Strategic Sample Selection*

Alfredo Di Tillio† Marco Ottaviani‡ Peter Norman Sørensen§

February 26, 2017

Abstract

This paper develops a framework to evaluate the impact of sample selection on the quality of statistical inference. An evaluator tests a hypothesis based on observation of a sample selected as the most favorable of several observations. The impact of this sample selection on the evaluator's payoff is characterized through a generalization of Lehmann's comparison of location experiments. The evaluator benefits from greater selection when the data distribution's quantile density function is less elastic than in Gumbel's extreme value distribution. The evaluator is harmed either when the data distribution has sufficiently thick tails and the hypothesis would be rejected at the prior, or when tails are sufficiently thin and the prior is to accept the hypothesis.

*Research funded by the European Research Council through Advanced Grant 29583 (EVALIDEA). We thank Matteo Camboni for outstanding research assistance. We also thank Pierpaolo Battigalli and participants at seminars in Bruxelles, Leicester, Milan, Nottingham, Surrey, and Toulouse for helpful comments.

†Department of Economics and IGIER, Bocconi University, Via Roberto Sarfatti 25, 20136 Milan, Italy. Phone: +39–02–5836–5422. E-mail: alfredo.ditillio@unibocconi.it.

‡Department of Economics and IGIER, Bocconi University, Via Roberto Sarfatti 25, 20136 Milan, Italy. Phone: +39–02–5836–3385. E-mail: marco.ottaviani@unibocconi.it.

§Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK–1353 Copenhagen K, Denmark. Phone: +45–3532–3056. E-mail: peter.sorensen@econ.ku.dk.

1 Introduction

As econometricians have long recognized, in many instances observational data are nonrandomly selected.[1] Two main selection mechanisms typically operate: either self selection is induced by choices made by the subjects under investigation, or selection originates from sample inclusion decisions made by data analysts carrying out the study. Experimental data can also suffer from selection problems, either because the study population is not representative of the population of interest (challenging external validity),[2] or because the randomized allocation to treatment rather than control is subverted (challenging internal validity).[3] Whatever the source of selection, it is natural to wonder about the impact of sample selection on the quality of statistical inference.

Building on classic hypothesis testing, this paper provides a framework to characterize the information value of a selected observation, relative to a random observation. Consider an evaluator testing a simple hypothesis. Acceptance or rejection of the hypothesis has payoff consequences for the evaluator, depending on the true state. In the hypothesis testing problem, a central role is played by the error rates of false acceptance, α, and false rejection, β. We first note (Lemma 1) that our evaluator's payoff can be rewritten as a linear function of these two error rates. The relative weight attached to α represents the evaluator's preference against false acceptance. Each strategy for mapping realized observations into the evaluator's decision determines a pair of error rates. Efficient use of the evidence provides lowest β for every α, yielding a convex relationship of the best available β as a function of α. We call this relationship the information constraint of the experiment, and we derive its convexity by assuming that

[1] See for example the large literature following Heckman (1979).

[2] In this regard, Alcott (2015) documents the presence of hard-to-control-for selection of the experimental sites in the context of experiments conducted by utility companies to evaluate the effectiveness of energy-savings policies.

[3] See Schulz (1995), Schulz et al. (1995), and Berger (2005) for extensive accounts and examples of subversion of randomization in clinical trials. As explained by Berger (2005), the practice of blocking to ensure an equal number of patients in the control and in the treatment group tends to make allocation to control/treatment more predictable toward the end of the block, allowing researchers to subvert the assignment of individual patients depending on the outcomes they expect for individual patients.

the experimental observation is drawn from a log-concave distribution.[4]

Selection of the observation changes the information constraint and hence the value of the evaluator's problem. In our first main result (Theorem 1) we characterize those distributions for which more selection induces an information constraint that is everywhere lower or everywhere higher than under less selection. We compare the value of a random realization to that of a selected observation of the highest out of k realizations. Our theorem shows that the evaluator's payoff is increasing/decreasing in the amount k of selection when the error distribution has a quantile density (a.k.a. Tukey's sparsity) function that is less/more elastic than Gumbel's extreme value distribution.

The elasticity condition in Theorem 1 allows for a partial comparison of experiments affected by selection via the well known Lehmann's (1988) dispersion order. The evaluator prefers more/less selection for any preference against false acceptance. For many distributions of interest, however, the criterion does not apply, as the evaluator prefers more or less selection depending on the preference parameters. Our second main contribution (Theorem 2) is a generalization of Lehmann's result. Our theorem provides a tool for comparing experiments whose information constraints may exhibit one or more crossings. To compare two experiments for a neighborhood of parameter values, it suffices to check for dispersion in a corresponding neighborhood of the distributions underlying the two experiments.

Using Theorem 2, we then extend the welfare analysis of selection to the cases where a uniform comparison of experiments fails (Proposition 1). If for instance greater selection depresses the quantile density at low quantiles but raises it at high quantiles, then selection harms evaluators with a strong preference against false acceptance, but otherwise benefits them. This means that the evaluator is harmed either when the data distribution has sufficiently thick tails (as for the Laplace distribution) and the hypothesis would be rejected at the prior, or when tails are sufficiently thin (as for the uniform distribution) and the prior is to accept the hypothesis.

[4] Di Tillio, Ottaviani and Sørensen (2015) address another set of questions for a toy model with binary errors (hence not satisfying log-concavity).

We then address how the evaluator is affected when the amount of selection grows extremely large, as k tends to infinity. Drawing on extreme value theory, our third main result (Theorem 3) characterizes the limit impact of selection. For an illustration, with normally distributed errors, we show that the evaluator in the limit is able to identify the true state on the basis of one, extremely selected observation. In this case, the evaluator thus obtains the highest possible payoff where both error rates are zero. By contrast, with exponentially distributed errors, the quantile density is more elastic than in the Gumbel case, so selection increasingly harms the evaluator (Theorem 1) and the limit information is less than full.

We conclude by applying our main results to the positive and normative analysis of strategic selection by a researcher aiming at demonstrating that a treatment is effective. The researcher's incentives to bias upward the evaluator's inference are anticipated in equilibrium by the evaluator. Under the global elasticity conditions of Theorem 1, satisfied for example if the errors are normally distributed, we show that equilibrium selective sampling benefits also the researcher in the empirically relevant case where the prior strongly favors rejection.[5] Instead, when the errors have thick tails and selection is mild (small k), equilibrium selective sampling harms not only the evaluator but also the researcher when the prior strongly favors rejection—generating a credibility crisis.[6]

In a complementary approach to modeling conflicts of interest in statistical testing, Tetenov (2016) analyzes a regulator's optimal commitment to a decision rule when privately informed proponents of innovations select into costly testing. Instead, we focus on the impact of a researcher's manipulation of the data on the welfare of an uncommitted evaluator. Henry and Ottaviani (2015) analyze a dynamic model of persuasion with costly information acquisition à la Wald (1950), where information is truthfully reported at the time of application.

[5] In this case, we also show that equilibrium selection harms the researcher when the prior strongly favors acceptance.

[6] In this case, we also show that equilibrium selection benefits the researcher when the prior strongly favors acceptance. The researcher is harmed by selection for intermediate priors.

2 Statistical Model

An evaluator is interested in the true value of a binary state $\theta \in \{\theta_L, \theta_H\}$. It is natural to interpret the two states as the true/false outcome of a hypothesis. The level of the state θ may correspond to a treatment effect, where $\theta = \theta_H$ is equivalent to a good treatment. We let $\theta_H > \theta_L$ be real numbers. The evaluator behaves like a Bayesian, and has prior belief $\Pr(\theta = \theta_H) = q \in (0,1)$. We compare two scenarios:

Random Sample In the first scenario, the evaluator observes one signal realization x of the location experiment

$$x = \theta + \varepsilon,$$

where θ is the true state and the unobservable sampling error ε is a real random variable drawn from a known distribution F.

Selected Sample In the second scenario, the signal realization observed by the evaluator is the maximum of k independent and identically distributed draws:

$$x_{(k)} = \theta + \varepsilon_{(k)},$$

with $\varepsilon_{(k)} = \max\{\varepsilon_1, ..., \varepsilon_k\}$, and each ε_i is drawn from the same distribution function F. Moreover, θ and $(\varepsilon_1, \ldots, \varepsilon_k)$ are stochastically independent. The distribution of $\varepsilon_{(k)}$ is F^k, with density kfF^{k-1}.

After having observed the signal, the evaluator updates the prior belief into a posterior, and then either accepts or rejects. Acceptance results in payoff θ, while rejection gives the safety payoff R, a real number satisfying $\theta_L < R < \theta_H$.

For an application, we may think of ε_i as the health outcome of an individual i when not receiving the treatment, and $x_i = \theta + \varepsilon_i$ as the health outcome of the same individual when receiving the treatment. The assumption that F is known then corresponds to assuming that many prior experiments have uncovered the distribution of baseline health outcomes in the population.

In the first scenario, the evaluator's problem is in line with Lehmann (1988). This setting is our benchmark to evaluate the impact of sample selection (second scenario). The main difference between the two scenarios is the distribution of the random error in the signal observed. Section 6 justifies the second scenario in terms of a game-theoretical setting in which the evidence is provided by a researcher who is interested in obtaining acceptance.

Our goal is to analyze how changes in k affect the evaluator's payoff in the second scenario, depending on the exogenously given distribution F. Again, the first scenario will be used as a benchmark both to formalize the evaluator's decision problem and to provide welfare results.

2.1 Evaluator's Decision Problem

We start by a reformulation of the evaluator's decision problem. We provide a general characterization of the evaluator's payoff when observing a given signal x. In our binary choice problem, we show that the decision may be viewed as a simple trade-off between correct choices—rejection in state θ_L and acceptance in state θ_H.

Assumption 1. *The c.d.f. F has a derivative f which is log-concave, i.e.,* $\log(f)$ *is a concave function.*

Our assumption implies that F is continuous with an interval support, and F^{-1} is well-defined. It also implies the monotone likelihood ratio property: the posterior belief that the state is θ_H after observing $x = \theta + \varepsilon$ with ε drawn from F,

$$\frac{qf(x-\theta_H)}{qf(x-\theta_H)+(1-q)f(x-\theta_L)},$$

is increasing in x.[7] The optimal decision rule is therefore also monotone in x, and the evaluator's problem thus amounts to choosing a threshold $\hat{x}$ such that there is acceptance if and only if $x \geq \hat{x}$.

[7]Since we are dealing with a location experiment, the monotone likelihood ratio property is not only necessary but also sufficient for log-concavity of the error distribution

Any given threshold $\hat{x}$ induces a (correct) rejection probability $1-\alpha = F(\hat{x}-\theta_L)$ in state θ_L and a (correct) acceptance probability $1-\beta = 1-F(\hat{x}-\theta_H)$. This defines a function relating α to β. We denote this function $\beta_F(\alpha)$ and call it the *information constraint* of experiment F. Clearly,

$$\beta_F(\alpha) = F\left(F^{-1}(1-\alpha)+\theta_L-\theta_H\right).$$

We refer to the function $\beta_F(\alpha)$ as the *information constraint* of experiment F.

Lemma 1. *The evaluator's problem is equivalent to choosing* $\alpha,\beta \in [0,1]$ *in order to minimize* $(1-q)(R-\theta_L)\alpha + q(\theta_H - R)\beta$ *subject to* $\beta = \beta_F(\alpha)$.

Proof. As a function of the threshold $\hat{x}$, the evaluator's payoff is

$$\begin{aligned}&\sum_\theta \Pr(\theta)\left\{\left[1-F\left(\hat{x}-\theta\right)\right]\theta + F(\hat{x}-\theta)R\right\}\\ &\qquad = qF\left(\hat{x}-\theta_H\right)\left(R-\theta_H\right) + (1-q)\left[1-F\left(\hat{x}-\theta_L\right)\right]\left(\theta_L - R\right) + q\theta_H + (1-q)R.\end{aligned}$$

The constant $q\theta_H + (1-q)R$ is irrelevant for the maximization of the payoff. Using the definitions of α and β, we conclude that the optimality criterion is to minimize $(1-q)(R-\theta_L)\alpha + q(\theta_H - R)\beta$. Finally, the definitions of α and β can be rewritten as the constraint $F^{-1}(1-\alpha)+\theta_L = \hat{x} = F^{-1}(\beta)+\theta_H$, that is, $\beta = \beta_F(\alpha)$. $\square$

Log-concavity guarantees that this minimization problem is well behaved, as the information constraint is convex. Along the constraint,

$$\frac{d\beta}{d\alpha} = -\frac{f\left[F^{-1}(\beta)\right]}{f\left[F^{-1}(1-\alpha)\right]} = -\frac{f\left(\hat{x}-\theta_H\right)}{f\left(\hat{x}-\theta_L\right)},$$

which is negative and increasing in α because α is decreasing in $\hat{x}$. Moreover, if this likelihood ratio is unbounded (tending to 0 as $\hat{x} \to -\infty$ and to ∞ as $\hat{x} \to \infty$) then $d\beta/d\alpha \to -\infty$ as $\alpha \to 0$ and $d\beta/d\alpha \to 0$ as $\alpha \to 1$. This is illustrated in the left panel of Figure 1, where $\theta_L = 0$, $\theta_H = 1$, $q = 1/2$, $R = 1/2$ and the signal distribution F is the standard normal, $N(0,1)$. The black lines are iso-payoff lines, the blue curve is the information constraint. The right panel is the analogous picture for an error uniformly

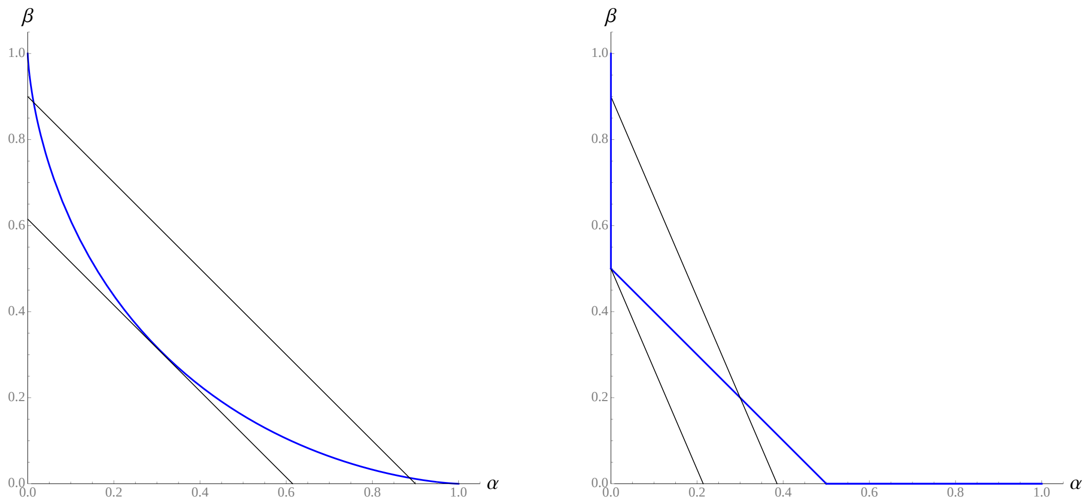

Figure 1: Normal (left) and uniform (right) signal.

distributed $U[-1,1]$, with the same parameters other than $R = 7/10$ instead of $R = 1/2$, which gives a corner solution with $\alpha = 0$.[8]

2.2 Comparing Experiments

Consider two location experiments with log-concave error distributions F and G, and let $\beta_F(\alpha)$ and $\beta_G(\alpha)$ denote their respective information constraints. An immediate separation argument shows that the evaluator is better off with the location experiment induced by G than with the one induced by F, uniformly over q and R, if and only if $\beta_G(\alpha) < \beta_F(\alpha)$ for every α. In this case, we say that G is *more effective* than F (relative to θ_L and θ_H).

Lehmann (1988) considered a similar problem—more general as θ need not be binary—and proved that G is more effective than F, relative to *every* pair θ_L, θ_H with

[8] The constraint connecting β to α is known in the statistical literature under various names. Jewitt (2007) refers to a probability-probability plot which shows β as a function of $1-\alpha$. Torgerson (1991) uses both labels β-functions and Neyman-Pearson functions to describe power $1-\beta$ as a function of significance α. The receiver operating characteristic curve likewise plots the true positive rate $1-\beta$ as a function of the false positive rate α.

$\theta_L < \theta_H$, if and only if the quantile difference $G^{-1}(u) - F^{-1}(u)$ is a weakly decreasing function of u, that is,

$$f\big(F^{-1}(u)\big) \leq g\big(G^{-1}(u)\big) \qquad \text{for all } 0 < u < 1. \tag{1}$$

This criterion defines that G is *less dispersed* than F, a notion of stochastic ordering already proposed by Bickel and Lehmann (1979).[9] Intuitively, let $0 < u < v < 1$ and suppose that in experiment F we use a threshold resulting in type I error $\alpha = 1 - v$ and type II error $\beta = u$. Then the inequality $G^{-1}(u) - F^{-1}(u) < G^{-1}(v) - F^{-1}(v)$ says that the threshold adjustment needed to induce the same α under G, is smaller than the one needed to induce the same β, and therefore it achieves a lower β.

3 Dispersion Ordering of Order Statistics

Recall that our goal is to compare the experiment induced by a generic F with the one induced by a selected sample from F, that is, the experiment generated by F^k. Towards this goal, we first seek a characterization of distributions F for which the number k has a *monotone* effect on the evaluator's payoff, uniformly over parameter constellations.

Lehmann's condition turns on the quantile function, and it is convenient to introduce notation for this. We denote it by $\Xi(u) = F^{-1}(u)$. Its derivative is the quantile density $\xi(u) = 1/f\left(F^{-1}(u)\right)$. It can be directly computed that, when ξ is differentiable,

$$-\frac{\xi'(u)}{\xi(u)} = \frac{f'\left[F^{-1}(u)\right]}{f^2\left[F^{-1}(u)\right]}.$$

By definition, the elasticity of the quantile density function is $-u\xi'(u)/\xi(u)$.

An important special role is played by Gumbel's extreme value distribution, $F(\varepsilon) = \exp(-\exp(-\varepsilon))$. Its quantile function is $\Xi(u) = -\log(-\log u)$ and hence $\xi(u) = -1/(u\log u)$. Thus, the elasticity of its quantile density function is

$$-\frac{u\xi'(u)}{\xi(u)} = \frac{u(1+\log u)/(u\log u)^2}{1/(u\log u)} = \frac{1+\log u}{\log u}.$$

[9]They, in turn, credit Brown and Tukey (1946) for the essence of this definition.

Theorem 1. *Let F have a differentiable quantile density function ξ. If the elasticity of ξ is everywhere greater than that of the Gumbel distribution, then F^k is more effective than F, the greater the k. If, instead, this elasticity is everywhere less than that of the Gumbel distribution, then F^k is less effective than F, the greater the k.*

Proof. Fix the natural number k, and denote $G = F^{k+1}$ and $H = F^k$. By Lehmann's condition (1), G is more effective than H if $G^{-1} - H^{-1}$ is decreasing. Observe that $\varepsilon = H^{-1}(v)$ if and only if $F^k(\varepsilon) = v$, or $F(\varepsilon) = v^{1/k}$. In particular, $H^{-1}(v) = \varepsilon = F^{-1}\left(v^{1/k}\right)$. By the implicit function theorem, the slope of H^{-1} is

$$\frac{d\varepsilon}{dv} = \frac{(1/k)\, v^{(1-k)/k}}{f\left(H^{-1}(v)\right)} = (1/k)\, v^{(1-k)/k} \xi\left(v^{1/k}\right).$$

It suffices to check whether this slope is a monotone function of the real number $k \geq 1$ — since natural numbers are real, it follows that the slope of G^{-1} can be compared to that of H^{-1}. Taking the logarithm, we need to check when

$$\log\left(\frac{1}{k}\right) + \frac{1-k}{k}\log(v) + \log\left(\xi\left(v^{1/k}\right)\right)$$

is monotone in k. The derivative of this is

$$-\frac{1}{k} - \frac{1}{k^2}\log(v) - \frac{\xi'\left(v^{1/k}\right)}{\xi\left(v^{1/k}\right)} \frac{v^{1/k}\log(v)}{k^2}$$

Change variables to $u = v^{1/k} \in (0,1)$, and note that this derivative is non-positive if and only if

$$-\frac{\xi'(u)}{\xi(u)} u \log(u) \leq 1 + \log(u).$$

□

An immediate implication of our characterization is that selection has no effect in the case of the Gumbel distribution. This has an intuitive explanation. When selecting the best of k, $F^k(\varepsilon) = \exp(-k\exp(-\varepsilon)) = F(\varepsilon - \log(k))$. The location experiment $x = \theta + \varepsilon$ is unaffected by this translation of the noise distribution.

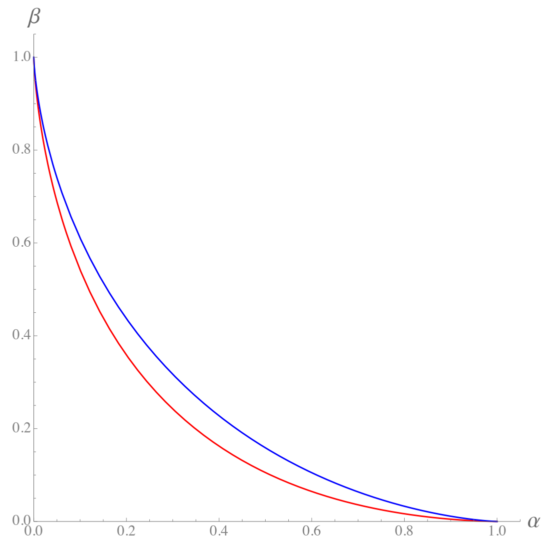

Figure 2: Comparison between F (blue) and F^2 (red) for the normal distribution.

For the normal distribution, it can be verified that the elasticity condition in the theorem is met, and hence more selection benefits the evaluator. A plot of the information constraints corresponding to F and F^2 illustrates this in Figure 2.

If the distribution is logistic, $F(\varepsilon) = 1/(1+e^{-\varepsilon})$, we have $\Xi(u) = \log[u/(1-u)]$ and hence $\xi(u) = 1/[u(1-u)]$. Thus

$$-\frac{u\xi'(u)}{\xi(u)} = \frac{1-2u}{1-u} > \frac{1+\log u}{\log u},$$

so ξ is more elastic than the Gumbel distribution. Here, too, any amount of selection benefits the evaluator, and benefits even more as k increases.

For the exponential distribution, instead, $F(\varepsilon) = 1-e^{-\varepsilon}$ where $\varepsilon \geq 0$, so $\Xi(u) = -\log(1-u)$ and hence $\xi(u) = 1/(1-u)$. Thus

$$-\frac{u\xi'(u)}{\xi(u)} = -\frac{u}{1-u} < \frac{1+\log u}{\log u}$$

so ξ is less elastic than the Gumbel distribution—any amount of selection harms the evaluator, ever more as k increases. It should be noted that signal realizations $x < \theta_H$

perfectly reveal state θ_L, so $\beta(\alpha) = 0$ where $\alpha \geq 1 - F^k(\theta_H - \theta_L)$.[10]

Finally, note that with the change of variable $u = F(\varepsilon)$, the condition for F^k to become more effective with k can be restated as

$$\frac{f'(\varepsilon)/f(\varepsilon)}{f(\varepsilon)/F(\varepsilon)} > \frac{1 + \log F(\varepsilon)}{\log F(\varepsilon)}.$$

4 Local Comparison of Experiments

Lehmann's comparison gives only a partial ordering of location experiments, because it imposes a uniform criterion. For all $\theta_L < \theta_H$ the information constraint $\beta_G(\alpha)$ must lie entirely below $\beta_F(\alpha)$, in order for G to be more effective than F. In many cases of interest, this property is not satisfied, even for fixed θ_L and θ_H. For example, as we shall see, the property fails when we compare F with $G = F^k$ and F is, e.g., the Laplace or the uniform distribution. In this section we show that a useful criterion for comparing experiments with crossing information constraints can still be given, in terms of a notion that we call *local dispersion*.

4.1 Local Dispersion

Recall that G is more effective than F relative to every θ_L and $\theta_H > \theta_L$ if and only if the quantile difference $G^{-1}(u) - F^{-1}(u)$ is decreasing in u. This equivalence is at the heart of the characterization provided in Theorem 1. What if the slope of the quantile difference is not everywhere negative or everywhere positive? In this case, for some values of θ_L and θ_H, the information constraints of G and F must cross at least once. This means that, for those values of θ_L and θ_H, whether G or F is preferred depends on the values of q and R.[11]

[10] Boland et al. (1995) compare dispersion of order statistics from the exponential distribution. Kochar (1996) extends this to distributions with decreasing failure rate.

[11] Graphically, it is easy to see when experiment G is preferred to F for fixed parameter values. Let $\beta_{F\vee G}$ denote the convex envelope of β_F and β_G, that is, the largest convex function that is no greater than β_F or β_G. Let α^* denote a point at which the objective function $(1-q)(R-\theta_L)\alpha + q(\theta_H - R)\beta$

Generalizing Lehmann's (1988) main theorem, our next result shows that even in those cases where Lehmann's uniform criterion is not applicable, the pattern of $G^{-1}(u) - F^{-1}(u)$ still determines the pattern of preference over G and F. The key is that, while the quantile difference may not be everywhere increasing or everywhere decreasing, its slope in a neighborhood around a particular value u still characterizes the preference over F and G in a certain corresponding region of the parameters θ_L, θ_H, q, and R. Intuitively, and in a sense that our theorem makes precise, the decreasing (increasing) parts of the quantile difference correspond to those regions of parameters where G is preferred to F (resp. F is preferred to G).

Theorem 2. *Let $N \geq 1$ and $0 = u_1 \leq \cdots \leq u_{2N+1} = 1$. Suppose that, for all $n = 1, \ldots, N$,*

$$\begin{aligned} f\big(F^{-1}(u)\big) &\leq g\big(G^{-1}(u)\big) \qquad \text{for all } u \in (u_{2n-1}, u_{2n}), \\ f\big(F^{-1}(u)\big) &\geq g\big(G^{-1}(u)\big) \qquad \text{for all } u \in (u_{2n}, u_{2n+1}), \end{aligned}$$

Suppose also that, for all $m = 1, \ldots, 2N$,

$$\theta_H - \theta_L \leq \max\Big\{F^{-1}(u_{m+1}) - F^{-1}(u_m)\,,\; G^{-1}(u_{m+1}) - G^{-1}(u_m)\Big\}.$$

Then there exist $0 = K_1 \leq \cdots \leq K_{2N+1} = \infty$ such that, for all $n = 1, \ldots N$,

$$\begin{aligned} &G \text{ is preferred to } F \quad \text{if} \quad \frac{1-q}{q}\frac{R-\theta_L}{\theta_H - R} \in \big(K_{2n-1}, K_{2n}\big), \\ &F \text{ is preferred to } G \quad \text{if} \quad \frac{1-q}{q}\frac{R-\theta_L}{\theta_H - R} \in \big(K_{2n}, K_{2n+1}\big). \end{aligned}$$

Proof. It suffices to show that there exist $0 = \alpha_{2N+1} \leq \cdots \leq \alpha_1 = 1$ such that, for all $n = 1, \ldots, N$, we have $\beta_G(\alpha) \leq \beta_F(\alpha)$ for all $\alpha \in (\alpha_{2n}, \alpha_{2n-1})$ and $\beta_G(\alpha) \geq \beta_F(\alpha)$ for all $\alpha \in (\alpha_{2n+1}, \alpha_{2n})$.

Define, for every $v \in (0,1)$,

$$u(v) = \max\big\{\beta_F(1-v)\,,\, \beta_G(1-v)\big\}.$$

is minimal on the curve $\beta_{F\vee G}(\alpha)$. Clearly, if $\beta_{F\vee G}(\alpha^*) = \beta_F(\alpha^*)$ then F is at least as good as G for the evaluator. Likewise, if $\beta_{F\vee G}(\alpha^*) = \beta_G(\alpha^*)$, then G is at least as good as F. Finally, if neither $\beta_{F\vee G}(\alpha^*) = \beta_F(\alpha^*)$ nor $\beta_{F\vee G}(\alpha^*) = \beta_G(\alpha^*)$, then F and G are equally good.

The bound on the difference $\theta_H - \theta_L$ implies that there exist $0 = v_1 \leq \cdots \leq v_{2N+1} = 1$ such that, for all $n = 1, \ldots, N$,

$$
\begin{aligned}
G^{-1}(v) - F^{-1}(v) &\leq G^{-1}(u(v)) - F^{-1}(u(v)) \qquad \text{for all } v \in (v_{2n-1}, v_{2n}), \\
G^{-1}(v) - F^{-1}(v) &\geq G^{-1}(u(v)) - F^{-1}(u(v)) \qquad \text{for all } v \in (v_{2n}, v_{2n+1}).
\end{aligned}
$$

But for each $\alpha \in (0,1)$ we have

$$
G^{-1}(1-\alpha) - F^{-1}(1-\alpha) = G^{-1}(\beta_G(\alpha)) - F^{-1}(\beta_F(\alpha))
$$

and hence the inequality $\beta_G(\alpha) \leq \beta_F(\alpha)$ is equivalent to both of the following:

$$
\begin{aligned}
G^{-1}(1-\alpha) - F^{-1}(1-\alpha) &\leq G^{-1}(\beta_F(\alpha)) - F^{-1}(\beta_F(\alpha)), \\
G^{-1}(1-\alpha) - F^{-1}(1-\alpha) &\leq G^{-1}(\beta_G(\alpha)) - F^{-1}(\beta_G(\alpha)).
\end{aligned}
$$

We conclude that, for all $n = 1, \ldots, N$,

$$
\begin{aligned}
\beta_G(\alpha) &\leq \beta_F(\alpha) \qquad \text{for all } \alpha \in (1 - v_{2n}, 1 - v_{2n-1}) =: (\alpha_{2n}, \alpha_{2n-1}), \\
\beta_G(\alpha) &\geq \beta_F(\alpha) \qquad \text{for all } \alpha \in (1 - v_{2n+1}, 1 - v_{2n}) =: (\alpha_{2n+1}, \alpha_{2n}).
\end{aligned}
$$

□

This theorem implies that, when the difference $\theta_H - \theta_L$ is not too large, the slope of $G^{-1}(u) - F^{-1}(u)$ for large, intermediate, or small values of u determines the evaluator's preferences over G and F when the relative weight attached to type I errors, the ratio $(1-q)(R-\theta_L)/[q(\theta_H - R)]$, is high, intermediate, or low, respectively. For example, suppose that we would like to check whether G is preferred to F for a given q and for all R above a certain value (or for a given R and for all q below some value). The theorem highlights that checking whether this preference indeed holds only requires us to verify that $G^{-1}(u) - F^{-1}(u)$ is decreasing for u above a certain bound. Perhaps remarkably, the same preference would then hold for any other distributions G' and F' with the same shape as G and F above that bound.

The bound on the difference $\theta_H - \theta_L$ in the theorem cannot be dispensed with. Indeed, recall that the fact that the slope of the quantile difference does not have a constant sign only implies that for *some* values of θ_L and θ_H the information constraints of G and F must cross. Our theorem shows that when θ_L and θ_H are sufficiently close, the information constraints do indeed cross. However, it is easy to construct examples where θ_L and θ_H are far apart, and one experiment is preferred to the other for all q, R (and those θ_L and θ_H) despite the fact that the quantile difference is not monotone.

4.2 Local Ordering of Order Statistics

Thanks to the local condition developed above, we can now refine Theorem 1 by considering the following immediate consequence of Theorem 2. Afterwards, we will illustrate this method for the case of Laplace and uniform distributions.

Proposition 1. *Let $\bar{u} \in (0,1)$ and suppose that $G^{-1}(u) - F^{-1}(u)$ is decreasing (resp. increasing) for $u \in (0,\bar{u})$ and increasing (resp. decreasing) for $u \in (\bar{u},1)$. Then there exists $\delta > 0$ and $a \geq 0$ such that, if $\theta_H - \theta_L \leq \delta$, then selection benefits (resp. harms) the evaluator for $(1-q)(R-\theta_L) < aq(\theta_H - R)$ and harms (resp. benefits) the evaluator for $(1-q)(R-\theta_L) > aq(\theta_H - R)$.*

Intuitively, we may describe $(F^k)^{-1}(u) - F^{-1}(u)$ as U-shaped when it has the first property assumed in the theorem, and bell-shaped in the second case. For an illustration of the first case, suppose that $k = 2$ and consider the Laplace distribution where $F(\varepsilon) = (1/2)e^{\varepsilon}$ for $\varepsilon < 0$ and $F(\varepsilon) = 1 - (1/2)e^{-\varepsilon}$ for $\varepsilon \geq 0$. In the left-hand side panel of Figure 3 we plot F (blue) and F^2 (red), which reveals the U-shape of $(F^2)^{-1}(u) - F^{-1}(u)$. In this example, there is a bound to the posterior belief after seeing any signal realization, and hence the information constraints are not tangent to the axes as $\alpha \to 0$ and $\alpha \to 1$. As shown in the right-hand side panel of Figure 3, drawn for $\theta_L = 0$ and $\theta_H = 1$, the evaluator prefers F to F^2 when the relative weight attached to type I errors—the ratio $(1-q)(R-\theta_L)/[q(\theta_H - R)]$—is large. Indeed, this correspond to steeper iso-payoff lines and hence lower optimal values of α. On the other hand, F^2 is better than F for smaller values of the ratio above.

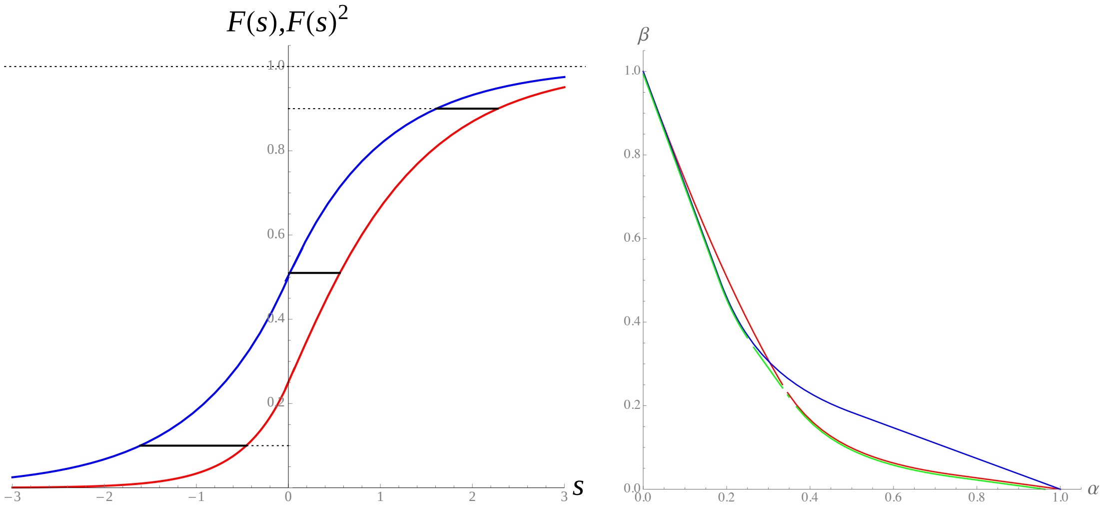

Figure 3: Comparison between F (blue) and F^2 (red) for the Laplace distribution.

Next, consider the uniform distribution, $F(\varepsilon) = \varepsilon$ for $\varepsilon \in [0,1]$. Suppose that $\theta_H < \theta_L + 1$, so there is some overlap of the supports of signal distributions $F(x - \theta_L)$ and $F(x - \theta_H)$. Since $F^2(\varepsilon) = \varepsilon^2$, it is simple to eyeball that $(F^2)^{-1}(u) - F^{-1}(u)$ is bell-shaped, as illustrated in the left-hand side panel of Figure 4. This is a case where greater selection is better for the evaluator for large values of $(1-q)(R-\theta_L)/[q(\theta_H - R)]$, but worse for small values, as illustrated in the right-hand side panel of Figure 4. It should be kept in mind that signals $x < \theta_H$ reveal state θ_L, so $\beta(\alpha) = 0$ where $\alpha \geq 1 - (\theta_H - \theta_L)^k$. Likewise, signals $x > \theta_L + 1$ reveal state θ_H, so we have a vertical part of the curve where $\alpha = 0$ when $\beta \geq (\theta_L + 1 - \theta_H)^k$.

5 Extreme Selection

In this section, we examine the effect of extreme selection, $k \to \infty$, on the evaluator's payoff. To gain some initial intuition, recall from Theorem 1 that in the Gumbel distribution, the evaluator's payoff is constant in k. Next, the fundamental result in extreme value theory says that the distribution of the maximum of k i.i.d. random variables, properly adjusted for location and scale inflation, either does not converge weakly to

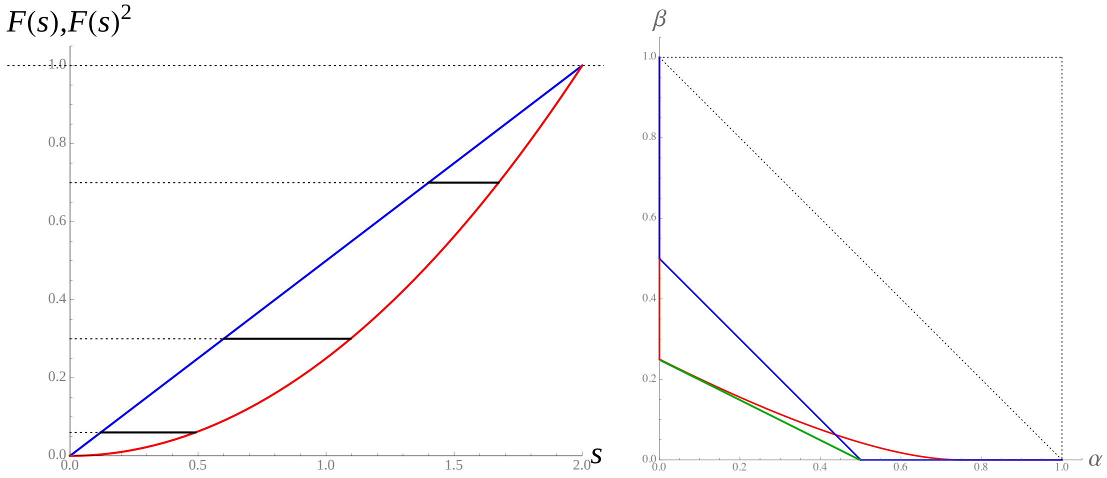

Figure 4: Comparison between F (blue) and F^2 (red) for the uniform distribution.

any nondegenerate distribution (for any choice of adjustment) or it converges weakly to a distribution $\bar{F}$ that must belong to one of the following three types: Gumbel, Extreme Weibull or Frechet. (See e.g. Leadbetter et al. (1983) for a reference.) More precisely, for some sequence $a_k > 0$ and b_k,

$$F^k(b_k + a_k\varepsilon) \to \bar{F}(\varepsilon)$$

for every continuity point ε of $\bar{F}$.

Note that F^k is decreasing in k wherever $F \in (0,1)$. Hence, the distribution of ε is systematically shifted upwards as k increases, in the sense of first-order stochastic dominance. Hence, the location adjustment sequence $-b_k$ is growing. However, the evaluator can adjust for any translation of the error distribution without any impact on payoff.

The limit impact of selection thus hinges on whether the sequence a_k shrinks to zero or not. If $a_k \to 0$, the error distribution is less and less dispersed, providing the evaluator with arbitrarily precise information about the state.

To formalize this claim, denote by V^k the payoff of the evaluator when selection is at level k.

Theorem 3. *If $F^k(b_k + a_k\varepsilon) \to \bar{F}(\varepsilon)$ at every continuity point of $\bar{F}$, with $a_k \to 0$, then $V^k \to q\theta_H + (1-q)R$, the full information payoff. If $a_k = 1$ for all k, then V^k converges to the payoff from observing one non-selected observation with error distribution $\bar{F}$.*

It is well known that many familiar distributions are in the basin of attraction of the Gumbel distribution. Specifically, when F is normal—or half-normal, which has the same right tails—then a_n must be decreasing to zero ($a_n = (2\log n)^{-1/2}$ works), and $\bar{F}$ is the Gumbel distribution. More generally, when F is exponential power (or folded exponential power) with shape parameter $b > 1$, a_n must be decreasing to zero ($a_n = (b\log n)^{-(b-1)/b}$ works, as we show in the appendix), and $\bar{F}$ is the Gumbel distribution. This result is striking, because it is also known that when F is the exponential distribution—or the Laplace distribution, since the two distributions have the same right tails—then it also converges to the Gumbel distribution, but we can take $a_k = 1$ for each k. That is, at least within the rather large exponential power family of distributions, extreme selection leads to full information whenever the parameter value $b > 1$, while it does not when $b = 1$. The uniformly negative impact of selection in the exponential case discussed earlier is, in this sense, non-generic, as any arbitrarily close distribution in the family reverses the conclusion.

6 Strategic Selection

Sample selection of the sort considered above naturally arises as an equilibrium phenomenon in a strategic setting where the experiment is carried out by a researcher who is less worried than the evaluator about false acceptances (the type I error α). Intuitively, a researcher biased toward acceptance may want to select an individual with a high error term—e.g., in the treatment effect setting, a good untreated outcome—in order to bias upward the experimental result and thus increase the chances of acceptance. As we have seen, an evaluator taking this behavior into account may suffer or benefit, compared to the case of a random sample. In this section we discuss how the researcher's ability to strategically select the sample affects the researcher's own welfare, too.

To illustrate the issues in a simple setup, consider the following timeline:

1. The researcher privately observes $\varepsilon_1, \dots, \varepsilon_k$ and then chooses $i \in \{1, \dots, k\}$.
2. The evaluator observes $x_i = \theta + \varepsilon_i$ and then chooses whether to accept or reject.

As before, the evaluator receives a fixed payoff R when rejecting, and θ when accepting. The researcher receives 0 if the evaluator rejects, and 1 if the evaluator approves.

The game described above has a unique (Bayes Nash) equilibrium where the evaluator follows a cutoff rule.[12] In this equilibrium, the researcher chooses $i \in \arg\max_{1 \le j \le k} \varepsilon_j$ and hence the evaluator's threshold $\hat{x}_k$ is defined by

$$\frac{F^{k-1}(\hat{x}_k - \theta_H) f(\hat{x}_k - \theta_H)}{F^{k-1}(\hat{x}_k - \theta_L) f(\hat{x}_k - \theta_L)} = K.$$

This is also the unique equilibrium in which the researcher selects the same order statistic (in equilibrium, the highest) for each $\varepsilon_1, \dots, \varepsilon_k$.[13] It is also worth remarking that, even without assuming equilibrium, the outcome we described is the only outcome compatible with the assumption that (i) both players are rational, (ii) the researcher believes that the evaluator follows a cutoff rule, and (iii) the evaluator believes in (i) and (ii).

How are the two players affected by the researcher's strategic sample selection? For fixed values of q, θ_L and θ_H, we can assess the impact of selection on both the evaluator's and the researcher's welfare, as R varies between θ_L and θ_H (and hence as the

[12] As usual, uniqueness has to be qualified by the fact that any other strategy profile where, e.g., the researcher behaves differently on a measure zero set of vectors $(\varepsilon_1, \dots, \varepsilon_k)$, is also an equilibrium.

[13] Since the density function $f(x)$ is log-concave, both the cumulative distribution function $F(x)$ and the reliability function $1 - F(x)$ are log-concave. Moreover, the product of log-concave functions is log-concave. Thus, the density function of the mth smallest of k such (iid) random variables,

$$\frac{k!}{(m-1)!(k-m)!} F^{m-1}(x) \left[1 - F(x)\right]^{k-m} f(x),$$

is log-concave, which implies that the evaluator follows a cutoff rule. But then the best response of the researcher is to select the individual with highest error term, as its distribution first-order stochastically dominates that of any other order statistic.

relative weight attached to type I errors, the ratio $(1-q)(R-\theta_L)/[q(\theta_H-R)]$, varies between 0 and ∞), by comparing the c.d.f. of the posterior expectation of θ corresponding to $k>1$ with the one corresponding to $k=1$. Given $k\geq 1$ and an observation x, the posterior expectation is

$$\pi_k(x) := \frac{qF^{k-1}(x-\theta_H)f(x-\theta_H)\theta_H + (1-q)F^{k-1}(x-\theta_L)f(x-\theta_L)\theta_L}{qF^{k-1}(x-\theta_H)f(x-\theta_H) + (1-q)F^{k-1}(x-\theta_L)f(x-\theta_L)},$$

with c.d.f.

$$\tilde{F}_k(p) := \int 1_{\{\pi_k(x)\leq p\}} dF^k(x).$$

The evaluator accepts if and only if $x \geq \hat{x}_k$, which is the same as $\pi_k(x)\geq R$. Thus, at safety payoff R, the researcher's expected payoff is

$$U^k(R) := 1-\tilde{F}_k(R),$$

while the evaluator's expected payoff is

$$V^k(R) := R + \int_R^{\theta_H}[1-\tilde{F}_k(r)]dr.$$

When the elasticity condition in Theorem 1 is satisfied, $V^k(R)$ is increasing in k for every R. But the expectation of π_k is the same for every k, because it must equal the prior expectation of θ. In other words, for every k we have

$$\int_{\theta_L}^{\theta_H}[1-\tilde{F}_k(r)]dr = q\theta_H + (1-q)\theta_L.$$

Thus, in terms of the c.d.f.s of the posterior expectations, we must have $\tilde{F}_{k'}(r)\geq \tilde{F}_k(r)$ for all $k'\geq k$ and all r sufficiently close to θ_L and $\tilde{F}_{k'}(r)\leq \tilde{F}_k(r)$ for all r sufficiently close to θ_H. Clearly, the opposite must hold when the elasticity condition is violated. The following result summarizes these facts.

Proposition 2. *Let F have a differentiable quantile density function ξ. If the elasticity of ξ is everywhere greater (resp. everywhere less) than that of the Gumbel distribution, then $V^k(R)$ is increasing (resp. decreasing) in k for every R. Moreover, for every k,k' with $k'\geq k$ there exist R',R'' such that $U^{k'}(R)\leq U^k(R)$ for $R\in[\theta_L,R']$ and $U^{k'}(R)\geq U^k(R)$ for $R\in[R'',\theta_H]$ (resp. $U^{k'}(R)\geq U^k(R)$ for $R\in[\theta_L,R']$ and $U^{k'}(R)\leq U^k(R)$ for $R\in[R'',\theta_H]$).*

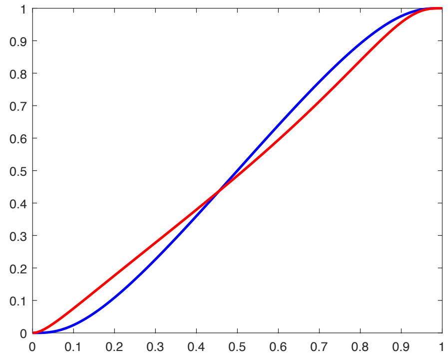

Figure 5: Welfare impact of selection in the normal case

In many special cases of interest, for $k' \geq k$ the c.d.f.s of $\pi_{k'}$ and π_k cross only once—other than at the extremes $R = \theta_L$ and $R = \theta_H$—and hence the thresholds R' and R'' in Proposition 2 coincide. When our elasticity condition holds, the crossing occurs from above. The evaluator benefits from a larger k for every R, whereas the researcher benefits or loses according to whether R is above or below a threshold. This happens, for instance, in the case of a normally distributed error (cf. Figure 2), as illustrated in Figure 5 for $k = 1$ (blue) and $k = 2$ (red), with $\theta_L = 0$ and $\theta_H = 1$.

When greater selection is not uniformly better for the evaluator, the c.d.f.s of the posterior expectation corresponding to k and $k' > k$ must cross more than once, and the effect on the researcher is more subtle. Consider again the case of a Laplace distribution. Recall that, in this case, for the evaluator F^2 is worse than F for large R but better for small R (cf. Figure 3). A consequence of this is that the impact on the researcher's welfare changes sign twice. The researcher benefits from selection for small or large values of R, but loses for intermediate values. The left-hand side panel of Figure 6 plots the c.d.f.s of the posterior expectations π^2 (red) and π^1 (blue), illustrating this.

Next, recall the case of the uniform distribution discussed earlier, where the evaluator benefits from selection for large R, but fares worse for small R (cf. Figure 4) Similarly to the case of Laplace distribution, here the impact on the researcher's welfare

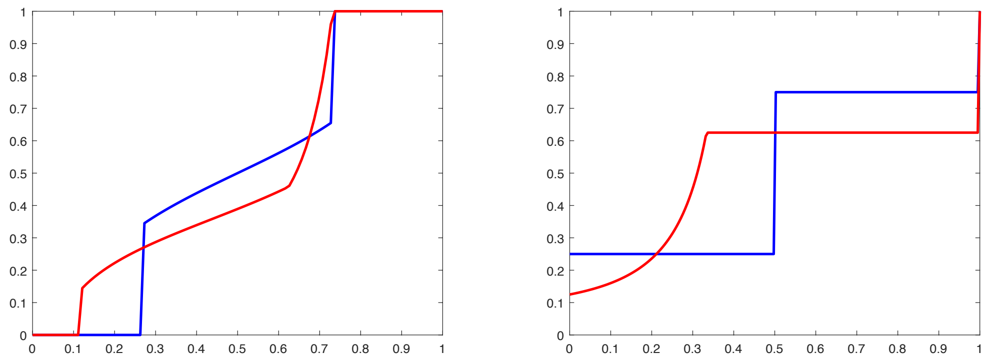

Figure 6: Comparison between F (blue) and F^2 (red) for the uniform distribution.

also changes sign twice as R varies between θ_L and θ_H. However, the impact on the researcher's welfare, just like that on the evaluator's, is exactly the opposite. Selection benefits the researcher for small or large values of R, but harms for intermediate values, as illustrated in the right-hand side panel of Figure 6.

Finally, note that in the Gumbel case, the researcher's selection has no impact on the error rates (α, β). In this case, properly anticipated selection has zero impact on the decision and thus on the information value (evaluator's payoff) and the approval probability (researcher's decision payoff). If there is any cost to selection, this cost is completely wasted, so that the researcher's net payoff is necessarily reduced by strategic selection. In this case, strategic selection is a pure rat race.

To the literature on optimal persuasion following Rayo and Segal (2010) and Kamenica and Gentzkow (2011) we contribute a signal-jamming model of persuasion.[14] The researcher's choice of the size k of the presample is akin to the agent's effort choice in Holmström's (1999) classic career concern model. The twist here is that this effort results in private information, which the researcher then uses to select the reported

[14]See also Henry (2009), Dahm, González, and Porteiro (2009), and Felgenhauer and Schulte (2014) for persuasion models with endogenous information acquisition. In our setting, the researcher is constrained to disclose a single observation, as in the limited-attention models proposed by Fishman and Hagerty (1990) and Hoffmann, Inderst, and Ottaviani (2014).

information. As we show, information manipulation induces positive skewness in the distribution of treated outcomes. Contrary to naive intuition, the evaluator is not necessarily hurt by information manipulation; actually, we characterize natural conditions under which the evaluator benefits. In addition, we characterize situations in which the researcher ends up suffering from information manipulation like in a rat race, even if we abstract away from the cost of acquiring information.

In a pioneering game theoretic analysis, Blackwell and Hodges (1957) analyze how an evaluator should optimally design a sequential experiment to minimize selection bias, defined as the number of times an optimizing researcher is able to correctly forecast the treatment assignment.[15] In the context of our single experiment, we characterize situations in which the selection bias that is present when the researcher is able to forecast the assignment actually benefits the evaluator, contrary to what Blackwell and Hodges (1957) stipulate.

7 Concluding Remarks

While in general selected data are not Blackwell comparable to random data, we characterized the welfare impact of selection on the basis of dispersion. Our results depend on the features of the environment (conditional signal distribution and parameters of decision problem). Our notion of local dispersion, like Lehmann's, applies to the case where the evaluator observes only one outcome x. It is natural to generalize. The construction of the $\beta(\alpha)$ curve is generally applicable, and the convex envelope generally allows for the comparison of experiments, as explained in Section 4.

In a natural extension of the selection model, the number k is random. Again, the

[15] Blackwell and Hodges (1957) argue that selection bias is minimized by a truncated binomial design, according to which the initial allocations to treatment and control are selected independently with a fair coin, until half of the subjects are allocated to either treatment or control; from that point on, allocation is deterministic. Efron (1971), instead, characterizes the selection bias resulting from a biased coin design, according to which the probability of current assignment to treatment is higher if previous randomizations resulted in excess balance of controls over treatments.

$\beta(\alpha)$ curve can be generally used to compare experiments. However, the application of dispersion is harder. Suppose that ε is drawn from $\lambda F^{k+1} + (1-\lambda)F^k$ where $\lambda \in (0,1)$ and F^{k+1} is less dispersed than F^k. It might be natural to conjecture that the evaluator is better off, the greater the weight λ attached to the more effective experiment. However, this is generally false. To see this note that when F is Gumbel, both F^k and F^{k+1} are Gumbel, but $\lambda F^{k+1} + (1-\lambda)F^k$ is not Gumbel. In fact, for every $\lambda \in (0,1)$, $\lambda F^{k+1} + (1-\lambda)F^k$ is worse than F^k, for it is Blackwell worse than informing the evaluator about the outcome of the lottery over F^k and F^{k+1}. Intuitively, the equivalence of F^k with F^{k+1} rests on being able to remove a constant bias from the distribution of ε, but this is not feasible when it is random whether ε derives from one distribution or the other.[16]

Finally, in a more elaborate game-theoretic model it is natural to ask how large is the equilibrium amount of manipulation. Considering a deviation from equilibrium, the researcher actually has a potential gain through the upward shift of the realized observation x. This is to be weighed against the cost of looking at more subjects, when already looking at k. The main point, however, is the ratchet effect: when the evaluator correctly anticipates a greater selection, the researcher's effort to manipulate the experiment is wasted.

We leave to future work the design of experiments and policy responses in the presence of strategic selection. A natural starting point in this direction is Chassang, Padró i Miquel, and Snowberg's (2012) characterization of experimental design when outcomes are affected by experimental subjects' unobserved actions.

[16] Some advance may still be feasible by using the arithmetic-mean to geometric-mean inequality, and observing that $F^{\lambda(k+1)}F^{(1-\lambda)k}$ is itself a power of F, as characterized in Theorem 1.

A Proofs

A.1 Proof of Theorem 3

Fix any $\delta \in (0, (\theta_H - \theta_L)/2)$. Let $\varepsilon_\delta > 0$ be such that $\bar{F}(\varepsilon_\delta) - \bar{F}(-\varepsilon_\delta) \geq 1 - \delta/2$. Choose $\hat{k}$ so that for all $k \geq \hat{k}$,

$$a_k \varepsilon_\delta < \delta, \qquad F^k(b_k + a_k \varepsilon_\delta) \geq \bar{F}(\varepsilon_\delta) - \frac{\delta}{4}, \qquad \text{and} \qquad F^k(b_k - a_k \varepsilon_\delta) \leq \bar{F}(-\varepsilon_\delta) + \frac{\delta}{4}.$$

Then, for each θ, since $x = \theta + b_k + a_k \varepsilon$,

$$\begin{aligned}
\Pr\left(\theta + b_k - \delta \leq x \leq \theta + b_k + \delta \,\middle|\, \theta\right) &\geq \Pr\left(\theta + b_k - a_k \varepsilon_\delta \leq x \leq \theta + b_k + a_k \varepsilon_\delta \,\middle|\, \theta\right) \\
&= F^k(b_k + a_k \varepsilon_\delta) - F^k(b_k - a_k \varepsilon_\delta) \\
&\geq \bar{F}(\varepsilon_\delta) - \frac{\delta}{4} - \bar{F}(-\varepsilon_\delta) - \frac{\delta}{4} \\
&\geq 1 - \delta.
\end{aligned}$$

In words, the distribution of observation x in state θ assigns at least probability $1 - \delta$ to a δ-ball around the point $\theta + b_k$. Now, rejecting if and only if $x < \hat{x} = \theta_H + b_k - \delta$ gives

$$\alpha = \Pr\left(x \geq \theta_H + b_k - \delta \,\middle|\, \theta_L\right) \leq 1 - \Pr\left(\theta_L + b_k - \delta \leq x \leq \theta_L + b_k + \delta \,\middle|\, \theta_L\right) \leq \delta$$

and

$$\beta = \Pr\left(x < \theta_H + b_k - \delta \,\middle|\, \theta_H\right) \leq 1 - \Pr\left(\theta_H + b_k - \delta \leq x \leq \theta_H + b_k + \delta \,\middle|\, \theta_H\right) \leq \delta.$$

As we can choose $\delta > 0$ arbitrarily small, we can make $K\alpha + \beta$ arbitrarily small. The first claim in the theorem now follows from Lemma 1.

To prove the second claim, consider the pair of error rates (α, β) that result with threshold $\hat{x}$ when the error is drawn from $\bar{F}$. Let (α_k, β_k) be the error rates that result with threshold $\hat{x} - b_k$ when the error is draw from F^k. The convergence $F^k(b_k + \varepsilon) \to \bar{F}(\varepsilon)$ at both $\varepsilon = \hat{x} - \theta_H$ and $\hat{x} - \theta_L$ implies that the sequence (α_k, β_k) converges to (α, β). Since this is true for every $\hat{x}$, every point on the $\beta(\alpha)$ curve generated from $\bar{F}$

is a limit point for the corresponding curves for experiments F^k. For the convex curves in the compact space, this implies convergence of the $\beta(\alpha)$ functions. This implies convergence of the evaluator's payoff.

A.2 Extreme Selection in the Exponential Power Case

In this appendix we show that if $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. exponential power with shape b, location 0 and scale 1, then $M_n = \max\{\varepsilon_1, \ldots, \varepsilon_n\}$ satisfies

$$\Pr(M_n \leq a_n \varepsilon + b_n) \to e^{-e^{-\varepsilon}} \quad \forall \varepsilon \qquad \text{(Gumbel)},$$

where

$$a_n = (b \log n)^{-\frac{b-1}{b}}$$

and

$$b_n = (b \log n)^{1/b} - \frac{\frac{b-1}{b} \log\log n + \log(2\Gamma[1/b])}{(b \log n)^{\frac{b-1}{b}}}.$$

Here, Γ denotes the Gamma function.

Remark 1. *For $b = 1$ (Laplace) we have $a_n = 1$ and*

$$b_n = \log n - \log(2\Gamma[1]) = \log n - \log 2.$$

Remark 2. *For $b = 2$ (normal) we have $a_n = (2 \log n)^{-1/2}$ and*

$$\begin{aligned} b_n &= (2 \log n)^{1/2} - \frac{\frac{1}{2} \log\log n + \log(2\Gamma[1/2])}{(2 \log n)^{1/2}} \\ &= (2 \log n)^{1/2} - \frac{\log\log n + 2\log\left(2\sqrt{\pi}\right)}{2(2 \log n)^{1/2}} \\ &= (2 \log n)^{1/2} - \frac{\log\log n + \log 4\pi}{2(2 \log n)^{1/2}}, \end{aligned}$$

as in Leadbetter et al. (Note that a_n in that book corresponds to $1/a_n$ here.)

Remark 3. *The calculations do not cover the case $b = \infty$ (uniform) because in this case the relevant extreme value distribution is not Type I (Gumbel) but rather Type III (Weibull).*

The proof follows closely and generalizes the one given by Leadbetter et al. (1983) for the normal case $b = 2$.

Start by noticing that

$$\frac{f(\varepsilon)}{\varepsilon^{b-1}[1-F(\varepsilon)]} \to 1 \quad \text{as } \varepsilon \to \infty.$$

Fix ε and define y_n for each $n \geq 1$ by

$$1 - F(y_n) = \frac{e^{-\varepsilon}}{n},$$

so that

$$\frac{e^{-\varepsilon}}{n} \frac{y_n^{b-1}}{f(y_n)} \to 1 \quad \text{as } n \to \infty. \tag{2}$$

We may assume $y_n > 0$ for all n. Then

$$f(y_n) = \frac{b^{\frac{b-1}{b}}}{2\Gamma[1/b]} e^{-y_n^b/b},$$

and hence, by (2),

$$-\log n - \varepsilon + (b-1)\log y_n - \frac{b-1}{b}\log b + \log(2\Gamma[1/b]) + \frac{y_n^b}{b} \to 0. \tag{3}$$

From (3) we see that

$$-\log n + (b-1)\log y_n + \frac{y_n^b}{b} = -\log n + o\left(y_n^b/b\right) + \frac{y_n^b}{b} \to \text{a constant},$$

and hence that

$$-\frac{b\log n}{u_n^b} + \frac{o\left(y_n^b/b\right)}{u_n^b/b} + 1 \to 0,$$

that is

$$\frac{b\log n}{u_n^b} \to 1,$$

or

$$b\log y_n - \log b - \log\log n \to 0,$$

that is

$$\log y_n = \frac{1}{b}(\log b + \log\log n) + o(1).$$

Using this fact in (3), we obtain

$$\begin{aligned}\frac{y_n^b}{b} &= \log n + \varepsilon - \frac{b-1}{b}\left(\log b + \log\log n\right) + \frac{b-1}{b}\log b - \log\left(2\Gamma[1/b]\right) + o\left(1\right) \\ &= \log n + \varepsilon - \frac{b-1}{b}\log\log n - \log\left(2\Gamma[1/b]\right) + o\left(1\right),\end{aligned}$$

or

$$y_n^b = b\log n\left[1 + \frac{\varepsilon - \frac{b-1}{b}\log\log n - \log\left(2\Gamma[1/b]\right)}{\log n} + o\left(\frac{1}{\log n}\right)\right],$$

or

$$\begin{aligned}y_n &= \left(b\log n\right)^{1/b}\left[1 + \frac{\varepsilon - \frac{b-1}{b}\log\log n - \log\left(2\Gamma[1/b]\right)}{\log n} + o\left(\frac{1}{\log n}\right)\right]^{1/b} \\ &= \left(b\log n\right)^{1/b}\left[1 + \frac{\varepsilon - \frac{b-1}{b}\log\log n - \log\left(2\Gamma[1/b]\right)}{b\log n} + o\left(\frac{1}{\log n}\right)\right] \\ &= \left(b\log n\right)^{1/b} + \frac{\varepsilon - \frac{b-1}{b}\log\log n - \log\left(2\Gamma[1/b]\right)}{\left(b\log n\right)^{\frac{b-1}{b}}} + o\left(\frac{1}{\left(\log n\right)^{\frac{b-1}{b}}}\right) \\ &= a_n\varepsilon + b_n + o\left(a_n\right).\end{aligned}$$

Thus

$$\Pr\left(M_n \le a_n\varepsilon + b_n + o\left(a_n\right)\right) \to e^{-e^{-\varepsilon}}.$$

References

Allcott, Hunt (2015). "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 130(3), 1117–1165.

Berger, Vance W. (2005). *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. Wiley.

Boland, Philip J., Emad El-Neweihi, and Frank Proschan (1994), "Applications of the Hazard Rate Ordering in Reliability and Order Statistics," *Journal of Applied Probability*, 31(1), 180–192.

Blackwell, David and Joseph L. Hodges (1957), "Design for the Control of Selection Bias," *Annals of Mathematical Statistics*, 28(2), 449–460.

Brown, George W. and John W. Tukey (1946), "Some Distributions of Sample Means," *Annals of Mathematical Statistics*, 17(1), 1–12.

Chassang, Sylvain, Gerard Padró i Miquel, and Erik Snowberg (2012), "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments," *American Economic Review*, 102(4), 1279–1309.

Dahm, Matthias, Paula González, and Nicolás Porteiro (2009), "Trials, Tricks and Transparency: How Disclosure Rules Affect Clinical Knowledge," *Journal of Health Economics*, 28(6), 1141–1153.

Di Tillio, Alfredo, Marco Ottaviani, and Peter N. Sørensen (2015), "Persuasion Bias in Science: Can Economics Help?" mimeo.

Efron, Bradley (1971), "Forcing a Sequential Experiment to be Balanced," *Biometrika*, 58(3), 403–417.

Felgenhauer, Mike and Elisabeth Schulte (2014), "Strategic Private Experimentation," *American Economic Journal: Microeconomics*, 6(4), 74–105.

Fishman, Michael J. and Hagerty, Kathleen M. (1990), “The Optimal Amount of Discretion to Allow in Disclosure,” *Quarterly Journal of Economics*, 105(2), 427–444.

Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions (2015), “The Extent and Consequences of P-hacking in Science,” *PLoS Biology*, 13(3), 1–15.

Heckman, James J. (1979), “Sample Selection Bias as a Specification Error,” *Econometrica*, 47(1), 153–161.

Henry, Emeric (2009), “Strategic Disclosure of Research Results: The Cost of Proving Your Honesty”, *Economic Journal*, 119(539), 1036–1064.

Henry, Emeric and Marco Ottaviani (2015), “Research and the Approval Process: The Organization of Persuasion”, Bocconi mimeo.

Hoffmann, Florian, Roman Inderst, and Marco Ottaviani (2014), “Persuasion through Selective Disclosure: Implications for Marketing, Campaigning, and Privacy Regulation,” Bocconi mimeo.

Holmström, Bengt (1999), “Managerial Incentive Problems: A Dynamic Perspective,” *Review of Economic Studies*, 66(1), 169–182.

Jewitt, Ian (2007), “Information Order in Decision and Agency Problems,” Nuffield College mimeo.

Kamenica, Emir and Matthew Gentzkow (2011), “Bayesian Persuasion,” *American Economic Review*, 101(6), 2590–2615.

Karlin, Samuel and Herman Rubin (1956), “The Theory of Decision Procedures for Distributions with Monotone Likelihood Ratio,” *Annals of Mathematical Statistics*, 27(2), 272–299.

Kochar, Subhash C. (1996), “Dispersive Ordering of Order Statistics,” *Statistics and Probability Letters*, 27(3), 271–274.

Leadbetter, M.R., Georg Lindgren, and Holger Rootzén (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

Lehmann, Erich L. (1988), “Comparing Location Experiments,” *Annals of Statistics*, 16(2), 521–533.

Rayo, Luis and Ilya Segal (2010), “Optimal Information Disclosure,” *Journal of Political Economy*, 118(5), 949–987.

Schulz, Kenneth F. (1995), “Subverting Randomization in Controlled Trials,” *Journal of American Medical Association*, 274(18), 1456–1458.

Schulz, Kenneth F., Iain Chalmers, Richard J. Hayes, and Douglas G. Altman (1995). “Empirical Evidence of Bias: Dimensions of Methodological Quality Associated with Estimates of Treatment Effects in Controlled Trials,” *Journal of the American Medical Association*, 273(5), 408–412.

Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons (2014), “*P*-curve: A Key to the File Drawer,” *Journal of Experimental Psychology: General*, 143(2), 534–547.

Tetenov, Alexey (2016), “An Economic Theory of Statistical Testing,” Bristol mimeo.

Torgersen, Erik (1991), *Comparison of Statistical Experiments*, Cambridge University Press.

Wald, Abraham (1950), *Statistical Decision Function*, Wiley.