 0$, the player with the highest performance obtains the highest prize, the player with the second-highest performance obtains the second-highest prize, and so on. Ties are resolved by a fair lottery. The utility of a player of type x from choosing performance t and obtaining prize y is

$$U(x, y, t) = xh(y) - c(t), \tag{1}$$

where $h(0) = c(0) = 0$, and prize valuation h and performance cost c are continuously differentiable and strictly increasing. Notice that functions h and c are common to all players. Notice also that the game is strategically equivalent to one in which players have private information about their performance cost, as in Spence's (1973) signaling model, by dividing the utility by x to obtain $h(y) - c(t)/x$. We assume that sufficiently high performance levels are prohibitively costly, i.e., $h(\bar{y}) < c(t)$ for large enough t, so no player chooses performance higher than $c^{-1}(h(\bar{y}))$. The functional form (1) and special cases thereof have been assumed in numerous existing papers on contests (see, e.g., Clark and Riis 1998, Moldovanu and Sela 2001, Bulow and Levin 2006, and Xiao 2016).

Our focus is on contests with a large (but finite) number of players and prizes. Olszewski and Siegel (2016) (henceforth, OS) showed that all the equilibria of such large contests with a fixed prize structure are closely approximated by the unique single-agent mechanism in a specific environment that implements the assortative allocation of prizes to agent types and gives the lowest type a utility of 0. More precisely, let F be a distribution of types with a continuous, strictly positive density f, and let G be some (not necessarily continuous) distribution of prizes. We interpret F as the average distribution of players' types in the large contest and G as the empirical distribution of prizes. It is important that we do not restrict G, since we will optimize over prize distributions, and any exogenous restriction on the optimal prize distribution would restrict the scope of our analysis. The *assortative allocation* assigns to each type x prize

$$y^A(x) = G^{-1}(F(x)),$$

where

$$G^{-1}(z) = \inf\{y : G(y) \geq z\} \quad \text{for } 0 \leq z \leq 1.$$

That is, the quantile in the prize distribution of the prize assigned to type x is the same as the quantile of type x in the type distribution. It is well known (see, e.g., Myerson 1981) that the unique incentive-compatible mechanism that implements the assortative allocation and gives type $x = 0$ utility 0 specifies for every type x performance

$$t^A(x) = c^{-1}\left(xh\big(y^A(x)\big) - \int_0^x h\big(y^A(\widetilde{x})\big)\, d\widetilde{x}\right). \tag{2}$$

Roughly speaking, the approximation shows that in any equilibrium of a large contest a player with type x with high probability chooses a performance close to $t^A(x)$ and obtains a prize close to $y^A(x)$.

The intuition for why this single-agent mechanism approximates the equilibria of large (finite) contests is that, given players' equilibrium strategies, with a large number of players the law of large numbers implies that each bid leads to an almost deterministic rank-order quantile (in the distribution of bids), and thus to an almost deterministic prize. In the limit, we obtain an "inverse tariff" that maps bids to prizes. Utility (1) implies that higher types choose higher bids from any tariff, so the mechanism induced by the inverse tariff implements the assortative allocation. Any player can bid 0 and obtain the lowest prize, so the utility of type 0 is 0. Appendix A.1 formalizes the approximation result of OS (see OS for additional details).

We point out that the single-agent setting is not a "limit contest" with a continuum of players who compete against each other, but rather a single privately-informed agent who faces a mechanism that implements the assortative allocation of a fixed distribution of prizes. The aforementioned approximation results of OS allow us to avoid having to study a limit contest, which would involve several conceptual and technical difficulties. In the rest of paper, we focus on the single-agent mechanism to investigate the performance-maximizing prize distributions.

The bound $\bar{y}$ is the highest possible value of any single prize. We require this bound because the approximation results of OS rely on the compactness of the spaces of types and prizes.[4] Such a bound is not an issue for the analysis of contests with a fixed prize structure. It is also not an issue for the analysis of the optimal prize structures in settings in which such a bound arises naturally, as the result of policy, fairness considerations, or technological limitations. But in other settings, the per-capita prize budget we impose in Section 2.1 does not naturally imply any bound $\bar{y}$ as the number of contestants grows large. To apply our results to such settings, we will find the optimal prize distributions for all $\bar{y}$, and then take the limit of the optimal structures as $\bar{y}$ diverges to infinity.

2.1 *Performance maximization*

We are interested in the prize distributions G that maximize the aggregate performance t^A subject to a prize budget of Y in the single-agent mechanism that implements the

---

[4]A maximal prize value is also required when it is optimal to spend the entire budget on the highest possible prizes, which we show is sometimes the case; otherwise, an optimal prize structure would not exist.

assortative allocation and gives the lowest type utility 0. The formal statement of the problem is as follows:

$$\max_{G \in \mathcal{G}} \int_0^1 t^A(x) f(x)\, dx$$

$$\text{s.t.} \quad \int_0^{\bar{y}} y^A(x) f(x)\, dx \leq Y, \tag{3}$$

where $\mathcal{G}$ is the set of all possible prize distributions. Appendix A.1 shows that the prize distributions that solve (3) approximate the prize structures that maximize aggregate performance in large (finite) contests, subject to a per-capita prize budget of Y,[5] in a lower- and upper-hemicontinuity sense. This result goes beyond the approximation results of OS, which apply to contests with a fixed prize structure. We begin by solving (3) in the relatively simple case of linear performance costs.

3. Linear cost functions

With linear costs $c(t) = t$, we can use (2) and integration by parts to rewrite (3) as

$$\max_{G \in \mathcal{G}} \int_0^1 h\big(y^A(x)\big)\left(x - \frac{1 - F(x)}{f(x)}\right) dx$$

$$\text{s.t.} \quad \int_0^{\bar{y}} y^A(x) f(x)\, dx \leq Y. \tag{4}$$

For some intuition for why the objective in (4) approximates the expected average performance in large contests, observe that this function coincides with the expected revenue from a bidder in a single-object independent private-value auction if we let $h(y^A(x))$ be the probability that the bidder wins the object when his type is x (Myerson 1981). In the auction setting, increasing the probability that type x obtains the object along with the price the type is charged allows the auctioneer to capture the entire increase in surplus for this type, but requires a decrease in the price that higher types are charged to maintain incentive compatibility. This net increase in revenue, or "virtual value," also coincides with a monopolist's marginal revenue (Bulow and Roberts 1989). In a large contest, increasing the prize that type x obtains also allows the designer to capture the entire surplus increase for this type, because the higher prize increases this type's competition with slightly lower types until the surplus increase from the higher prize is exhausted. But the prize increase also decreases the competition of higher types for their prizes, since the prize of type x becomes more attractive to them.

To get a sense for the solution of (4), suppose that instead of solving for the optimal prize distribution G we wanted to solve for the optimal assortative allocation y^A of prizes to types (which determines G by $y^A(x) = G^{-1}(F(x))$). Then, if we assumed that the optimal y^A was sufficiently smooth, we could put a Lagrange multiplier λ in the

---

[5]For example, a contest with 10,000 contestants and a budget of $1,000,000 would have a per-capita prize budget of $Y = 100$. As we consider larger and larger contests, we require the overall budget to scale proportionally in order to maintain the per-capita prize budget.

constraint and obtain that for any type x such that $y^A(x)$ takes an interior value (i.e., in $(0, \bar{y})$), we would have the first-order condition

$$h'(y^A(x))\left(x - \frac{1 - F(x)}{f(x)}\right) = \lambda.$$

To guarantee that the solution y^A is nondecreasing, we make the following monotonicity assumption, which we maintain throughout Section 3.

ASSUMPTION 1. $x - (1 - F(x))/f(x)$ *strictly increases in* $x \in [0, 1]$.

This monotonicity condition leads to the objective in (4) having increasing differences in $y^A(x)$ and x, and corresponds to Myerson's (1981) "regular case."[6]

Since we are interested in conditions that more directly characterize the optimal prize distribution G, we rewrite the first-order condition by substituting $z = F(x)$ to obtain

$$h'(G^{-1}(z))J(z) = \lambda, \tag{5}$$

where

$$J(z) = F^{-1}(z) - (1 - z)/f(F^{-1}(z)).$$

It is helpful to think of G^{-1} as mapping every quantile in the type distribution to its corresponding prize in the assortative allocation.

The following lemma characterizes the optimal distributions G. Despite the straightforward derivation of (5) above, the proof of the lemma is not completely straightforward because we cannot assume that the optimal G is differentiable (or even continuous) or that G^{-1} takes interior values (i.e., in $(0, \bar{y})$). For the lemma, we note that for any prize distribution G (optimal or not), in the assortative allocation there are quantiles $z_{\min} \leq z_{\max}$ in $[0, 1]$ such that types in quantiles lower than $z_{\min}$ are allocated prize 0 (no prize), types in quantiles higher than $z_{\max}$ are allocated the highest possible prize $\bar{y}$, and types in intermediate quantiles are allocated positive, nonmaximal prizes.

LEMMA 1. *Given a prize distribution* G, *let* $z_{\min} \leq z_{\max}$ *in* $[0, 1]$ *be such that* $G^{-1}(z) = 0$ *for* $z \leq z_{\min}$, $G^{-1}(z) = \bar{y}$ *for* $z > z_{\max}$, *and* $G^{-1}(z) \in (0, \bar{y})$ *for* $z \in (z_{\min}, z_{\max})$. *If* G *is an optimal prize distribution, then it satisfies the following conditions:*

1. *If* $z_{\min} < z_{\max}$ (*Case 1*), *then there exists a* $\lambda \geq 0$ *such that* $h'(G^{-1}(z))J(z) = \lambda$ *for* $z \in (z_{\min}, z_{\max}]$; *in addition,* $h'(0)J(z_{\min}) \leq \lambda$, *and* $h'(\bar{y})J(z_{\max}) \geq \lambda$ *if* $z_{\max} < 1$.

2. *If* $z_{\min} = z_{\max}$ (*Case 2*), *then* $h'(0)J(z_{\min}) \leq h'(\bar{y})J(z_{\max})$.

In the special case of $h'(0) = \infty$, it is understood that $J(z_{\min}) = 0$ and $h'(0)J(z_{\min}) = 0$.

The parameter λ in Case 1 is the shadow price of the budget constraint, i.e., by how much the aggregate performance increases if the budget is increased slightly. This shadow price can be used to determine the optimal budget when the budget is endogenous, as discussed in Section 5.

[6]Without Assumption 1, Myerson's (1981) ironing technique can be used to show that the main qualitative conclusions with linear costs still hold.

3.1 *Concave, linear, and convex prize valuations*

When the budget is large enough, the curvature of the prize valuation h does not affect the optimal prize distribution. To see this, observe that maximizing the objective in (4) pointwise while ignoring the budget constraint leads to choosing the highest possible prize, $\bar{y}$, for types $x > x^*$ and the lowest possible prize, 0, for types $x < x^*$, where $x^* \in (0, 1)$ is the unique type (by Assumption 1) that satisfies $x^* - (1 - F(x^*))/f(x^*) = 0$. This gives us the following result.

PROPOSITION 1. *If $Y \geq \bar{y}(1 - F(x^*))$, then for any function h the optimal prize distribution consists of a mass $1 - F(x^*) \in (0, 1)$ of the highest possible prize, $\bar{y}$, and a mass $F(x^*)$ of prize 0.*

Proposition 1 shows that with a sufficiently large budget it is optimal to award a set of identical prizes, as in the all-pay auctions studied by Clark and Riis (1998), rather than heterogeneous prizes, as in, for example, the all-pay auctions studied by Bulow and Levin (2006), or a combination of identical and heterogeneous prizes. Notice that the optimal mass of prizes, $1 - F(x^*)$, is independent of the value of the highest possible prize, $\bar{y}$.

When the budget is lower than $\bar{y}(1 - F(x^*))$, the curvature of the prize valuation h affects the optimal prize distribution. We first present the simpler result for linear or convex functions h.

PROPOSITION 2. *If $Y < \bar{y}(1 - F(x^*))$ and h is weakly convex, then the optimal prize distribution consists of a mass $Y/\bar{y}$ of the highest possible prize, $\bar{y}$, and a mass $1 - Y/\bar{y}$ of prize 0.*

Proposition 2 shows that awarding identical maximal prizes remains optimal when the budget is low, provided that agents' marginal prize utility is nondecreasing. If the highest possible prize is increased, fewer of these prizes are optimally awarded. The limit as $\bar{y}$ grows arbitrarily large corresponds to a single grand prize.

Proposition 1 and Proposition 2 show that when the budget is large (regardless of the curvature of h) and when the marginal prize utility is nondecreasing (regardless of the size of the budget), the optimal prize distribution does not depend on the precise functional form of h. With a small budget and decreasing marginal prize utility, the optimal prize distribution depends more heavily on h. We first provide a qualitative characterization of the optimal prize distribution in this case, and then a full characterization for strictly concave functions h.

PROPOSITION 3. *Suppose that $Y < \bar{y}(1 - F(x^*))$.*

1. *If h is weakly concave (but not linear on $[0, \bar{y}]$), then any optimal prize distribution assigns positive probability to the set of intermediate prizes $(0, \bar{y})$ and may have atoms only at prize 0 and prize $\bar{y}$.*

2. *If h is strictly concave, then any optimal prize distribution awards all prizes up to the highest prize awarded. That is, the optimal G strictly increases on $[0, G^{-1}(1)]$.*

Proposition 3 shows that decreasing marginal prize utility optimally leads to awarding intermediate prizes, whose values gradually decrease with players' performance ranking. Among the (positive) prizes, only the highest possible prize, $\bar{y}$, may optimally be awarded to multiple players. The following proposition shows that this generally does not occur when $\bar{y}$ is sufficiently large, so the constraint that no prize can exceed $\bar{y}$ does not bind for sufficiently large $\bar{y}$.

PROPOSITION 4. *Suppose that $Y < \bar{y}(1 - F(x^*))$, and let $G_{\max}^{\bar{y}}$ be an optimal prize distribution when $\bar{y}$ is the highest possible prize. If h is weakly concave (but not linear on $[0, \bar{y}]$), and $h'(y) \to 0$ as $y \to \infty$, then there exists a $\bar{y}$ such that $G_{\max}^{\bar{y}'} = G_{\max}^{\bar{y}}$ for any $\bar{y}' \geq \bar{y}$, and this $G_{\max}^{\bar{y}}$ may have an atom only at prize 0.*

We now provide a full characterization of the optimal prize distribution when $Y < \bar{y}(1 - F(x^*))$ and h is strictly concave. Since the optimal G is strictly increasing (part 2 of Proposition 3), G^{-1} is continuous, so we have $h'(0)J(z_{\min}) = \lambda$. Thus,

$$z_{\min} = J^{-1}\big(\lambda / h'(0)\big).$$

Since $h'(G^{-1}(z_{\max}))J(z_{\max}) = \lambda$ and h' is decreasing, $h'(\bar{y})J(z_{\max}) \leq \lambda$. If $z_{\max} < 1$, then we also have $h'(\bar{y})J(z_{\max}) \geq \lambda$ (because we are in Case 1 of Lemma 1), so we obtain $h'(\bar{y})J(z_{\max}) = \lambda$. Thus,

$$z_{\max} = 1 \quad \text{or} \quad J^{-1}\big(\lambda / h'(\bar{y})\big).$$

In addition,

$$G^{-1}(z) = \big(h'\big)^{-1}\big(\lambda/J(z)\big) \quad \text{for } z \in (z_{\min}, z_{\max}] \tag{6}$$

and

$$G^{-1}(z) = \begin{cases} 0 & z \leq z_{\min}, \\ \bar{y} & z > z_{\max}. \end{cases}$$

Thus, G^{-1} is pinned down by λ. The value of λ is determined by the fact that the budget constraint holds as an equality (because $Y < \bar{y}(1 - F(x^*))$).

3.1.1 *An example*   To demonstrate the usefulness of the above characterization, we now derive the optimal G^{-1} for contests with prize valuations $h(y) = y^{1/\eta}$ for $\eta > 1$ (and any type distribution F). We assume that the highest possible prize $\bar{y}$ is large enough that $z_{\max} = 1$ (see Proposition 4), which also implies that $Y < \bar{y}(1 - F(x^*))$, so the entire budget is used. Since $h'(0) = \infty$, we have $z_{\min} = z^*$, where $z^* = F(x^*)$. Since $(h')^{-1}(r) = (\eta r)^{\eta/(1-\eta)}$, by (6) we have

$$G^{-1}(z) = \big(h'\big)^{-1}\big(\lambda/J(z)\big) = \frac{1}{\lambda^{\eta/(\eta-1)}\eta^{\eta/(\eta-1)}}J(z)^{\eta/(\eta-1)} \tag{7}$$

for $z \in (z^*, 1]$. Thus,

$$Y = \int_{z^*}^1 G^{-1}(z)\, dz = \frac{1}{\lambda^{\eta/(\eta-1)}\eta^{\eta/(\eta-1)}} \int_{z^*}^1 J(z)^{\eta/(\eta-1)}\, dz,$$

so

$$\lambda^{\eta/(\eta-1)} = \frac{1}{Y\eta^{\eta/(\eta-1)}} \int_{z^*}^1 J(\widetilde{z})^{\eta/(\eta-1)} \, d\widetilde{z}.$$

Substituting this expression for $\lambda^{\eta/(\eta-1)}$ into (7), we obtain

$$G^{-1}(z) = Y \frac{J(z)^{\eta/(\eta-1)}}{\displaystyle\int_{z^*}^1 J(\widetilde{z})^{\eta/(\eta-1)} \, d\widetilde{z}} \quad \text{for } z \in (z^*, 1] \quad \text{and} \quad G^{-1}(z) = 0 \quad \text{for } z \le z^*. \quad (8)$$

4. More general cost functions

With nonlinear costs c the analysis is substantially more complicated. Nevertheless, conditions similar to the ones in Lemma 1 can be derived and used to characterize the optimal prize distributions. We provide these conditions in Lemma 3 in Appendix A.2. To do so, we make the following assumption, which plays a similar role to that of Assumption 1 in the case of linear costs and facilitates an analytical characterization of the solution.[7] For the assumption, recall that no player chooses performance higher than $c^{-1}(h(\bar{y}))$, and let $\underline{c} = \min\{c'(t) : t \in [0, c^{-1}(h(\bar{y}))]\}$ and $\bar{c} = \max\{c'(t) : t \in [0, c^{-1}(h(\bar{y}))]\}$. We restrict attention to continuously differentiable density functions f.

Assumption 2. *For all z in $[0, 1]$,*

$$\frac{2}{\bar{c}f(F^{-1}(z))} + \frac{f'(F^{-1}(z))(1-z)}{\underline{c}f^3(F^{-1}(z))} > 0, \quad (9)$$

where if $\underline{c} = 0$ the second fraction is equal to ∞, $-\infty$, or 0 when its numerator is positive, negative, or 0, respectively.

Assumption 2 generalizes Assumption 1, because when $\underline{c} = \bar{c} = 1$ the left-hand side of (9) is equal to $J'(z)$. Assumption 2 is satisfied, for example, whenever f is nondecreasing (for any cost function c). The assumption imposes no restrictions on the prize valuation h. In addition, the set of primitives for which Assumption 2 holds is generic in the sense that if it holds for some pair of a continuous derivative of a cost function and a continuous derivative of a density function, then it holds for all such pairs that are sufficiently close to it in the sup norm.

4.1 *Concave and linear prize valuations with convex costs*

We now characterize the optimal prize distribution when the prize valuation h is concave or linear and the costs are convex. This will generalize Proposition 3 and highlight additional features of the optimal prize distribution implied by convex costs. As we will see, the effects of concave prize valuations and convex costs on the optimal prize

---

[7]We strongly conjecture that, similar to the case of linear costs, many of the results for general costs also hold without Assumption 2.

distribution are qualitatively similar, but not identical. It is also possible to generalize the results from Section 3.1 for convex prize valuations to concave costs, but this case seems less relevant for economic applications. Lemma 3 in Appendix A.2, which is the analogue of Lemma 1 for nonlinear costs, can also be used to study the optimal prize distribution for convex prize valuations and convex costs, but no general results exist in this case, because the effects of convex prize valuations and convex costs go in opposite directions.

PROPOSITION 5. *Suppose that h is weakly concave and c is weakly convex but not linear on any interval with lower bound* 0.

1. *Any optimal prize distribution assigns positive probability to the set of intermediate prizes* $(0, \bar{y})$ *and may have atoms only at prize* 0 *and prize* $\bar{y}$.

2. *If c is strictly convex, then any optimal prize distribution awards all prizes up to the highest prize awarded. That is, G strictly increases on* $[0, G^{-1}(1)]$.

3. *If the marginal cost of the first unit of performance is* 0, *i.e.,* $c'(0) = 0$, *then every type $x > 0$ is optimally awarded a positive prize.*

Proposition 5 highlights some similarities and differences between the effects on the optimal prize distribution of convex costs and concave prize valuations. When the budget is large ($Y \geq \bar{y}(1 - F(x^*))$), linear costs with strictly concave prize valuations lead to awarding only a mass of the highest possible prize, but strictly convex costs with linear prize valuations lead to awarding intermediate prizes whose values gradually decrease with players' performance ranking (since Proposition 5 holds for any budget). This is because with convex costs a slight change in the prize a type is awarded induces a larger change in that type's performance when the prize is 0 than when the prize is $\bar{y}$. It therefore cannot be optimal to award prize 0 to some type and prize $\bar{y}$ to a slightly higher type, since a slight increase in the former type's prize along with a corresponding decrease in the latter type's prize increases these types' aggregate performance without significantly affecting all higher types' performance. When the budget is not large ($Y < \bar{y}(1 - F(x^*))$), both concave prize valuations and convex costs optimally lead to awarding intermediate prizes. But with concave valuations types lower than x^* do not obtain a prize, whereas with convex costs whose marginal at 0 is 0 every positive type is optimally awarded a positive prize. This is because a marginal cost of 0 implies that the marginal of the inverse function, i.e., the marginal increase in performance associated with a marginal increase in cost, is infinity. This in turn means that a slight increase from 0 in the prize awarded to a positive type, which leads to an increase in the performance cost this type incurs, generates an increase in that type's performance that infinitely outweighs the decrease in the performance of higher types. It is therefore optimal to have almost every type participate in the contest and obtain a positive prize, unlike with linear costs.

4.2 *An example*

With convex costs and concave valuations, we cannot provide a full characterization of the optimal prize distribution similar to the one following Proposition 4 for linear costs

and concave valuations. This is because that characterization uses Lemma 1, which relies on the marginal prize valuation J being independent of the optimal prize distribution, whereas the corresponding marginal prize valuation K for convex costs (defined in Appendix A.2) depends on the optimal prize distribution. Nevertheless, Proposition 5 and Lemma 3 in Appendix A.2 can be used to explicitly derive the optimal prize distribution once functional forms are specified. We illustrate this with an example, which also highlights the two differences described above between the optimal prize distribution with linear costs and concave valuations and the optimal prize distribution with convex costs.

Suppose that the prize valuation is linear ($h(y) = y$), the costs are quadratic ($c(t) = t^2$), and the type distribution F is uniform. Proposition 5 shows that the optimal prize distribution G assigns positive probability to the set of intermediate prizes $(0, \bar{y})$ and may have atoms only at 0 and $\bar{y}$. In Appendix A.3, we use the conditions in Case 1 of Lemma 3 to derive G^{-1}. Suppose first that $\bar{y}$, the highest possible prize, is at least $4Y$. Then $G^{-1}(z) = 4z^3Y$ for z in $[0, 1]$. This distribution is independent of $\bar{y}$, so the bound $\bar{y}$ on the highest possible prize is not binding, and the associated aggregate performance is $\sqrt{Y/3}$. Consistent with Proposition 5, every positive type obtains a positive prize, the prizes increase gradually from 0 to $4Y$, and there are no atoms (see Figure 1). Now suppose that $\bar{y}$ is less than $4Y$. Then, as long as $\bar{y} \geq 8Y/5$, we have $G^{-1}(z) = 27z^3\bar{y}^4/(64(\bar{y} - Y)^3)$ for z in $[0, 4(\bar{y} - Y)/(3\bar{y})]$ and $G^{-1}(z) = \bar{y}$ for z in $[4(\bar{y} - Y)/(3\bar{y}), 1]$. The associated aggregate performance is $\sqrt{\bar{y} - Y}(1 - 8(\bar{y} - Y)/(9\bar{y}))$. Every positive type still obtains a positive prize, and the prizes increase gradually from 0 to $\bar{y}$, but there is also a mass $(4Y - \bar{y})/(3\bar{y})$ of prize $\bar{y}$ (see Figure 1). If $\bar{y}$ falls below $8Y/5$, then the budget in excess of $5\bar{y}/8$ is optimally not used, so the optimal prize distribution coincides with the one for $\bar{y} = 8Y/5$. Notice that unlike the case of linear costs, and consistent with Proposition 5, even when the budget is large ($Y \geq 5\bar{y}/8$) the optimal prize distribution still awards all prizes between 0 and $\bar{y}$, and every positive type obtains a prize.
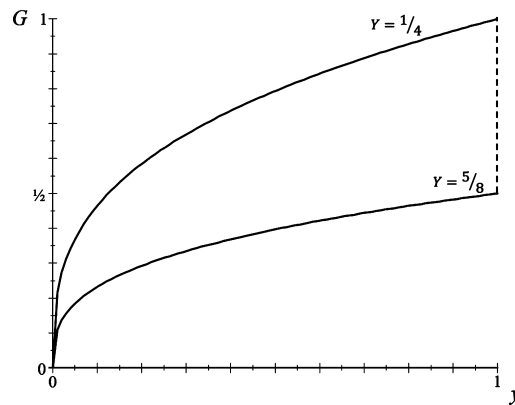

FIGURE 1. The optimal prize distribution for $\bar{y} = 1$ and $Y = 1/4$ and $Y = 5/8$.

5. Discussion

In addition to characterizing the optimal prize distribution, our methods can also be used to derive comparative statics. We provide one result here, which shows that a first-order stochastic dominance (FOSD) shift in the type distribution increases the aggregate performance for any prize distribution.

PROPOSITION 6. *If $\widetilde{F}$ FOSD F, then for any prize distribution G the aggregate performance is higher under $\widetilde{F}$ than under F. In particular, for the optimal prize distributions the aggregate performance is higher under $\widetilde{F}$ than under F.*

While it may seem intuitive that a more able pool of players will generate higher equilibrium performance, this is not always the case in contests with a small number of players. To see this, consider a two-player all-pay auction with complete information and one prize. The prize is $y = 1$, the prize valuation function satisfies $h(1) = 1$, and the cost function is $c(t) = t$. Players' publicly observed types satisfy $0 < x_1 < x_2 < 1$. It is well known (Hillman and Riley 1989) that in the unique equilibrium player 2 chooses a bid by mixing uniformly on the interval $[0, x_1]$ and player 1 bids 0 with probability $1 - x_1/x_2$ and with the remaining probability mixes uniformly on the interval $[0, x_1]$. The resulting expected aggregate bids are $x_1/2 + (x_1)^2/(2x_2)$, which monotonically increase in x_1 and monotonically decrease in x_2. Thus, an increase in player 2's type, even when accompanied by a small increase in player 1's type, decreases the expected aggregate bids. The intuition is that the increased asymmetry between the players, which discourages competition, outweighs the increase in their types, which encourage higher bids. The intuition for Proposition 6 is that in a large contest competition is "localized" in the sense that players compete against players with similar types.[8] Therefore, any decrease in local competition between some types resulting from a FOSD shift in players' type distribution is more than compensated for by an increase in local competition between some higher types.

Our approach can also illustrate how to determine the optimal budget Y (recall that Y was so far assumed exogenous). Suppose that there is a budget cost, which is strictly increasing, continuously differentiable, and takes high enough values for large Y to make the designer never choose such values. To compare the marginal budget cost to the marginal budget benefit, consider the most relevant case of concave h and convex c. Propositions 3 and 5 show that Case 1 of Lemmas 1 and 3 applies. The shadow cost λ is then the marginal budget benefit, so the optimal budget Y can be identified by comparing the marginal budget cost to λ.

As an example, consider the contest with $\bar{y} = 1$, $h(y) = \sqrt{y}$, $c(t) = t$, and F uniform for $Y \le 1/2$. Figure 2 depicts the optimal prize distributions for different budgets $Y \le 1/2$. The prize distributions were computed by using (8) for $Y \le 1/6$ (which implies

---

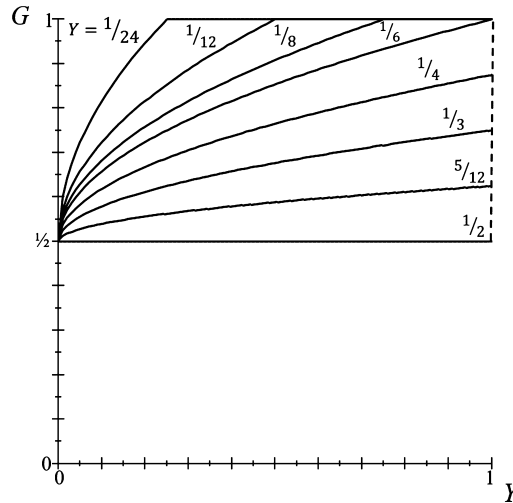[8]A discussion of this phenomenon appears in Bulow and Levin (2006).

FIGURE 2. The optimal prize distributions as the budget varies from 0 to 1/2.

that $z_{\max} = 1$ and there is no mass of prize $\bar{y} = 1$) and by using the more general characterization from Section 3.1 for $Y > 1/6$ (which implies that $z_{\max} < 1$ and there is a mass of prize $\bar{y} = 1$).[9]

The characterization from Section 3.1 shows that $\lambda = 1/\sqrt{24Y}$ for $Y \leq 1/6$, and $\lambda = (3/4 - 3Y/2)$ for $Y > 1/6$.[10] Consistent with Proposition 1, $\lambda = 0$ for $Y = 1/2$, so the budget will never optimally exceed $1/2$. With a linear budget cost of Y, for example, the optimal budget is $Y = 1/24$, and with a quadratic cost of Y^2 the optimal budget is $Y = 3/14$.

6. EXISTING WORK

Several previous papers consider maximizing the expected aggregate output (or effort) in contests. The two closest to our work, Glazer and Hassin (1988) and Moldovanu and Sela (2001) examine this maximization with respect to the prize structure subject to a budget constraint.[11] Both papers study contests in which players' utilities are special cases of (1).

---

[9]The optimal prize distributions are

$$G(y) = \begin{cases} \dfrac{1}{2} + \sqrt{\dfrac{y}{24Y}} & y \in [0, 6Y], \\ 1 & y \in [6Y, 1] \end{cases} \quad \text{and} \quad G(y) = \begin{cases} \dfrac{1}{2} + \sqrt{\dfrac{y(3 - 6Y)^2}{16}} & y \in [0, 1), \\ 1 & y = 1 \end{cases}$$

for $0 < Y \leq 1/6$ and $1/6 < Y \leq 1/2$, respectively.

[10]The maximal average performance is $\sqrt{Y/6}$ for $Y \leq 1/6$ and $(12Y(1 - Y) + 1)/16$ for $Y \geq 1/6$. The difference between the functional forms is due to the atom at the highest possible prize, 1, which appears when the budget exceeds $1/6$.

[11]Moldovanu et al. (2007) and Immorlica et al. (2015) study this maximization in the context of social status. Xiao (2018) numerically computes the optimal division of a budget between two prizes when contestants have linear costs and no private information.

Glazer and Hassin (1988) analyze contests in which contestants are randomly drawn from a population, and use a somewhat specific concept of equilibrium,[12] which facilitates their analysis in a manner similar to that in which our limit approach facilitates the analysis of large contests. They derive an optimal prize structure in two cases. First, when contestants' ability is uniformly distributed in the population, the costs are linear, and prize valuations are weakly concave, they obtain a result that corresponds to our Propositions 2 and 3. Second, when all contestants have identical abilities, they show that the optimal prize structure has $n − 1$ equal prizes and one prize of 0.

Moldovanu and Sela (2001) restrict attention to the symmetric equilibria of discrete contests with ex-ante symmetric contestants, incomplete information, and linear prize valuations, but their results apply to any number of contestants. They show that for weakly concave costs, it is optimal to award the entire budget as a single prize.[13] Moldovanu and Sela (2001) also show that with convex costs awarding the entire budget as a single prize may be inferior to splitting the budget between two prizes.

Proposition 2 is an analogue of the result of Moldovanu and Sela (2001) for linear costs. Although Proposition 2 was established under Assumption 1, it can be shown that this result does not require Assumption 1. Note that Proposition 2 holds for weakly convex (not necessarily linear) prize valuations. In addition, Proposition 2 can be generalized to weakly concave costs by using the conditions in Case 2 of Lemma 3 instead of those in Case 2 of Lemma 1.

Proposition 5 is related to the result of Moldovanu and Sela (2001) that shows that with convex costs splitting the budget into two prizes is sometimes better than awarding the entire budget as a single prize. Proposition 5 goes beyond this, and characterize the optimal prize structure. In addition, our results apply to all equilibria of contests with a large, but finite, number of players. The players may be ex ante symmetric or asymmetric, may or may not have private information, and their prize valuations need not be linear.

Several other papers consider the added value of running several simultaneous or sequential subcontests instead of a single grand contest. Moldovanu and Sela (2006) and Fu and Lu (2012) study multiple rounds of competition in an environment with ex ante identical players. The former restrict attention to identical prizes, and the latter consider Tullock (noisy) contests and focus on symmetric equilibria. Fu and Lu (2009) find conditions under which a grand contest generates higher effort than multiple subcontests when contestants are homogeneous.[14] Xiao (2017) shows that with heterogeneous contestants performance is maximized when contestants with the same ability are assigned to the same subcontest, and characterizes the optimal prize structure in this case. More

---

[12]They disregard the consistency condition between the distribution of abilities in the population and the equilibrium distribution of output of a randomly chosen contestant.

[13]Kaplan and Zamir (2015) notice that this result for linear costs is implied by a result from auction theory, which says that if an auction with ex ante symmetric bidders maximizes revenue, the object must be allocated (if it is allocated at all) to the highest bidder. (The auction-theory result also holds when the object must be allocated and cannot be kept by the auctioneer.)

[14]Fu et al. (2015) identify contest environments in which the temporal and informational structure does not affect the total expected effort or overall outcome.

recently, Hinnosaar (2018) studies effort-maximizing information disclosure policies in sequential Tullock contests. Dasgupta and Nti (1998) and Polishchuk and Tonis (2013) consider the optimal choice of contest success functions, and Nti (2004) compares optimal designs for different contest technologies.

More recently Fang et al. (forthcoming) consider all-pay auctions with complete information and any number of symmetric players with linear valuations, and find that when players' costs are convex, more unequal prize structures lead to lower effort. Similar to the findings of Glazer and Hassin (1988), the effort-maximizing prize structure has all but one of the players obtaining a prize, and all prizes are identical. In contrast, our characterization of the optimal prize structure when players have convex costs and linear valuations shows that when players are asymmetric, captured by distribution F having full support, gradual prize structures are optimal. However, as the asymmetry between players decreases, the optimal prize structure becomes closer to that of Fang et al. (forthcoming). This is easy to see with linear costs and a distribution F "close" to uniform on $[1 - \varepsilon, 1]$ for small $\varepsilon > 0$. In this case, J is "almost constant," so (5) implies that the prizes allocated to types x in $[1 - \varepsilon, 1]$ are nearly identical. Since F is required to have full support, our results for the linear case show that some small measure of low or 0 prizes are also allocated, and this gives rise to the competition by the high types.

The constrained maximization problem (3) is also related to the theory of optimal taxation when the social planner chooses the tax schedule to maximize the tax revenue (Salanié 2003, p. 84). To see the relationship, consider an optimal taxation setting with a continuum of agents, each of whom is characterized by his privately-known productivity x. The population productivity distribution is F. An agent with productivity x who chooses output t incurs cost $c(t/x)$. Before the agents make their output choices, the social planner commits to a tax schedule, which specifies a tax amount for every level of output. Denote by $t(x)$ the output chosen by type x and by $y(x)$ the consumption of type x, so the utility of an agent of type x is $y(x) - c(t(x)/x)$. Suppose that we impose the constraint that agents' aggregate consumption must equal some amount Y. This constraint is not a natural one in an optimal taxation setting. Under this constraint, maximizing the tax revenue is equivalent to maximizing the aggregate output. In addition, instead of choosing the tax schedule the designer can choose the output for each type (which by incentive compatibility pins down each type's consumption up to a constant) or the consumption for each type (which by incentive compatibility pins down each type's output up to a constant). This gives a maximization problem similar to (3), in which the agents' utility $y - c(t/x)$ is somewhat different from the utility $h(y) - c(t)/x$ in our setting.

7. Concluding remarks

This paper investigates the performance-maximizing prize structures in contests with many contestants. Our key qualitative finding is that concave prize valuations and convex performance costs call for numerous prizes of different values. This shows that many sales and workplace competitions, as well as some research grant competitions, can be improved by increasing the number of prizes and making them more heterogeneous.

The analysis facilitates comparative statics and enables deriving closed-form approximations of the performance-maximizing prize distributions for concrete utility functions and distributions of player types.

Our approach can also be used to investigate many other contest design questions. One example is maximizing the expected aggregate performance when not all prize structures are available (for instance, all prizes must be identical), or when the budget is also determined optimally (as discussed in Section 5). Another example is maximizing a weighted sum of contestants' performance. For example, the designer may want to maximize the aggregate performance of contestants whose type exceeds a certain cutoff, or the aggregate performance of the highest-performing q percent of contestants. Our analysis can be extended to capture both scenarios.[15] As an example, consider mathematical olympiads, and suppose that the goal is to identify and encourage the development of the most mathematically gifted individuals. This would correspond in our setting to maximizing the aggregate performance of a top fraction of the contestants. In this case, a minor modification of our analysis implies that it is optimal to spend the entire budget on prizes for this top fraction of contestants. Conditional on this top fraction, the optimal prize distribution is the one that maximizes the aggregate performance for a distribution of types that is the conditional distribution of the types in the top fraction.

More generally, our techniques can be used to investigate various contest design objectives. This can be done by identifying the corresponding objectives in the limit setting and showing that they approximate the ones in large contests. An optimization problem would then be formulated and solved, possibly using the methods we developed here.

Appendix

A.1 *Approximation results*

Approximating the equilibria of large contests with a fixed prize structure by a single-agent mechanism   Consider a sequence of contests parameterized by the number n of players and prizes (since a prize $y = 0$ corresponds to "no prize" it is without loss of generality to have the same number of players and prizes). In the nth contest, let player i's type x_i be distributed according to a cdf F_i that does not have an atom at 0, and let $y_1^n \leq y_2^n \leq \cdots \leq y_n^n$ be the set of prizes. Olszewski and Siegel (2019) show that every contest has an equilibrium in distributional strategies (Milgrom and Weber 1985). The equilibrium approximation technique requires the contests in the sequence to become increasingly similar in some sense as n increases. To formalize this requirement, let $F^n = (\sum_{i=1}^n F_i^n)/n$, so $F^n(x)$ is the expected percentile ranking of type x in the nth contest given the random vector of players' types. Denote by G^n the empirical prize distribution of the n prizes, which assigns a mass of $1/n$ to each prize y_j^n (recall that there is no uncertainty about the prizes). We require that F^n converge pointwise to a distribution F that has a continuous, strictly positive density f, and that G^n converge pointwise to some (not necessarily continuous) distribution G at all points of continuity of G. OS

---

[15]In the limit setting, there is no distinction between the two scenarios, because higher types choose higher performance.

provide several examples to illustrate the convergence of F^n and G^n. The following approximation result appears as Corollary 2 in OS.[16]

THEOREM 1. *(OS) For any $\varepsilon > 0$, there is an N such that for all $n \geq N$, in any equilibrium of the n-th contest each of a fraction of at least $1 - \varepsilon$ of the players i obtains with probability at least $1 - \varepsilon$ a prize that differs by at most ε from $y^A(x_i^n)$, and chooses performance that is with probability at least $1 - \varepsilon$ within ε of $t^A(x_i^n)$.*

Approximating the optimal prize structures in large contests by the optimal prize distributions in a single-agent setting   When $t_i^n(x)$ is the performance of player i of type x in the nth contest, the expected average performance is

$$\frac{1}{n} \sum_{i=1}^{n} \int_0^1 t_i^n(x)\, dF_i^n(x). \tag{10}$$

The budget per capita is Y:

$$\frac{1}{n} \sum_{j=1}^{n} y_j^n \leq Y.$$

The reason that we work with averages is to avoid the quantities becoming infinitely large as n tends to infinity.

We first observe that given a converging sequence of contests, Theorem 1 implies that the expected average performance for large n is approximated by the aggregate performance in the mechanism that implements the assortative allocation in the limit setting:

$$\int_0^1 t^A(x) f(x)\, dx. \tag{11}$$

COROLLARY 1. *For any $\varepsilon > 0$, there is an N such that for all $n \geq N$, in any equilibrium of the nth contest the expected average performance (10) is within ε of (11).*

Corollary 1 applies to a given limit distribution G. Consider a sequence of type distributions that converges to distribution F with a continuous, strictly positive density f, and denote by $G_{\max}^n$ the empirical distribution of prizes that maximizes the equilibrium expected average performance in the nth contest over all equilibria and all sets of prizes $y_1^n \leq \cdots \leq y_n^n$ whose average is no greater than Y. We denote by $M_{\max}^n$ the maximal expected average performance attained by $G_{\max}^n$. For the limit setting, we denote by $\mathcal{M}$ the set of prize distributions that maximize (11) subject to the budget constraint $\int_0^{\bar{y}} y\, dG(y) \leq Y$. An upper hemicontinuity argument shows that a maximizing distribution exists.

---

[16] OS also provide a result on the rates of convergence, which roughly says that it suffices for $1/N$ to be smaller than an expression of order $-\varepsilon^2/\ln \varepsilon$. We refer the reader to their Section 6 for the precise statement of the result.

Lemma 2. *The set $\mathcal{M}$ is not empty.*

Denote by M the corresponding maximal value of (11) subject to the budget constraint. The next proposition formalizes the convergence using any metrization of the weak*-topology on the space of prize distributions.[17]

Proposition 7. *Choose some metrization of the weak*-topology on the space of prize distributions.*

1. *For any $\varepsilon > 0$, there is an N such that for every $n \geq N$, $G_{\max}^n$ is within ε of some distribution in $\mathcal{M}$. In particular, if there is a unique prize distribution $G_{\max}$ that maximizes (11) subject to the budget constraint, then $G_{\max}^n$ converges to $G_{\max}$.*

2. *$M_{\max}^n$ converges to M.*

3. *For any $\varepsilon > 0$, there are an N and a $\delta > 0$ such that for any $n \geq N$ and any empirical prize distribution G^n of n prizes that is within δ of some G in $\mathcal{M}$, the expected average performance in any equilibrium of the nth contest with empirical prize distribution G^n is within ε of $M_{\max}^n$.*

Part 1 of Proposition 7 shows that the optimal prize distributions in large contests are approximated by the prize distributions that maximize (11) subject to the budget constraint. Part 2 shows that the maximal expected average performance is approximated by the maximal value of (11) subject to the budget constraint.[18] Part 3 shows that any prize distribution that is close to a prize distribution that maximizes (11) subject to the budget constraint generates an expected average performance (in any equilibrium) that is close to maximal. For example, given a prize distribution G that maximizes (11) subject to the budget constraint, the set of n prizes defined by $y_j^n = G^{-1}(j/n)$ for $j = 1, \ldots, n$ generates, for large n, an expected average performance that is close to maximal; moreover, the average prize Y^n for the so defined distributions G^n converges to the average prize Y for the distribution G.

A.2 *Optimality conditions for general costs*

To develop the conditions satisfied by an optimal prize distribution for nonlinear costs, it is useful to substitute $x = F^{-1}(z)$ and (2) into the aggregate performance $\int_0^1 t^A(x)f(x)$ to obtain

$$\int_0^1 c^{-1}\bigl(\widetilde{L}(z)\bigr)\,dz, \tag{12}$$

---

[17]Convergence in weak* topology is equivalent to the convergence of CDFs at the points in which the limit is continuous. Roughly speaking, a prize distribution G' is close to a prize distribution G if the graph of the CDF of G' lies in a small neighborhood of the graph of the CDF of G.

[18]The rates of convergence result in OS also indicates that if the optimal prize distribution satisfies the conditions in their Section 6, then the maximal expected average performance is within ε of the maximal value of (11) if $1/N$ is smaller than an expression of order $-\varepsilon^2/\ln \varepsilon$.

where

$$\widetilde{L}(z) = F^{-1}(z)L\big(F^{-1}(z)\big) - \int_0^{F^{-1}(z)} L(\widetilde{x})\, d\widetilde{x} \tag{13}$$

is the cost of the performance of type x in quantile $z = F(x)$ and $L(x) = h(y^A(x))$. Notice that $\widetilde{L}(z)$ is well-defined even when function $G^{-1}(z)$ is not monotone. We will consider such functions in some of our proofs. It is also useful to rewrite the budget constraint as

Similar to Section 3, to derive the conditions for optimality it is useful to consider the effect of a slight increase Δ in the value of G^{-1} at quantile z on the aggregate performance (12). In the case of linear cost c, the effect was to generate an increase of $h'(G^{-1}(z))J(z)\Delta$, where $J(z)$ was the marginal virtual performance. With nonlinear costs, the marginal virtual performance $K(z)$ (given by (14) below) in the corresponding expression for the increase will involve G^{-1} and the derivative of c^{-1}. The expression will be instrumental in formulating conditions that characterize the optimal G^{-1} and generalize the conditions in Lemma 1. But because the expression involves G^{-1}, it cannot be used directly in formulating Assumption 2 below, which guarantees that any optimizer of the relaxed problem is nondecreasing and generalizes Assumption 1, because such an assumption must refer only to the primitives of the model, i.e., only to functions h, c, and F.

To estimate the effect of a slight increase in $G^{-1}(z)$ on (12), consider a function G^{-1} that takes values only in the set $\{0, 1/2^n, 2/2^n, \ldots, (2^n-1)/2^n, 1\}$, and is constant on each interval $(0, 1/2^n], (1/2^n, 2/2^n], \ldots, ((2^n-1)/2^n, 1]$. Suppose that we increase the value of G^{-1} on an interval $(l/2^n, (l+1)/2^n]$ by $\Delta = 1/2^n$. Since $L(F^{-1}(z)) = h(G^{-1}(z))$, this change increases the value of L on $(F^{-1}(l/2^n), F^{-1}((l+1)/2^n)]$ by $h'(G^{-1}((l+1)/2^n))\Delta$, to a first-order approximation. In Figure 3, this corresponds to shifting the graph of L on $(F^{-1}(l/2^n), F^{-1}((l+1)/2^n)]$ to the right by the width of the shaded square. This change does not affect $\tilde{L}$, and thus the integrand in (12), on intervals $(k/2^n, (k+1)/2^n]$ for $k < l$. It increases the integrand for $z \in (l/2^n, (l+1)/2^n]$, to a first-order approximation, by

$$\big(c^{-1}\big)'\big(\widetilde{L}((l+1)/2^n)\big)F^{-1}\big((l+1)/2^n\big)h'\big((G)^{-1}((l+1)/2^n)\big)\Delta$$

(the union of the shaded and darkened rectangles in Figure 3). For any $k > l$, it decreases the integrand for $z \in (k/2^n, (k+1)/2^n]$, to a second-order approximation, by

$$\big(c^{-1}\big)'\big(\widetilde{L}((k+1)/2^n)\big)h'\big((G)^{-1}((l+1)/2^n)\big)\Delta\big[F^{-1}\big((l+1)/2^n\big) - F^{-1}\big(l/2^n\big)\big]$$

(the shaded square in Figure 3).

Since $F^{-1}((l+1)/2^n) - F^{-1}(l/2^n) = \Delta/f(F^{-1}((l+1)/2^n))$, to a first-order approximation, letting $z = (l+1)/2^n$, we express the total increase in (12) as

$$h'\big(G^{-1}(z)\big)\left(F^{-1}(z)\big(c^{-1}\big)'\big(\widetilde{L}(z)\big) - \frac{\displaystyle\int_z^1 \big(c^{-1}\big)'\big(\widetilde{L}(\widetilde{z})\big)\, d\widetilde{z}}{f\big(F^{-1}(z)\big)}\right)\Delta^2.$$
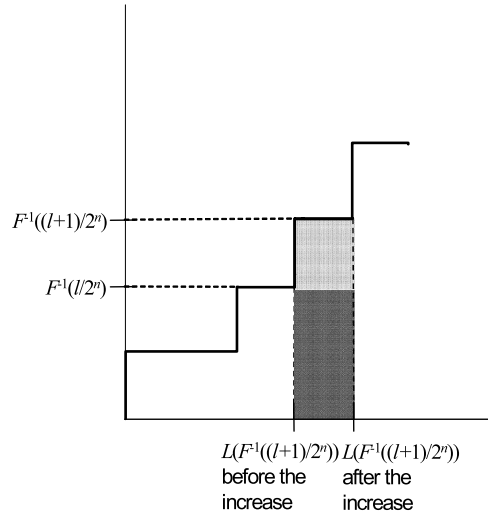
Figure 3. Increasing G^{-1}.

Recalling that $\widetilde{L}(z)$ is the cost of the performance of type x in quantile $z = F(x)$, we can interpret

$$K(z) = \left(F^{-1}(z)\big(c^{-1}\big)'\big(\widetilde{L}(z)\big) - \frac{\int_z^1 \big(c^{-1}\big)'\big(\widetilde{L}(\widetilde{z})\big)\, d\widetilde{z}}{f\big(F^{-1}(z)\big)} \right) \tag{14}$$

as the marginal virtual performance of the type in quantile z: a marginal increase in the prize allocated to this type intensifies competition for this prize and exhausts the corresponding increase in allocation utility (the first term on the right-hand side of (14)), but reduces competition by all higher types (the second term on the right-hand side of (14)). With nonlinear costs these effects depend on the prizes allocated to lower types (through $\widetilde{L}(z)$), because they determine the current performance, which affects the marginal cost of performance. This dependency disappears with linear cost $c(t) = t$, in which case $K(z)$ coincides with $J(z)$.

With linear costs, Assumption 1 guarantees that the maximizer G^{-1} of the relaxed problem is nondecreasing. Indeed, if $G^{-1}(z)$ were lower on an interval $(l/2^n, (l+1)/2^n]$ than on an interval $(j/2^n, (j+1)/2^n]$ for some $j < l$, we could exchange the two values, generating a higher increase on $(l/2^n, (l+1)/2^n]$ than a decrease on $(j/2^n, (j+1)/2^n]$. The assumption that $K(z)$ is strictly increasing would be a natural counterpart of Assumption 1 in the more general setting. Unfortunately, $K(z)$ involves the endogenous variable G^{-1} (through $\widetilde{L}(z)$), which would make the assumption unattractive; moreover, it would no longer serve its purpose, because exchanging the values of G^{-1} on $(l/2^n, (l+1)/2^n]$ and $(j/2^n, (j+1)/2^n]$ would affect the value of $K(z)$. Instead, Assumption 2 in Section 4, which is expressed only in terms of the primitives of the model, guarantees directly that exchanging the values of G^{-1} on the two intervals is beneficial and, therefore, guarantees that the maximizer of the relaxed problem is nondecreasing.

Equipped with (14), we obtain the following analogue of Lemma 1.

LEMMA 3. *Given a prize distribution G, let $z_{\min} \leq z_{\max}$ in $[0, 1]$ be such that $G^{-1}(z) = 0$ for $z \leq z_{\min}$, $G^{-1}(z) = \bar{y}$ for $z > z_{\max}$, and $G^{-1}(z) \in (0, \bar{y})$ for $z \in (z_{\min}, z_{\max})$. If G is an optimal prize distribution, then it satisfies the following conditions:*

1. *If $z_{\min} < z_{\max}$ (Case 1), then there exists a $\lambda \geq 0$ such that*

$$h'\big(G^{-1}(z)\big)K(z) = \lambda \tag{15}$$

for $z \in (z_{\min}, z_{\max}]$; in addition,

$$h'(0)K(z_{\min}) \leq \lambda, \tag{16}$$

and

$$h'(\bar{y})K(z_{\max}) \geq \lambda \tag{17}$$

if $z_{\max} < 1$.

2. *If $z_{\min} = z_{\max}$ (Case 2), then*

$$h'(0)K(z_{\min}) \leq \lim_{z \downarrow z_{\max}} h'(\bar{y})K(z). \tag{18}$$

The difference between Case 2 in Lemma 3 and Case 2 in Lemma 1 arises because $J(z)$ is continuous at every z, whereas $K(z)$ is left continuous at every z but changes discontinuously at quantiles z at which $G^{-1}(z)$ increases discontinuously. In particular, if $z_{\min} = z_{\max}$, then $\widetilde{L}(z_{\min}) = 0$ (type $z_{\min}$ obtains prize 0 and chooses performance 0) but $\widetilde{L}(z) = G(0)h(\bar{y})$ for all $z > z_{\max}$ (types above $z_{\max}$ obtain prize $\bar{y}$ and choose the performance with cost $G(0)h(\bar{y})$, which makes type $z_{\min} = z_{\max} = G(0)$ indifferent between choosing this performance and obtaining prize $\bar{y}$ and choosing performance 0 and obtaining prize 0).

A more subtle difference from Lemma 1 relates to (16). The intuition for (16) is that if the inequality were reversed, then in the relaxed problem increasing $G^{-1}(z)$ for z slightly below $z_{\min}$ by decreasing $G^{-1}(z)$ for z in $(z_{\min}, z_{\max})$ would increase the aggregate performance. This relies on $z_{\min} > 0$, which is always the case with linear costs (because $J(0) < 0$). More generally, however, it can be that $z_{\min} = 0$ (see part 3 of Proposition 5). But in this case (16) follows from (15) directly, because $G^{-1}(z)$ and, therefore, $K(z)$, are continuous at $z = 0$. Otherwise, $y_{\min} = \lim_{z \downarrow 0} G^{-1}(z) > 0$, so G^{-1} could be "shifted down" to reduce $h(G^{-1}(z))$ by $h(y_{\min})$,[19] which would reduce the cost of providing the prizes without changing each type's performance. The prizes $G^{-1}(z)$ for z close to 1 could then be increased, which would increase the aggregate performance.

A.3 *The example from Section 4.2*

Proposition 5 shows that $z_{\min} < z_{\max}$ and G may have atoms only at 0 and $\bar{y}$. We now use the conditions in Case 1 of Lemma 3 to derive G. Define an auxiliary function $q(z) = (c^{-1})'(\widetilde{L}(z))$, plug $q(z)$ into (14) and the resulting expression for $K(z)$ into

---

[19]The new inverse prize distribution would assign prize $G^{-1}(z) - h^{-1}(h(G^{-1}(z)) - h(y_{\min}))$ to quantile z. This maintains the same value of (2) for every x.

(15), and differentiate with respect to z to obtain the differential equation $q'(z)z + 2q(z) = 0$ for $q(z)$.[20] Solving this equation, and substituting back into (15), we obtain $(c^{-1})'(\widetilde{L}(z)) = \lambda/z^2$. By the definition (13) of $\widetilde{L}(z)$, we obtain $((c^{-1})')^{-1}(\lambda/z^2) = zG^{-1}(z) - \int_0^z G^{-1}(t)\,dt$. If G^{-1} is differentiable, differentiating the last equality gives $(G^{-1})'(z) = (-2\lambda/z^4)(((c^{-1})')^{-1})'(\lambda/z^2)$.[21]

Since $c^{-1}(z) = \sqrt{z}$, we have $(c^{-1})'(z) = 1/(2\sqrt{z})$, $((c^{-1})')^{-1}(z) = 1/(4z^2)$, and $(((c^{-1})')^{-1})'(z) = -1/(2z^3)$. Thus, $G^{-1}(z) = z^3/(3\lambda^2) + y_{\min}$, where $y_{\min}$ is the "lowest prize" awarded. By parts 2 and 3 of Proposition 5, $z_{\min} = 0$ and $y_{\min} = 0$.

Consider first $\bar{y} \geq 4Y$. Suppose that $z_{\max} = 1$ and the entire budget is used. Substituting the expression for $G^{-1}(z)$ into the budget constraint with equality, we obtain $\lambda = 1/\sqrt{12Y}$, which gives $G^{-1}(z) = 4z^3Y$. Thus, G^{-1} does not exceed $\bar{y} \geq 4Y$. Substituting G^{-1} into the objective, the aggregate performance is $\sqrt{Y/3}$, which increase in the budget Y, so it is indeed optimal to use the entire budget. Moreover, we cannot have $z_{\max} < 1$, because the budget constraint would be violated: on $[z_{\max}, 1]$ the prize would be $\bar{y}$, higher than with $z_{\max} = 1$, and in order to have $G^{-1}(z_{\max}) = \bar{y}$, the value of λ would have to be lower than that with $z_{\max} = 1$, which implies a pointwise higher value of G^{-1} on $[0, z_{\max}]$ than with $z_{\max} = 1$.

Now suppose that $\bar{y} < 4Y$ and the entire budget is used. Then we still have $G^{-1}(z) = z^3/(3\lambda^2)$ for $z \leq z_{\max}$, but this new λ is different from that for $\bar{y} = 4Y$. (Otherwise, since $G^{-1}(z) = 4Y$ at $z = 1$ for the old λ, the entire budget would not be used.) This implies that $z_{\max} < 1$. Since the budget constraint is satisfied with equality, $\lambda = z_{\max}^2/(12(Y - \bar{y}(1 - z_{\max})))^{1/2}$. Substituting this λ into $\bar{y} = G^{-1}(z_{\max}) = z_{\max}^3/(3\lambda^2)$ gives that $z_{\max} = 4(\bar{y} - Y)/(3\bar{y})$. Substituting the expression for $z_{\max}$ into the expression for λ, and substituting the resulting expression for λ into the expression for G^{-1} for $z \leq z_{\max}$, gives $G^{-1}(z) = 27z^3\bar{y}^4/(64(\bar{y} - Y)^3)$. Substituting this G^{-1} into the objective, the aggregate performance is $\sqrt{\bar{y} - Y}(1 - 8(\bar{y} - Y)/(9\bar{y}))$.

This expression increases for Y in $[\bar{y}/4, 5\bar{y}/8]$, and decreases for Y in $[5\bar{y}/8, \bar{y}]$. Therefore, this expression is the maximal aggregate performance for Y in $[\bar{y}/4, 5\bar{y}/8]$. Any budget in excess of $5\bar{y}/8$ will optimally not be used.

A.4 *Proofs*

PROOF OF LEMMA 1. We first substitute $z = F(x)$ to rewrite the problem as maximizing

$$\int_0^1 h(G^{-1}(z))J(z)\,dz \tag{19}$$

subject to

$$\int_0^1 G^{-1}(z)\,dz \leq Y. \tag{20}$$

---

[20]The solution can be verified to be differentiable.

[21]We will show that an optimal prize distribution G with differentiable inverse G^{-1} exists. No other prize distribution will lead to higher average performance, since the average performance corresponding to any prize distribution can be approximated arbitrarily closely by the average performance corresponding to a prize distribution with a differentiable inverse.

Let G be an optimal distribution, and suppose that $z_{\min} < z_{\max}$. We now show that $h'(G^{-1}(\underline{z}))J(\underline{z}) = h'(G^{-1}(\overline{z}))J(\overline{z})$ for all $\underline{z}, \overline{z} \in (z_{\min}, z_{\max})$. The idea is that if this were not the case, e.g., if we had $<$ instead of $=$, then we could increase G^{-1} around $\overline{z}$ and decrease G^{-1} around $\underline{z}$, thereby increasing the aggregate performance. We must be careful, however, not to violate the budget constraint, and to maintain the monotonicity of G^{-1}. These properties will be easier to control if we first approximate G^{-1} by a piecewise constant function.

To simplify notation, we assume that $\bar{y} = 1$. We approximate G^{-1} by a sequence of inverse distribution functions $((G^n)^{-1})_{n=1}^{\infty}$. To define $(G^n)^{-1}$, partition interval $[0, 1]$ into intervals of size $1/2^n$, and set the value of $(G^n)^{-1}$ on interval $(j/2^n, (j+1)/2^n]$ to be constant and equal to the highest number in the set $\{0, 1/2^n, 2/2^n, \ldots, (2^n - 1)/2^n, 1\}$ that is no higher than $G^{-1}(j/2^n)$. By left continuity of G^{-1}, $(G^n)^{-1}$ converges pointwise to G^{-1}. By definition of $(G^n)^{-1}$ and monotonicity of G^{-1}, $(G^n)^{-1}$ satisfies the budget constraint (20).

Suppose that $h'(G^{-1}(\underline{z}))J(\underline{z}) < h'(G^{-1}(\overline{z}))J(\overline{z})$ for some $\underline{z}, \overline{z} \in (z_{\min}, z_{\max})$. By left continuity of G^{-1}, and continuity of h' and J, the previous inequality also holds for points slightly smaller than $\underline{z}$ and $\overline{z}$. Thus, there are $\delta > 0$, N, and intervals $(l/2^N, (l+1)/2^N]$ and $(j/2^N, (j+1)/2^N]$, such that for every $n \geq N$ we have $h'((G^n)^{-1}(\overline{z}'))J(\overline{z}') - h'((G^n)^{-1}(\underline{z}'))J(\underline{z}') > \delta$ for any $\overline{z}' \in (j/2^N, (l+1)/2^N]$ and $\underline{z}' \in (l/2^N, (j+1)/2^N]$.

Denote by I the infimum of the values $h'((G^n)^{-1}(z))J(z)$ for $n \geq N$ and z in the former interval, and by S the supremum of the values $h'((G^n)^{-1}(z))J(z)$ for $n \geq N$ and z in the latter interval. Thus, we have that $I - S \geq \delta$. Define functions $(H^n)^{-1}$ by increasing the value of $(G^n)^{-1}$ on $(j/2^N, (j+1)/2^N]$ by ε, and decreasing the value of $(G^n)^{-1}$ on $(l/2^N, (l+1)/2^N]$ by ε, so the budget constraint is maintained. For sufficiently small $\varepsilon > 0$, the former change increases (19) at least by $(\varepsilon/2^N)(I - \delta/3)$, and the latter change decreases (19) at most by $(\varepsilon/2^N)(S + \delta/3)$. This increases the value of (19) by at least $\delta\varepsilon/(3 \cdot 2^N)$ (for all $n \geq N$), since $I - S \geq \delta$.

If functions $(H^n)^{-1}$ are monotone, they are inverse distribution functions, and the value of (19) with $(H^n)^{-1}$ instead of G^{-1} exceeds, for large enough n, the value of (19) for G^{-1}. If functions $(H^n)^{-1}$ are not monotone, define $(\widetilde{H}^n)^{-1}$ by setting its value on interval $(0, 1/2^n]$ to the lowest value of $(H^n)^{-1}$ over intervals $(0, 1/2^n], (1/2^n, 2/2^n], \ldots, ((2^n - 1)/2^n, 1]$, setting its value on interval $(1/2^n, 2/2^n]$ to the second lowest value of $(H^n)^{-1}$ on these intervals, etc. The value of (19) with $(\widetilde{H}^n)^{-1}$ instead of G^{-1} is higher than with $(H^n)^{-1}$ instead of G^{-1}, because J is an increasing function.

The second condition in Case 1 and the condition in Case 2 are obtained by analogous arguments, noticing that $z_{\min} > 0$ (since $J(0) < 0$) and, since J is increasing and continuous, the inequality $h'(\bar{y})J(z) \geq \lambda$ for $z > z_{\max}$ is equivalent to $h'(\bar{y}) \times J(z_{\max}) \geq \lambda$. $\qquad\blacksquare$

Proof of Proposition 2. Weak convexity implies that $z_{\min} = z_{\max}$, so only prizes 0 and $\bar{y}$ are awarded. Otherwise, since h' and G^{-1} are weakly increasing and J is strictly increasing, for any $z' < z''$ in $(z_{\min}, z_{\max})$ we would have $h'(G^{-1}(z'))J(z') < h'(G^{-1}(z''))J(z'')$, which would violate the condition $h'(G^{-1}(z'))J(z') = h'(G^{-1}(z'')) \times J(z'') = \lambda$ in Case 1 of Lemma 1. $\qquad\blacksquare$

PROOF OF PROPOSITION 3. Observe that $z_{\min} < z_{\max}$. Indeed, since $h'(0) > h'(\bar{y})$, we cannot have that $z_{\min} = z_{\max}$ and $h'(0)J(z_{\min}) \leq h'(\bar{y})J(z_{\max})$, unless $J(z_{\min}) = J(z_{\max}) \leq 0$. But $J(z_{\max}) \leq 0$ implies that $z_{\max} \leq z^*$. Since $G^{-1}(z) = \bar{y}$ for $z > z_{\max}$, we obtain that $\int_0^1 G^{-1}(z)\,dz \geq \bar{y}(1 - F(x^*)) > Y$ violates the budget constraint (20). This yields the first part of 1. For the second part, notice that $G^{-1}(z)$ strictly increases in z on interval $(z_{\min}, z_{\max})$, so G does not have atoms there. This follows from the fact that $h'(G^{-1}(z))J(z) = \lambda$ on $(z_{\min}, z_{\max}]$ and the fact that $J(z)$ strictly increases in z.

To see 2, note that h' is strictly decreasing and, by assumption, continuous. Thus, $h'(G^{-1}(z))J(z) = \lambda$ also implies that G^{-1} is continuous on $(z_{\min}, z_{\max})$. If G^{-1} were not right continuous at $z_{\min}$, then the fact that $h'(0)J(z_{\min}) \leq \lambda$ and the assumption that h' is strictly decreasing would violate the condition $h'(G^{-1}(z))J(z) = \lambda$ for z slightly higher than $z_{\min}$.[22] Thus, G^{-1} is continuous on $[z_{\min}, z_{\max}]$, which means that G strictly increases on $[0, G^{-1}(z_{\max})]$. If $G^{-1}(z_{\max}) = G^{-1}(1)$, this completes the proof. If $G^{-1}(z_{\max}) < G^{-1}(1)$, which can happen when $z_{\max} < 1$, then $G^{-1}(z_{\max}) < G^{-1}(1) = \bar{y}$, which also completes the proof, as otherwise the fact that $h'(\bar{y})J(z_{\max}) \geq \lambda$ and the assumption that h' is strictly decreasing would violate the condition $h'(G^{-1}(z_{\max}))J(z_{\max}) = \lambda$. □

PROOF OF PROPOSITION 4. Let $z_{\min}^{\bar{y}}$, $z_{\max}^{\bar{y}}$, and $\lambda^{\bar{y}}$ denote $z_{\min}$, $z_{\max}$, and λ for a given $\bar{y}$. The proof of Proposition 3 shows that $z_{\min}^{\bar{y}} < z_{\max}^{\bar{y}}$ for all $\bar{y}$. We claim that $\lambda^{\bar{y}}$ weakly increases with $\bar{y}$. Suppose to the contrary that $\lambda^{\bar{y}'} > \lambda^{\bar{y}''}$ for some $\bar{y}' < \bar{y}''$.

Since $h'((G_{\max}^{\bar{y}})^{-1}(z))J(z) = \lambda^{\bar{y}}$ for all $z \in (z_{\min}^{\bar{y}}, z_{\max}^{\bar{y}}]$ and h' is decreasing, $h'(0)J(z) \geq \lambda^{\bar{y}}$ for all $z \in (z_{\min}^{\bar{y}}, z_{\max}^{\bar{y}}]$, and since J is continuous, we have $h'(0)J(z_{\min}^{\bar{y}}) \geq \lambda^{\bar{y}}$. Since we also have $h'(0)J(z_{\min}^{\bar{y}}) \leq \lambda^{\bar{y}}$ (because we are in Case 1 of Lemma 1), we obtain $h'(0)J(z_{\min}^{\bar{y}}) = \lambda^{\bar{y}}$. Since J is increasing, this implies that $z_{\min}^{\bar{y}'} > z_{\min}^{\bar{y}''}$. In particular, we have (a): $(G_{\max}^{\bar{y}'})^{-1}(z) = 0 \leq (G_{\max}^{\bar{y}''})^{-1}(z)$ for all $z \leq z_{\min}^{\bar{y}'}$, and the inequality is strict for $z \in (z_{\min}^{\bar{y}''}, z_{\min}^{\bar{y}'})$. Since $h'((G_{\max}^{\bar{y}})^{-1}(z))J(z) = \lambda^{\bar{y}}$ for all $z \in (z_{\min}^{\bar{y}}, z_{\max}^{\bar{y}}]$ and h' is decreasing, we have (b): $(G_{\max}^{\bar{y}'})^{-1}(z) \leq (G_{\max}^{\bar{y}''})^{-1}(z)$ for all $z \in (z_{\min}^{\bar{y}'}, \min\{z_{\max}^{\bar{y}'}, z_{\max}^{\bar{y}''}\}]$. If $z_{\max}^{\bar{y}'} \geq z_{\max}^{\bar{y}''}$, then we have (c): $(G_{\max}^{\bar{y}'})^{-1}(z) \leq \bar{y}' < (G_{\max}^{\bar{y}''})^{-1}(z) = \bar{y}''$ for $z > \min\{z_{\max}^{\bar{y}'}, z_{\max}^{\bar{y}''}\}$. If $z_{\max}^{\bar{y}'} < z_{\max}^{\bar{y}''} \leq 1$, then $h'(\bar{y}')J(z_{\max}^{\bar{y}'}) \geq \lambda^{\bar{y}'}$ (because we are in Case 1 of Lemma 1). But $h'((G_{\max}^{\bar{y}''})^{-1}(z_{\max}^{\bar{y}'}))J(z_{\max}^{\bar{y}'}) = \lambda^{\bar{y}''}$, so $\lambda^{\bar{y}'} > \lambda^{\bar{y}''}$ implies that $(G_{\max}^{\bar{y}''})^{-1}(z_{\max}^{\bar{y}'}) \geq \bar{y}'$. Thus, as the inverse of any *cdf* is increasing, we again obtain (c), except that this time $(G_{\max}^{\bar{y}''})^{-1}(z) \leq \bar{y}''$. Now, (a), (b), and (c) imply that the budget constraint cannot be satisfied with equality by both $G_{\max}^{\bar{y}'}$ and $G_{\max}^{\bar{y}''}$, which completes the proof that $\lambda^{\bar{y}}$ weakly increases with $\bar{y}$.

By $h'(0)J(z_{\min}^{\bar{y}}) = \lambda^{\bar{y}}$, we obtain that $z_{\min}^{\bar{y}}$ also weakly increases with $\bar{y}$. If $z_{\max}^{\bar{y}'} > z_{\max}^{\bar{y}''}$ for $\bar{y}' < \bar{y}''$, then $h'((G_{\max}^{\bar{y}'})^{-1}(z_{\max}^{\bar{y}''}))J(z_{\max}^{\bar{y}''}) = \lambda^{\bar{y}'}$ and $h'(\bar{y}'')J(z_{\max}^{\bar{y}''}) \geq \lambda^{\bar{y}''} \geq \lambda^{\bar{y}'}$, which

---

[22]More precisely, the argument delivering the right continuity at $z_{\min}$ applies only to cases in which $h'(0) < \infty$. The case when $h'(0) = \infty$ requires a somewhat special treatment.

If $h'(0) = \infty$, then $J(z_{\min}) = 0$, so if G^{-1} were not right continuous at $z_{\min}$ the product $h'(G^{-1}(z))J(z)$ would be strictly positive for any $z \in (z_{\min}, z_{\max}]$, but would approach 0 as $z \downarrow z_{\min}$, so could not be constant on $(z_{\min}, z_{\max}]$.

would imply that $(G_{\max}^{\bar{y}'})^{-1}(z_{\max}^{\bar{y}''}) \geq \bar{y}'' > \bar{y}'$. Thus, $z_{\max}^{\bar{y}}$ also weakly increases with $\bar{y}$. Moreover, $z_{\max}^{\bar{y}}$ converges to 1 as $\bar{y}$ diverges, because otherwise the budget constraint would be violated for large enough values of $\bar{y}$. Because $h'((G_{\max}^{\bar{y}})^{-1}(z))J(z) = \lambda^{\bar{y}}$, we have that $(G_{\max}^{\bar{y}''})^{-1}(z) \leq (G_{\max}^{\bar{y}'})^{-1}(z)$ for all $z \leq z_{\max}^{\bar{y}'}$.[23]

Notice that $z_{\max}^{\bar{y}} = 1$ for sufficiently large $\bar{y}$. Otherwise, the condition $h'(\bar{y})J(z_{\max}) \geq \lambda$ cannot be satisfied for large enough $\bar{y}$, by the assumption that $h'(y) \to 0$ as $y \to \infty$. And if $z_{\max}^{\bar{y}} = 1$ for some $\bar{y}$, then $(G_{\max}^{\bar{y}'})^{-1} \equiv (G_{\max}^{\bar{y}})^{-1}$ for all $\bar{y}' \geq \bar{y}$, because $(G_{\max}^{\bar{y}'})^{-1}(z) \leq (G_{\max}^{\bar{y}})^{-1}(z)$ for all $z \leq z_{\max}^{\bar{y}}$ and both $G_{\max}^{\bar{y}}$ and $G_{\max}^{\bar{y}'}$ satisfy the budget constraint with equality. This completes the proof.   □

PROOF OF PROPOSITION 5. The first claim in part 1 of the proposition is true because $z_{\min} < z_{\max}$. Indeed, if $z_{\min} = z_{\max}$, then $\widetilde{L}(z) = G(0)h(\bar{y})$ for all $z > z_{\max}$ (as explained immediately after Lemma 3), and since $\widetilde{L}(z_{\min}) = 0$ and $(c^{-1})'(0) > (c^{-1})'(G(0)h(\bar{y}))$, we obtain that $K(z_{\min}) > \lim_{z \downarrow z_{\max}} K(z)$. Together with $h'(0) \geq h'(\bar{y})$, (18) is violated.[24] For the second claim in part 1, an atom at some intermediate prize would mean that $G^{-1}(\underline{z}) = G^{-1}(\bar{z})$ for some $z_{\min} < \underline{z} < \bar{z} < z_{\max}$. We would then have $h'(G^{-1}(\underline{z})) = h'(G^{-1}(\bar{z}))$ and $\widetilde{L}(z)$ constant on $[\underline{z}, \bar{z}]$. The derivative of $K(z)$ on $[\underline{z}, \bar{z}]$ would then be

$$\frac{2}{f(F^{-1}(z))}(c^{-1})'(\widetilde{L}(z)) + \frac{f'(F^{-1}(z))\int_z^1 (c^{-1})'(\widetilde{L}(\tilde{z}))\,d\tilde{z}}{f^3(F^{-1}(z))},$$

which is strictly positive if $f'(F^{-1}(z)) \geq 0$, and also if $f'(F^{-1}(z)) < 0$ (by Assumption 2). We could then not have (15) for both $z = \underline{z}$ and $z = \bar{z}$.

For part 2, notice that $\widetilde{L}(z)$ increases discontinuously when $G^{-1}(z)$ increases discontinuously. So, if $(c^{-1})'$ is strictly decreasing, a discontinuity in $G^{-1}(z)$ would lead to a discontinuous decrease in the left-hand side of (15). Thus, G^{-1} is continuous on $(z_{\min}, z_{\max}]$. If G^{-1} were not right continuous at $z_{\min}$, then (15) and (16) could not both be satisfied, because of the discontinuous decrease of $(c^{-1})'$ at $z_{\min}$ (and, if h is strictly concave, also a discontinuous decrease of h'). Thus, G strictly increases on $[0, G^{-1}(z_{\max})]$. If $z_{\max} < 1$, then $G^{-1}(z_{\max}) = G^{-1}(1) = \bar{y}$. Indeed, if $G^{-1}(z_{\max}) < G^{-1}(1)$, then (15) and (17) could not both be satisfied, because of the discontinuous decrease in the left-hand side of (15) at $z_{\max}$.

For part 3, suppose that $z_{\min} > 0$. If G^{-1} is discontinuous at $z_{\min}$, then (16) cannot hold. And if G^{-1} is continuous at $z_{\min} > 0$, then $K(z)$, and so the left-hand side of (15), diverge to ∞ for z that tends to $z_{\min}$ from the right, so (15) is violated for z's close to $z_{\min}$.   □

PROOF OF PROPOSITION 6. Choose some prize distribution G. By looking at the areas below the graphs of L and L^{-1} (where $L(x) = h(y^A(x))$) in the square $[0, x] \times [0, L(x)]$,

---

[23]Therefore, $(G_{\max}^m)^{-1}$ converges pointwise to some G^{-1} on $[0, 1)$, even when $h'(y) \not\to 0$ as $y \to \infty$. We cannot conclude, however, that this G^{-1} is an inverse *cdf*. For example, G^{-1} can be a constant function equal to 0.

[24]If h were strictly convex, we would have $h'(0) < h'(\bar{y})$, so we could be in Case 1 or Case 2 of Lemma 3.

we obtain that the cost of the performance of type x in the mechanism that implements the assortative allocation satisfies

$$xL(x) - \int_0^x L(\tilde{x}) \, d\tilde{x} = \int_0^{L(x)} L^{-1}(u) \, du.$$

Thus, the aggregate performance is equal to

$$\int_0^1 \left(c^{-1} \left(\int_0^{L(x)} L^{-1}(u) \, du \right) \right) f(x) \, dx = \int_0^1 \left(c^{-1} \left(\int_0^{h(G^{-1}(z))} L^{-1}(r) \, dr \right) \right) dz, \qquad (21)$$

where the equality follows from the change of variables $z = F(x)$ and the identity $L(F^{-1}(z)) = h(G^{-1}(z))$.[25] Since a FOSD shift in F decreases F and, therefore, L pointwise, it increases L^{-1} pointwise and, therefore, increases (21).   □

PROOF OF COROLLARY 1.   Theorem 1 shows that for large n, in any equilibrium of the nth contest the expected average performance is within $\varepsilon/2$ of

$$\frac{\sum_{i=1}^n \int_0^1 t^A(x) \, dF_i^n(x)}{n} = \int_0^1 t^A(x) \, dF^n(x),$$

where the equality follows from the definition of F^n. In addition,

$$\int_0^1 t^A(x) \, dF^n(x) \to_n \int_0^1 t^A(x) \, dF(x),$$

which follows from the fact that t^A is monotonic and the assumption that F is continuous, because $\int g \, dF^n \to_n \int g \, dF$ for any bounded and measurable function g for which distribution F assigns measure 0 to the set of points at which function g is discontinuous. (This fact is established as the first claim of the proof of Theorem 25.8 in Billingsley 1995.) Thus, for large n, $\int_0^1 t^A(x) \, dF^n(x)$ is within $\varepsilon/2$ of $\int_0^1 t^A(x) \, dF(x)$.   □

PROOF OF LEMMA 2.   Let $(G^n)_{n=1}^\infty$ be a sequence on which (11) converges to its supremum, and which satisfies the budget constraint. By passing to a convergent subsequence (in the weak*-topology) if necessary, assume that G^n converges to some G. We will show below that $(G^n)^{-1}$ converges almost surely to G^{-1}. This will imply that $(y^n)^A(x) = (G^n)^{-1}(F(x))$ converges almost surely to $y^A(x) = G^{-1}(F(x))$, and since functions h and c^{-1} are continuous, also that $(t^n)^A(x)$ given by (2) with G replaced with G^n converges almost surely to $t^A(x)$ given by (2). This will in turn imply that the value of (11) with $(G^n)^{-1}$ instead of G^{-1} converges to the value of (11). Finally, as G^n satisfies the budget constraint, G satisfies the budget constraint as well. Indeed, the budget constraints are integrals of a continuous function (mapping y to y) with respect to distributions G and G^n, respectively, and weak*-topology may be alternatively defined as convergence of integrals of continuous functions.

---

[25]Even though L^{-1} may be discontinuous, because G^{-1} may be discontinuous, it is monotonic, so the change of variables applies.

Thus, it suffices to show that $(G^n)^{-1}$ converges to G^{-1}, except perhaps on the (at most) countable set $R = \{r \in [0, 1]:$ there exist $y' < y''$ such that $G(y) = r$ for $y \in (y', y'')\}$.

Suppose first that for some $r \in [0, 1]$ and $\delta > 0$ we have that $(G^n)^{-1}(r) \leq G^{-1}(r) - \delta$ for arbitrarily large n. Passing to a subsequence if necessary, assume that the inequality holds for all n, and that $(G^n)^{-1}(r)$ converges to some $y \leq G^{-1}(r) - \delta$. Then there exists a prize z such that $y < z < G^{-1}(r)$ and G is continuous at z. We cannot have that $G(z) = r$, since this would imply that $G^{-1}(r) \leq z$. Thus, $G(z) < r$. Since $G^n(z)$ converges to $G(z)$, as G is continuous at z, we have that $G^n(z) < r$ for large enough n. This yields $z \leq (G^n)^{-1}(r)$, contradicting the assumption that $(G^n)^{-1}(r)$ converges to $y < z$.

Suppose now that for some $r \in [0, 1] - R$ and $\delta > 0$ we have that $(G^n)^{-1}(r) \geq G^{-1}(r) + \delta$ for arbitrarily large n. Passing to a subsequence if necessary, assume that the inequality holds for all n, and that $(G^n)^{-1}(r)$ converges to some $y \geq G^{-1}(r) + \delta$. Then there exists a prize z such that $G^{-1}(r) < z < y$ and G is continuous at z. We have that $r < G(z)$, as $r \notin R$. Since $G^n(z)$ converges to $G(z)$, as G is continuous at z, we have that $r \leq G^n(z)$ for large enough n. This yields $(G^n)^{-1}(r) \leq z$, contradicting the assumption that $(G^n)^{-1}(r)$ converges to $y > z$. □

PROOF OF PROPOSITION 7. Since every sequence of distributions has a converging subsequence in weak*-topology, suppose without loss of generality that $G^n_{\max}$ converges to some distribution G. Denote the value of (11) under distribution G by V. If part 1 is false, then $G \notin \mathcal{M}$, so $V < M$. The distribution G satisfies the budget constraint, since distributions $G^n_{\max}$ satisfy the budget constraint.

Consider a distribution $G_{\max} \in \mathcal{M}$, and for every n consider an empirical distribution G^n of a set of n prizes, such that G^n converges to $G_{\max}$ in weak*-topology. For example, such a set of n prizes is defined by $y^n_j = G^{-1}_{\max}(j/n)$ for $j = 1, \ldots, n$.

Corollary 1 shows that for large n the expected average performance in any equilibrium of the nth contest with empirical prize distribution G^n exceeds $(V + M)/2$. On the other hand, Corollary 1 also shows that for large n the expected average performance in any equilibrium of the nth contest with empirical prize distribution $G^n_{\max}$ falls below $(V + M)/2$. This contradicts the definition of $G^n_{\max}$ for large n.

For part 2, Corollary 1 applied to the sequence G^n defined above implies that $\liminf M^n_{\max} \geq M$. If $\limsup M^n_{\max} > M$, then there is a corresponding subsequence of $G^n_{\max}$. A converging subsequence of this subsequence has a limit G. For this G, the value of (11) is by Corollary 1 strictly larger than M, a contradiction.

For part 3, notice that the proof of Lemma 2 also implies that the set M is closed in weak*-topology. Thus, if part 3 were false, there would exist a sequence of contests with empirical prize distributions G^n converging to some G in M, such that the expected average performance in an equilibrium of the nth contest with empirical prize distribution G^n would be lower than $M^n_{\max} - \varepsilon$. This would contradict part 2 and Corollary 1. □

PROOF OF LEMMA 3. The idea of the proof is analogous to that of the proof of Lemma 1. As in that proof, suppose that $z_{\min} < z_{\max}$, i.e., we are in Case 1; the condition in Case 2 is obtained by analogous arguments. For an optimal distribution G, approximate G^{-1} by a sequence of inverse distribution functions $((G^n)^{-1})^\infty_{n=1}$ that are constant on intervals $(j/2^n, (j + 1)/2^n]$ and with values in the set $\{0, 1/2^n, 2/2^n, \ldots, (2^n - 1)/2^n, 1\}$. If (15)

is violated, we construct functions $(H^n)^{-1}$ (also constant on intervals $(j/2^n, (j+1)/2^n]$ and with values in the set $\{0, 1/2^n, 2/2^n, \ldots, (2^n-1)/2^n, 1\}$) such that the value of the objective (12) with $(H^n)^{-1}$ instead of G^{-1} exceeds, for large enough n, that of (12) for G^{-1}. This part of the proof replicates the argument from the corresponding part of the proof of Lemma 1, and will be omitted. If for a large enough n function $(H^n)^{-1}$ is non-decreasing, it is an inverse distribution function. We then obtain a contradiction to the optimality of G^{-1}, which completes the proof. If function $(H^n)^{-1}$ is not monotone, we define another function $(\widetilde{H}^n)^{-1}$ whose value on interval $(0, 1/2^n]$ is equal to the lowest value of $(H^n)^{-1}$ over intervals $(0, 1/2^n], (1/2^n, 2/2^n], \ldots, ((2^n-1)/2^n, 1]$, whose value on interval $(1/2^n, 2/2^n]$ is equal to the second lowest value of $(H^n)^{-1}$ on these intervals, and so on. We will complete the proof by showing that the value of (12) is no lower for $(\widetilde{H}^n)^{-1}$ than for $(H^n)^{-1}$ for sufficiently large n's.

To show this, we will consider only two adjacent intervals $(j/2^n, (j+1)/2^n]$ and $(l/2^n, (l+1)/2^n]$ (i.e., $j+1 = l$) such that $(H^n)^{-1}(z) = U$ on $(j/2^n, (j+1)/2^n]$ and $(H^n)^{-1}(z) = D$ on $(l/2^n, (l+1)/2^n]$, where $D < U$, and estimate the effect on (12) of changing the value of $(H^n)^{-1}$ on $(j/2^n, (j+1)/2^n]$ to D and changing the value of $(H^n)^{-1}$ on $(l/2^n, (l+1)/2^n]$ to U. We will use the same symbol $(\widetilde{H}^n)^{-1}$ to denote the function obtained from $(H^n)^{-1}$ as a result of this change, and we will sometimes use symbol Δ to denote $1/2^n$.

The exchange of D and U does not affect the integrand of (12) on the intervals lower than $(j/2^n, (j+1)/2^n]$. It affects the value of $\widetilde{L}$ on interval $(j/2^n, (j+1)/2^n]$, increasing it by some $\widetilde{\Delta}_j$, as well as the value of $\widetilde{L}$ on interval $(l/2^n, (l+1)/2^n]$, increasing it by some $\widetilde{\Delta}_l$. As a result of the change in $\widetilde{L}$ on the two intervals, (12) increases by

$$\Delta\big[c^{-1}\big(\widetilde{L}((j+1)/2^n) + \widetilde{\Delta}_j\big) - c^{-1}\big(\widetilde{L}((j+1)/2^n)\big)\big]$$
$$+ \Delta\big[c^{-1}\big(\widetilde{L}((l+1)/2^n) + \widetilde{\Delta}_l\big) - c^{-1}\big(\widetilde{L}((l+1)/2^n)\big)\big]$$
$$= \Delta\big[c^{-1}\big(\widetilde{L}((l+1)/2^n) + \widetilde{\Delta}_l\big) - c^{-1}\big(\widetilde{L}((j+1)/2^n)\big)\big]$$
$$+ \Delta\big[c^{-1}\big(\widetilde{L}((j+1)/2^n) + \widetilde{\Delta}_j\big) - c^{-1}\big(\widetilde{L}((l+1)/2^n)\big)\big].$$

Observe that

$$\big[\widetilde{L}((l+1)/2^n) + \widetilde{\Delta}_l\big] - \big[\widetilde{L}((j+1)/2^n)\big] = \big[F^{-1}((j+1)/2^n) - F^{-1}(j/2^n)\big]\big[h(U) - h(D)\big].$$

This is easiest to see by looking at Figure 4, in which the graph of L for $(\widetilde{H}^n)^{-1}$ is obtained from the graph of L for $(H^n)^{-1}$ by moving it to the right by the darkened rectangle, and moving it to the left by the shaded rectangle. By definition, $\widetilde{L}$ for $(H^n)^{-1}$ on $(j/2^n, (j+1)/2^n]$ is equal to the area of the rectangle $[0, h(U)] \times [0, F^{-1}(j/2^n)]$ minus the area to the left of the graph of L for $(H^n)^{-1}$ on the interval $[0, F^{-1}(j/2^n)]$. Similarly, $\widetilde{L}$ for $(\widetilde{H}^n)^{-1}$ on $(l/2^n, (l+1)/2^n]$ is equal to the area of the rectangle $[0, h(U)] \times [0, F^{-1}(l/2^n)]$ minus the area to the left of the graph of L for $(\widetilde{H}^n)^{-1}$ on the interval $[0, F^{-1}(l/2^n)]$. So, the difference between the latter and the former areas consists only of the shaded rectangle.

Similarly,

$$\big[\widetilde{L}((j+1)/2^n) + \widetilde{\Delta}_j)\big] - \big[\widetilde{L}((l+1)/2^n)\big] = \big[F^{-1}((j+1)/2^n) - F^{-1}(j/2^n)\big]\big[h(U) - h(D)\big].$$
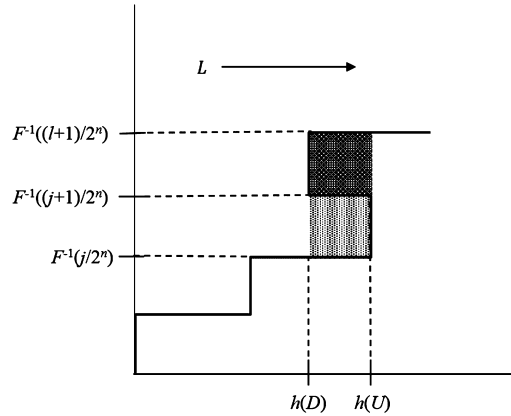
FIGURE 4. Making H^{-1} monotonic.

Using the mean value theorem, the increase in (12) caused by changing the value of $\widetilde{L}$ on intervals $(j/2^n, (j+1)/2^n]$ and $(l/2^n, (l+1)/2^n]$ can be expressed as

$$\Delta(c^{-1})'(\widetilde{L}_l)[F^{-1}((j+1)/2^n) - F^{-1}(j/2^n)][h(U) - h(D)]$$
$$+ \Delta(c^{-1})'(\widetilde{L}_j)[F^{-1}((j+1)/2^n) - F^{-1}(j/2^n)][h(U) - h(D)]$$

for some $\widetilde{L}_l$ between $\widetilde{L}((l+1)/2^n) + \widetilde{\Delta}_l$ and $\widetilde{L}((j+1)/2^n)$, and some $\widetilde{L}_j$ between $\widetilde{L}((j+1)/2^n) + \widetilde{\Delta}_j$ and $\widetilde{L}((l+1)/2^n)$.

The exchange of D and U also affects the integrand of (12) on the intervals higher than $(l/2^n, (l+1)/2^n]$. We can estimate this change in the integrand, as we did for the intervals $(j/2^n, (j+1)/2^n]$ and $(l/2^n, (l+1)/2^n]$, by using the mean value theorem. On each interval $(\bar{y}/2^n, (\bar{y}+1)/2^n]$, where $\bar{y} > l$, the integrand increases by

$$(c^{-1})'(\widetilde{L}_{\bar{y}})\{[F^{-1}((j+1)/2^n) - F^{-1}(j/2^n)][h(U) - h(D)]$$
$$- [F^{-1}((l+1)/2^n) - F^{-1}(l/2^n)][h(U) - h(D)]\}$$

for some $\widetilde{L}_{\bar{y}}$.

Setting $z = (j+1)/2^n$ and dividing the aggregate increase in (12) by $[h(U) - h(D)]$ (which appears in all expressions), we obtain

$$\Delta[(c^{-1})'(\widetilde{L}_l) + (c^{-1})'(\widetilde{L}_j)][F^{-1}(z) - F^{-1}(z - \Delta)]$$
$$- \{[F^{-1}(z + \Delta) - F^{-1}(z)] - [F^{-1}(z) - F^{-1}(z - \Delta)]\} \sum_{\bar{y}=l+1}^{2^n-1} \Delta(c^{-1})'(\widetilde{L}_{\bar{y}}). \qquad (22)$$

By using the mean value theorem twice on (22), once in the first line and once in the second line (for function $g(z) = F^{-1}(z + \Delta) - F^{-1}(z)$ on interval $[z - \Delta, z]$), we obtain

$$\Delta^2 \big[(c^{-1})'(\widetilde{L}_l) + (c^{-1})'(\widetilde{L}_j) \big] \left[\frac{1}{f(F^{-1}(z - \zeta))} \right]$$

$$- \Delta \left[\frac{1}{f(F^{-1}(z + \eta))} - \frac{1}{f(F^{-1}(z - \Delta + \eta))} \right] \sum_{\bar{y}=l+1}^{2^n - 1} \Delta(c^{-1})'(\widetilde{L}_{\bar{y}}), \qquad (23)$$

where $0 \le \zeta, \eta \le \Delta$.

Applying the mean value theorem again, (23) is equal to

$$\Delta^2 \big[(c^{-1})'(\widetilde{L}_l) + (c^{-1})'(\widetilde{L}_j) \big] \left[\frac{1}{f(F^{-1}(z - \zeta))} \right] + \Delta^2 \left[\frac{f'(F^{-1}(z'))}{f^3(F^{-1}(z'))} \right] \sum_{\bar{y}=l+1}^{2^n - 1} \Delta(c^{-1})'(\widetilde{L}_{\bar{y}}) \quad (24)$$

for some $z' \in [z - \Delta + \eta, z + \eta]$. By continuity of f and f', (24) is equal to

$$\Delta^2 \big[(c^{-1})'(\widetilde{L}_l) + (c^{-1})'(\widetilde{L}_j) \big] \left[\frac{1}{f(F^{-1}(z))} \right]$$

$$+ \Delta^2 \left[\frac{f'(F^{-1}(z))}{f^3(F^{-1}(z))} \right] \sum_{\bar{y}=l+1}^{2^n - 1} \Delta(c^{-1})'(\widetilde{L}_{\bar{y}}) + o(\Delta^2), \qquad (25)$$

where $o(\Delta^2)$ is an expression that tends to zero faster than Δ^2.

To determine the sign of (25), consider two cases: (1) If $f'(F^{-1}(z)) \ge 0$, then (25) is positive for sufficiently small Δ's, since its first component is strictly positive, and the second component is nonnegative; (2) If $f'(F^{-1}(z)) < 0$, then the first component is no smaller than

$$\frac{2\Delta^2}{\overline{c} f(F^{-1}(z))},$$

and the second component is no smaller than

$$\Delta^2 \left[\frac{f'(F^{-1}(z))}{f^3(F^{-1}(z))} \right] \sum_{\bar{y}=l+1}^{2^n - 1} \frac{\Delta}{\underline{c}} = \Delta^2 \left[\frac{f'(F^{-1}(z))}{f^3(F^{-1}(z))} \right] \frac{(1 - z)}{\underline{c}} + o(\Delta^2).$$

So, (25) is positive for sufficiently small Δ's by Assumption 2. This completes the proof of (15). □

References

Billingsley, Patrick (1995), *Probability and Measure*. Wiley, New York. [82]

Bulow, Jeremy and Jonathan Levin (2006), "Matching and price competition." *American Economic Review*, 96, 652–668. [59, 63, 68]

Bulow, Jeremy and John Roberts (1989), "The simple economics of optimal auctions." *Journal of Political Economy*, 97, 1060–1090. [61]

Clark, Derek J. and Christian Riis (1998), "Competition over more than one prize." *American Economic Review*, 88, 276–289. [59, 63]

Dasgupta, Ani and Kofi Nti (1998), "Designing an optimal contest." *European Journal of Political Economy*, 14, 587–603. [71]

Fang, Dawei, Thomas Noe, and Philipp Strack (forthcoming), "Turning up the heat: The demoralizing effect of competition in contests." *Journal of Political Economy*. [71]

Fu, Qiang and Jingfeng Lu (2009), "The beauty of 'bigness': On optimal design of multi-winner contests." *Games and Economic Behavior*, 66, 146–161. [70]

Fu, Qiang and Jingfeng Lu (2012), "The optimal multi-stage contest." *Economic Theory*, 51, 351–382. [70]

Fu, Qiang, Jingfeng Lu, and Yue Pan (2015), "Team contests with multiple pairwise battles." *American Economic Review*, 105, 2120–2140. [70]

Glazer, Amihai and Refael Hassin (1988), "Optimal contests." *Economic Inquiry*, 26, 133–143. [69, 70, 71]

Hillman, Arye L. and John G. Riley (1989), "Politically contestable rents and transfers." *Economics and Politics*, 1, 17–39. [68]

Hinnosaar, Toomas (2018), "Optimal sequential contests." Unpublished paper, SSRN 3119014. [71]

Immorlica, Nicole, Greg Stoddard, and Vasilis Syrgkanis (2015), "Social status and badge design." In *Proceedings of the 24th International Conference on World Wide Web*, 473–483, ACM. [69]

Kaplan, Todd R. and Shmuel Zamir (2015), "Advances in auctions." In *Handbook of Game Theory With Economic Applications* (H. Peyton Young and Shmuel Zamir, eds.), 381–453, Chapter 7, North-Holland, Amsterdam, The Netherlands. [70]

Lazear, Edward P. and Sherwin Rosen (1981), "Rank-order tournaments as optimum labor contracts." *Journal of Political Economy*, 89, 841–864. [58]

Milgrom, Paul R. and Robert J. Weber (1985), "Distributional strategies for games with incomplete information." *Mathematics of Operations Research*, 10, 619–632. [72]

Moldovanu, Benny and Aner Sela (2001), "The optimal allocation of prizes in contests." *American Economic Review*, 91, 542–558. [59, 69, 70]

Moldovanu, Benny and Aner Sela (2006), "Contest architecture." *Journal of Economic Theory*, 126, 70–96. [70]

Moldovanu, Benny, Aner Sela, and Xianwen Shi (2007), "Contests for status." *Journal of Political Economy*, 115, 338–363. [69]

Myerson, Roger B. (1981), "Optimal auction design." *Mathematics of Operations Research*, 6, 58–73. [60, 61, 62]

Nti, Kofi (2004), "Maximum efforts in contests with asymmetric valuations." *European Journal of Political Economy*, 20, 1059–1066. [71]

Olszewski, Wojciech and Ron Siegel (2016), "Large contests." *Econometrica*, 84, 835–854. [58, 59]

Olszewski, Wojciech and Ron Siegel (2019), "Equilibrium existence in games with ties." Unpublished paper, Department of Economics, Northwestern University. [72]

Polishchuk, Leonid and Alexander Tonis (2013), "Endogenous contest success functions: A mechanism design approach." *Economic Theory*, 52, 271–297. [71]

Salanié, Bernard (2003), *The Economics of Taxation*. MIT Press. [71]

Spence, Michael A. (1973), "Job market signaling." *Quarterly Journal of Economics*, 87, 355–374. [59]

Tullock, Gordon (1980), "Efficient rent seeking." In *Toward a Theory of the Rent Seeking Society* (James Buchanan, Gordon Tullock, and Robert Tollison, eds.), 269–282, Texas A&M University Press, College Station. [58]

Xiao, Jun (2016), "Asymmetric all-pay contests with heterogeneous prizes." *Journal of Economic Theory*, 163, 178–221. [59]

Xiao, Jun (2017), "Ability grouping in all-pay contests." Unpublished paper, Department of Economics, University of Melbourne. [70]

Xiao, Jun (2018), "Equilibrium analysis of the all-pay contest with two nonidentical prizes: Complete results." *Journal of Mathematical Economics*, 74, 21–34. [69]