SIER Working Paper Series No. 116

Testing for Overconfidence Statistically: A Moment Inequality Approach

By

Yanchun Jin and Ryo Okui

March, 2019

Institute of Economic Research
Seoul National University

Testing for Overconfidence Statistically: A Moment Inequality Approach*

Yanchun Jin[†] and Ryo Okui[‡]

March 2019

Abstract

We propose an econometric procedure to test for the presence of overconfidence using data collected by ranking experiments. Our approach applies the techniques from the moment inequality literature. Although a ranking experiment is a typical way to collect data for the analysis of overconfidence, Benoît and Dubra (2011) show that a ranking experiment may generate data that indicate overconfidence even if participants are purely rational Bayesian updaters. Instead, the authors provide a set of inequalities that are consistent with purely rational Bayesian updaters. We propose the application of the tests of moment inequalities developed by Romano et al. (2014) to test such a set of inequalities. Then, we examine the data from Svenson (1981) on driving safety. Our results indicate the presence of overconfidence with respect to safety among US subjects tested by Svenson. However, other cases tested do not show evidence of overconfidence. We also apply our method to re-examine and confirm the results of Benoît et al. (2015).

Keywords: overconfidence; ranking experiments; moment inequality; driving safety.

JEL Classification: C12; D03; D81; R41.

1 Introduction

In this paper, we propose an econometric procedure to test for the presence of overconfidence. We consider settings for which data are obtained from ranking experiments.

*The authors would like to thank Hiroaki Kaido, Kengo Kato, Michael Wolf, Yoichi Hizen, and seminar participants at Vrije Universiteit Amsterdam, Kyoto University, Kansai Econometrics Research Group Meeting 2017 and NYU Shanghai for their helpful comments. Juan Dubra provided us with useful comments and also helped us better understand the data set used in Benoît et al. (2015). Part of this research was conducted while Okui was at Vrije Universiteit Amsterdam and Kyoto University. Yanchun Jin acknowledges financial support from the Japan Society for the Promotion of Science (JSPS) under KAKENHI Grant Nos. 15J10229 and 18H05695. Okui acknowledges financial support from the Japan Society for the Promotion of Science (JSPS) under KAKENHI Grant Nos. 16K03598 and 15H03329.

†The University of Tokyo, Hongo 7-3-1, Bunkyo, Tokyo, Japan. Email: yanchunjin@e.u-tokyo.ac.jp

‡Department of Economics, Seoul National University, Building 16, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, South Korea; and Department of Economics, University of Gothenburg, P.O. Box 640, SE-405 30 Gothenburg, Sweden. Email: okuiryo@snu.ac.kr

Our procedure is based on tests for moment inequalities. As an example of ranking experiments, we reexamine the data of Svenson (1981) and find that although we confirm the presence of overconfidence concerning driving safety among US subjects tested by Svenson, we cannot reject the null hypothesis of no overconfidence in the other cases tested. We also re-examine the data of Benoît et al. (2015) and confirm the presence of overconfidence.

A large body of studies shows that people tend to overestimate their capabilities and that overconfidence is a common phenomenon.[1] Examples include Mobius et al. (2011) and many others regarding intelligence, Barber and Odean (2001) regarding investment skills, Klar and Giladi (1999) regarding happiness, and Zuckerman and Jost (2001) regarding popularity. Alicke and Govorun (2005) collect many other examples.

However, a recent study by Benoît and Dubra (2011) advocates caution when interpreting these empirical results. The authors argue that a typical research design used in the overconfidence literature—the ranking experiment—may generate data that misleadingly indicates overconfidence even if respondents follow purely Bayesian updating. We call this the apparent overconfidence problem after the title of Benoît and Dubra (2011).

In a ranking experiment, we ask participants to rank themselves according to their beliefs about their skills relative to other members of the groups under question. In the experiment conducted by Svenson (1981) (details of the research design will be presented in Section 4), participants were asked to indicate one of 10 equally sized intervals of a distribution of participants' driving abilities to which they considered they belonged. Two abilities were examined in this research: driving safety and driving skills. Svenson conducted this experiment in the United States and Sweden and obtained the data shown in Table 1. Among the 40 US participants who were tested on driving safety, 22.5% responded that they belonged in the top 10% of participants for driving safety. At first glance, the data imply the existence of substantial overconfidence.

Benoît and Dubra (2011) argue that data obtained from a ranking experiment cannot be directly used to investigate the presence of overconfidence. Nonetheless, they derive a set of inequalities that must be satisfied by responses from purely rational Bayesian updaters. For example, in the data from Svenson (1981), 51.4% of the Swedish subjects considered that they belonged in the top 30% for driving safety. However, this result can be consistent with purely rational Bayesian updating. Benoît and Dubra (2011) show that a fraction under 60% can be consonant with the absence of overconfidence. On the other hand, 46.3% of US subjects considered that they belonged in the top 20% for skill. This result indicates overconfidence, according to Benoît and Dubra (2011); however, the threshold above which overconfidence can be inferred is 40%, not 20%. The question that

[1]For example, Taylor and Brown (1988) argue that "considerable research evidence suggests that overly positive self-evaluations, exaggerated perceptions of control or mastery, and unrealistic optimism are characteristics of normal human thought."

Table 1: The Svenson (1981) data

decile	1	2	3	4	5	6	7	8	9	10	N
Safety											
US	2.5	0.0	5.0	0.0	5.0	2.5	2.5	22.5	37.5	22.5	40
Sweden	0.0	5.7	0.0	14.3	2.9	11.4	14.3	28.6	17.1	5.7	35
Skill											
US	0.0	2.4	2.4	2.4	0.0	12.2	22.0	12.2	26.8	19.5	41
Sweden	2.2	6.7	2.2	4.4	15.5	17.7	11.1	24.4	13.3	2.2	45

Note: Reformatted version of Table 1 in Svenson (1981). The numbers in the cells in columns 1 to 10 are percentages. N is the number of observations. For example, 2.5% of the 40 US participants ranked themselves in the first decile of the distribution for driving safety. Note that there was no overlap among participants between the safety experiment and the skills experiment.

we wish to address in this paper is whether the 6.3% difference observed in the data is a statistically significant difference.

We construct a null hypothesis with a set of moment inequalities according to the theory of Benoît and Dubra (2011) and apply recently developed methods in the moment inequality literature to conduct the test. In particular, we use the tests of Romano et al. (2014). We employ the approach by Romano et al. (2014) for the following two reasons. First, the authors tests are relatively powerful because they involve moment recentering. The null hypothesis of moment inequalities testing does not uniquely specify the values of moments. If we consider the null distribution in which all moment inequalities are binding, the test may not be powerful. The moment recentering approach of Romano et al. (2014) allows us to consider a null distribution in which some moments are negative, which enhances the testing power. Second, the authors approach allows us to compute p-values relatively easily. Other approaches with good power properties (e.g., Andrews and Barwick, 2012) require nontrivial amounts of computation to examine many different sizes other than 0.05, and the computation of p-values is not convenient.

Using state–of–the–art techniques is important in testing a set of moment inequalities. Note that testing a single inequality restriction is an easy one-sided testing problem and is explained in most elementary statistics textbooks. However, testing multiple inequality restrictions is a non-trivial task. It is possible to test for each restriction separately, but a multiple testing problem could arise: The null hypothesis may be rejected more often than the specified size (for example, if we test two hypotheses, rejecting at least one of the two is more likely than rejecting a single hypothesis when only one hypothesis is tested).[2] A simple solution to this problem is to apply the Bonferroni adjustment. For this, when the size is 5%, we use a $5/q$% critical value where q is the number of restrictions. However,

[2]Multiple testing problems have caught attention from experimental economists recently. See, e.g., List et al. (2019).

such an approach tends to make the test unnecessarily conservative. Our approach enables us to conduct a test of multiple inequality restrictions simultaneously with good power properties.

Our reexamination of Svenson (1981) indicates the presence of overconfidence among the US subjects in Svenson's experiment on safety, but overconfidence is not indicated for the skill experiment. Recall that the result for the US subjects on driving skill is only slightly above the theoretical threshold. We find that this deviation from the threshold is not statistically significant; that is, there is no statistical evidence that the US subjects exhibit overconfidence concerning skill. However, we do obtain statistical evidence that the US subjects are overconfident in relation to their driving safety. This result is important because Svenson (1981) is one of the first to provide evidence of the presence of overconfidence based on a relatively clean and credible research design, and it has been cited frequently as evidence of overconfidence in driving safety.

We also apply our method to test the data used in Benoît et al. (2015). The authors extend the theory of Benoît and Dubra (2011) and develop testable restrictions for overconfidence that consider the apparent overconfidence problem discussed in Benoît and Dubra (2011). The two new sets of restrictions provide more powerful tests but require more data, namely, the type of each subject. The authors designed and conducted experiments to meet this data requirement and found the presence of overconfidence using the new tests although the test given in Benoît and Dubra (2011) failed to find the same. Here, we consider the new set of inequality restrictions (the other concerns an equality restriction, and we do not consider it in the present paper). In Benoît et al. (2015), each inequality is examined separately. Our method examines all the inequalities simultaneously and avoids the multiple hypothesis testing problem. Our testing results confirm the authors findings and provide evidence of the presence of overconfidence.

This study contributes to the literature on behavioral economics by providing a statistical procedure to test for overconfidence and by reexamining the data from the seminal paper on this subject (Svenson, 1981). Moreover, we demonstrate that our statistical approach is useful when examining data from more recent studies (Benoît et al., 2015). As we noted above, many studies use ranking experiments to investigate and detect overconfidence. The theoretical implications of Benoît and Dubra (2011) and Benoît et al. (2015) apply in this case, as does our proposed procedure.

As we noted previously, the body of research on overconfidence is growing. The main argument is that overconfidence can explain a wide range of economic problems and that it has many important policy implications. For example, Camerer and Lovallo (1999) find that most subjects who enter an industry consider that the total profit earned by all entrants will be negative, but that their own profit will be positive, which leads to excessive business entry. Gervais and Goldstein (2007) show that in firms, the bias generated by an agent's over-estimation of the marginal product of his or her effort, and thus works harder, makes all agents more productive, including the overconfident agents. Dubra (2004) studies the conditions under which unbiased job searchers whose beliefs about a first offer are correct have higher welfare than overconfident decision makers who believe that the distribution that generates the offers is better than it really is. A symposium of the *Journal of Economic Perspectives, Volume 29, Number 4* collected review papers on overconfidence and its practical and policy implications (Daniel and Hirshleifer, 2015; Grubb, 2015; Malmendier and Taylor, 2015; Malmendier and Tate, 2015). For these implications to be useful, valid statistical evidence of overconfidence is essential. This study considers the fundamental question of how to test for the presence of overconfidence.

This paper also contributes to the literature on moment inequalities. Although the literature on moment inequalities is voluminous, most of the applications relate to empirical industrial organization (e.g., Ciliberto and Tamer, 2009). This paper demonstrates that statistical techniques developed for moment inequalities are useful in behavioral economics. Another paper that uses methods for moment inequalities in behavioral economics is Montiel Olea and Strzalecki (2014), who estimate partially identified distributions of the discount rate and hyperbolic parameters.

The reminder of the paper is organized as follows. Section 2 explains the theory of apparent overconfidence. Section 3 describes methods of moment inequality testing. In Section 4, we reexamine the Svenson (1981) data and find the presence of overconfidence concerning driving safety among US subjects but not in the other cases examined by Svenson. In Section 5, we reexamine the Benoît et al. (2015) data and find evidence of the presence of overconfidence. Section 6 concludes.

2 Theory of apparent overconfidence

In this section, we provide a brief explanation of the theory of apparent overconfidence by Benoît and Dubra (2011). We provide an intuitive outline of the theory and present the formal statement of the theoretical result.

The theory of apparent overconfidence states that, for example, even when the data indicate that the number of people who consider themselves good drivers is greater than the number of people who actually possess good driving skills, the outcome remains consistent with purely rational Bayesian updating.

Central to this theory is that, typically, people's beliefs about their abilities are not degenerate. This is the case because events that cause people to update their beliefs are not frequent enough to achieve a convergence of belief for most people. Rational Bayesian updating requires only that the average of beliefs weighted by their strength be correct. Suppose that people express that their abilities are in the top half when their beliefs of this event are larger than 0.5 (we call it median rationalizability). In this case, Bayesian rationality implies that the fraction of people who think they are in the top half weighted by the strength of belief (in this case, 0.5) should be less than the fraction of people who are actually in the top half (by definition, it is 0.5). This demonstrates that Bayesian rationality cannot be refuted even if everyone in the population thinks that they are in the top half if they express it for the case where their belief of being in the top half is more than 0.5.

For example, suppose that we are interested in overconfidence with respect to driving skill, and car accidents are the events that update peoples beliefs regarding driving skills. A car accident is relatively rare. If an individual has not experienced a car accident, the individual may consider their driving skills to be relatively good, but that person does not possess a strong degree of certainty regarding this belief. On the other hand, if the individual does experience an accident, they may acquire a strong belief that their driving skills are poor. In this example, there are a small number of people who confidently consider their driving skills poor and a large number of people whose beliefs cause them to place a slightly higher likelihood on the possibility that their skills are good.

Now, we introduce the following setting to describe the results of Benoît and Dubra (2011). A rationalizing model is $(\Theta, p, S, \{f_\theta\}_{\theta\in\Theta})$, where $\Theta \subset \mathbb{R}$ is a type space, p is a prior distribution over Θ, S is a set of signals, and f_θ is a distribution over S. In our application, $\theta \in \Theta$ represents driving skill, $s \in S$ indicates whether a participant has experienced a car accident, and f_θ is the probability of experiencing a car accident that depends on driving skill θ. Prior p is equal to the true distribution of skill levels, and it is known to the drivers. However, in the initial stages, none of the drivers have information on their own particular skill levels, and their beliefs about their own skill levels are the same as p. We divide Θ into K-deciles. Let Θ_k be the k-th K-cile; that is, $\Theta_k = \{\theta \in \Theta \mid (k-K)/K \leq p(\theta' < \theta) < k/K\}$ for $k \leq K-1$ and $\Theta_K = \{\theta \in \Theta \mid (K-1)/K \leq p(\theta' < \theta)\}$. Let $p(\cdot \mid s)$ be the posterior over Θ conditional on $s \in S$: for a measurable $A \in \Theta$, $p(A \mid s) = \int_{\theta\in A} f_\theta(s)dp(\theta)/\int_{\theta\in\Theta} f_\theta(s)dp(\theta)$. After the first stage, each participant drives and learns about their skill level from the driving experience. Then, the participants evaluate their skills using Bayes' rule. $p(\cdot \mid s)$ describes the subjective belief about driving skills.

Next, we define the median rationalizability of data. Note that the data from ranking experiments can be represented as $x \in \Delta^K$, where $\Delta^K = \{x = (x_1, \ldots, x_K) \in [0,1]^K \mid \sum_{k=1}^K x_k = 1\}$ and where x_k, $k = 1, \ldots, K$ is the fraction of people who rank themselves in the kth K-cile. Let

$$S_k = \left\{ s \in S \mid p\left(\bigcup_{n=k}^{K} \Theta_n \mid s\right) \geq \frac{1}{2} \text{ and } p\left(\bigcup_{n=1}^{k} \Theta_n \mid s\right) \geq \frac{1}{2} \right\}.$$

S_k is the set of signals that make the median of the posterior belong to Θ_k. Let $F(\cdot)$ be the marginal distribution of S so that $F(S_k)$ is the population fraction of participants for whom the medians of the posteriors are in Θ_k. We say that $x \in \Delta^K$ is median rationalized for (Θ, p) if there exists a rationalizing model $(\Theta, p, S, \{f_\theta\}_{\theta\in\Theta})$ such that $x_k = F(S_k)$ for $k = 1, \ldots, K$, this means that every driver places themselves in a certain K-cile if they believe that their actual skill places them in that K-cile or above with a probability of at least 1/2 and that it also places them in that K-cile or below with a probability of at least 1/2.

The following theorem states that a wide range of data from ranking experiments can be median rationalized even if the data show apparent overconfidence (or underconfidence).

Theorem 1 (Theorem 1 of Benoît and Dubra (2011))**.** *Suppose that $\Theta \subseteq \mathbb{R}$ and p is a distribution over Θ such that $p(\Theta_k) = 1/K$ for all k. Then, the population ranking data $x \in \Delta^K$ can be median rationalized for (Θ, p) if and only if for $i = k, \ldots, K$:*

$$\sum_{j=k}^{K} x_j < \frac{2}{k}(K - k + 1)$$

and

$$\sum_{j=1}^{k} x_j < \frac{2}{K}k.$$

In our application, we have $K = 10$, and x_k is the fraction of subjects who place themselves in the k-th decile. After eliminating the redundant items, we have the following eight inequalities:

$$\begin{aligned}
&x_1 < 0.2, \quad x_1 + x_2 < 0.4,\\
&x_1 + x_2 + x_3 < 0.6, \quad x_1 + x_2 + x_3 + x_4 < 0.8,\\
&x_7 + x_8 + x_9 + x_{10} < 0.8, \quad x_8 + x_9 + x_{10} < 0.6,\\
&x_9 + x_{10} < 0.4, \quad x_{10} < 0.2.
\end{aligned}$$

3 Test for moment inequalities

In this section, we describe the tests for moment inequalities developed by Romano et al. (2014). Here, we present the procedure only. For the theoretical arguments underlying this testing procedure, refer to Romano et al. (2014).

We describe the tests in a general setting. Let $W_i = (W_{i1}, \ldots, W_{iJ})', i = 1, \ldots, N$ be an i.i.d. sequence of random vectors with mean $\mu \in \mathbb{R}^k$. For $1 \leq j \leq J$, let μ_j be the j-th component of μ. Romano et al. (2014) consider the following testing problem:

$$H_0 : \mu_j \leq 0, \text{ for all } j = 1, \ldots, J, \tag{1}$$

against the alternative:

$$H_1 : \mu_j > 0, \text{ for some } j = 1, \ldots, J.$$

(1) is the hypothesis that all J moment inequalities are satisfied. When $J = 1$—that is, when there is only one moment inequality to test—a standard one-sided t test may be used. For multidimensional inequality hypotheses, the testing problem becomes more complicated.

We define some notations. Let $\bar{W} = \sum_{i=1}^{N} W_i/N$ be the sample average of W_i, and let $\hat{\Sigma} = \sum_{i=1}^{N}(W_i - \bar{W})(W_i - \bar{W})'/N$ be the sample variance covariance matrix of W_i. However, in what follows, we use a regularized version $\tilde{\Sigma}$ such that $\tilde{\Sigma}$ is always invertible:

$$\tilde{\Sigma} = \hat{\Sigma} + \max\{\epsilon - \det(\hat{\Omega}), 0\}\hat{D},$$

where $\hat{D} = Diag(\hat{\Sigma})$, $\hat{\Omega} = \hat{D}^{-1/2}\hat{\Sigma}\hat{D}^{-1/2}$, and $\epsilon = 0.012$. The presence of ϵ guarantees that $\tilde{\Sigma}$ is invertible.[3] For $j = 1, \ldots, J$, let $\bar{W}_j$ and $\hat{\sigma}_j$ denote the sample mean and sample standard deviation of $W_{1j}, \ldots, W_{Nj}$, respectively; that is, $\bar{W}_j = \sum_{i=1}^{N} W_{ij}/N$, $\hat{\sigma}_j = \left(\sum_{i=1}^{N}(W_{ij} - \bar{W}_j)^2/N\right)^{1/2}$.

We consider the following three test statistics.

$$T^{\mathrm{MAX}} = \max_{1\le j\le J} \frac{\sqrt{N}\bar{W}_j}{\hat{\sigma}_j} \tag{2}$$

$$T^{\mathrm{QLR}} = \inf_{t\in\mathbb{R}^J : t\le 0} (\sqrt{N}\bar{W} - t)'\tilde{\Sigma}^{-1}(\sqrt{N}\bar{W} - t) \tag{3}$$

$$T^{\mathrm{MMM}} = \sum_{j=1}^{J}\left(\frac{\sqrt{N}\bar{W}_j}{\hat{\sigma}_j}\right)^2 \mathbf{1}\{\bar{W}_j > 0\} \tag{4}$$

The MAX statistic, (2), is given by the maximum over J t- statistics. The QLR statistic in (3) is of quadratic form and measures the distance to the region that satisfies the inequalities. The MMM statistic in (4) may be considered a special case of QLR, but it ignores the correlation across the elements of $\bar{W}$.

Next, we discuss how to obtain critical values. The critical values are computed by bootstrap, but this involves moment recentering. That is, we can improve the power of a test by adjusting the bootstrap distribution of moments for which the associated inequalities are obviously satisfied.[4] Let α be the nominal size of the test. In principal, we can compute critical values based on a distribution that satisfies $E(W_i) = 0$ (this is called the least favorable distribution approach). However, when the number of moment

[3]The choice of the value of ϵ is somewhat arbitrary. $\epsilon = 0.012$ is used in Andrews and Barwick (2012) and Romano et al. (2014). We confirm that our results are robust to smaller values of ϵ.

[4]The literature also considers an alternative approachmoment selection, which drops moment inequalities that are obviously satisfied. See Allen (2018) for the difference between moment recentering and moment selection. The author argues that moment recentering provides a more powerful test. We adopt the term moment recentering from Allen (2018).

inequalities increases, so does the critical value, which causes the test to lose its power. To improve the power of the test, we consider the test incorporating moment recentering procedures. We allow some of the moments to be negative when we compute the distribution of the test statistics under the null hypothesis. This has a noticeable effect on the power of the test. Romano et al. (2014) provides a two-step method for achieving this goal. Note that this constitutes the main contribution of Romano et al. (2014).

First, we construct a confidence interval for each $E(W_{ij})$ at a confidence level $(1-\beta)$ where $\beta = \alpha/10$ in our application. If the upper bound of the confidence set is below zero, the moment inequality may be considered satisfied. The confidence set is obtained by bootstrap. Let $W_i^{(b)}$, $i = 1, \ldots, N$ be a sample from the bootstrap distribution in the b-th nonparametric bootstrap repetition, where $b = 1, \ldots, B$, and B is the number of bootstrap repetitions. We compute $\bar{W}_j^{(b)}$ and $\hat{\sigma}_j^{(b)}$, $j = 1, \ldots, J$, for each bootstrap repetition. Then, we obtain the empirical distribution of $\max_{1\leq j\leq J}(\sqrt{N}(\bar{W}_j - \bar{W}_j^{(b)})/\hat{\sigma}_j^{(b)})$. Let $\hat{L}^{-1}(1-\beta)$ be the $(1-\beta)$ quantile of this empirical distribution. The confidence set is:

$$\begin{aligned}\hat{M}(1-\beta) &= \left\{\mu \in \mathbb{R}^J : \max_{1\leq j\leq J} \frac{\sqrt{N}(\mu_j - \bar{W}_j)}{\hat{\sigma}_j} \leq \hat{L}^{-1}(1-\beta)\right\}, \\ &= \left\{\mu \in \mathbb{R}^J : \mu_j \leq \bar{W}_j + \frac{\hat{\sigma}_j \hat{L}^{-1}(1-\beta)}{\sqrt{N}}, \text{ for all } 1 \leq j \leq k\right\}.\end{aligned}$$

Then, we form the upper confidence bound for each μ_j by $\bar{W}_j + \hat{\sigma}_j \hat{L}^{-1}(1-\beta)/\sqrt{N}$.

We then compute the bootstrap distributions of the test statistics. Roughly speaking, we adjust the means of moment inequalities for which the upper confidence bounds are below 0. Let:

$$\lambda_j^* = \min\left\{\bar{W}_j + \frac{\hat{\sigma}_j \hat{L}^{-1}(1-\beta)}{\sqrt{N}}, 0\right\}.$$

and $\lambda^* = (\lambda_1^*, \ldots, \lambda_J^*)'$. We substitute the mean value μ_j under the bootstrap distribution, which is $\bar{W}_j$ with $\bar{W}_j - \lambda_j^*$. For example, for T^{MAX}, we compute the empirical distribution of:

$$T^{\text{MAX},(b)} = \max_{1\leq j\leq J} \frac{\sqrt{N}(\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^*)}{\hat{\sigma}_j^{(b)}}.$$

Similarly, for T^{QLR} and T^{MMM}, we compute the empirical distributions of:

$$\begin{aligned}T^{\text{QLR},(b)} &= \inf_{t\in\mathbb{R}^J : t\leq 0} (\sqrt{N}(\bar{W}^{(b)} - \bar{W} + \lambda^*) - t)'(\tilde{\Sigma}^{(b)})^{-1}(\sqrt{N}(\bar{W}^{(b)} - \bar{W} + \lambda^*) - t), \\ T^{\text{MMM},(b)} &= \sum_{j=1}^{J} \left(\frac{\sqrt{N}(\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^*)}{\hat{\sigma}_j^{(b)}}\right)^2 \cdot \mathbf{1}\{\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^* > 0\},\end{aligned}$$

where $\tilde{\Sigma}^{(b)}$ is the bootstrap version of $\tilde{\Sigma}$. Note that when $\lambda_j^* < 0$, $\bar{W}_j^{(b)} - \bar{W}_j + \lambda^* = \bar{W}_j^{(b)} + \hat{\sigma}_j \hat{L}^{-1}(1-\beta)/\sqrt{N}$. Because $\hat{L}^{-1}(1-\beta)$ is the $(1-\beta)$ quantile of the bootstrap distribution of $\max_{1 \le j \le J}(\sqrt{N}(\mu_j - \bar{W}_j)/\hat{\sigma}_j)$, it is unlikely that the moment with $\lambda_j^* < 0$ affects the bootstrap distribution of a test statistic. In this way, we do not use the least favorable null distribution, and we allow some moments to be negative.

The critical values are computed using the bootstrap distributions, but we also need to consider the effects of moment recentering. In particular, we use the $(1-\alpha+\beta)$ quantile of the bootstrap distribution rather than the $(1-\alpha)$ quantile. For example, the critical value for T^{MAX} is the $(1-\alpha+\beta)$ quantile of the empirical distribution of $T^{\text{MAX},(b)}$. Roughly speaking, because we use the $(1-\beta)$ confidence bound for moment selection, the probability of making a mistake in moment recentering is β, and we need to use a slightly higher critical value than the $(1-\alpha)$ quantile.

A test rejects the null hypothesis if the upper confidence bound is above 0 and if the statistic exceeds the critical value. Let T be one of T^{MAX}, T^{QLR}, and T^{MMM}, and let $\hat{c}(1-\alpha+\beta)$ be the corresponding critical value. The null hypothesis H_0 is rejected at size α when $\{\hat{M}(1-\beta) \nsubseteq \mathbb{R}_-^J\}$ and when $T > \hat{c}(1-\alpha+\beta)$.

4 Reexamination of Svenson's (1981) data

This section presents the results of our reexamination of Svenson's (1981) data. First, we review the data. Then, we present the results of our tests. We find that the US subjects are overconfident regarding their driving safety, but we find no statistical evidence for the presence of overconfidence in the other cases tested by Svenson.

4.1 Data

We reexamine the data provided in Svenson (1981), which Svenson collected in the United States and Sweden. In each country, the author gathered two collections of data. For each collection, he gathered participants in a room and asked them to respond to the following questionnaire:

> Svenson's (1981, page 144) questionnaire.
>
> We would like to know about what you think about how safely you drive an automobile. All drivers are not equally safe drivers. We want you to compare your own skill to the skills of the other people in this experiment. By definition, there is a least safe and a most safe driver in this room. We want you to indicate your own estimated position in this experimental group (and not, e.g., in Eugene, Oregon, or in the US) (or (and not, e.g., people in Stockholm or in Sweden)). Of course, this is a difficult question because you

do not know all of the people gathered here today, much less how safely they drive. But please make the most accurate estimate you can.

Each participant was asked to mark one of the deciles that they considered corresponded with their position in the distribution of participants' safety (or skill) levels. Note that there were different participants involved in the two data collections related to driving skills and safety in each country. Thus, Svenson obtained four sets of data, which we label as follows: US safety data, US skill data, Sweden safety data, and Sweden skill data. The data are documented in Table 1.

In view of the theory of apparent overconfidence by Benoît and Dubra (2011), our null hypothesis is composed of the following moment inequalities. Let $D(j)_i$ denote the dummy variable, which takes a value of one if subject i places themselves in the j-th decile. The set of moment inequalities is:[5]

$$E(D(1)_i - 0.2) < 0, \tag{5}$$

$$E(D(1)_i + D(2)_i - 0.4) < 0, \tag{6}$$

$$E(D(1)_i + D(2)_i + D(3)_i - 0.6) < 0, \tag{7}$$

$$E(D(1)_i + D(2)_i + D(3)_i + D(4)_i - 0.8) < 0, \tag{8}$$

$$E(D(7)_i + D(8)_i + D(9)_i + D(10)_i - 0.8) < 0, \tag{9}$$

$$E(D(8)_i + D(9)_i + D(10)_i - 0.6) < 0, \tag{10}$$

$$E(D(9)_i + D(10)_i - 0.4) < 0, \tag{11}$$

$$E(D(10)_i - 0.2) < 0. \tag{12}$$

The number of bootstrap replications is 5,000.[6] The procedure is implemented by R 3.3.0 (R Core Team, 2016) with Mac OS 10.10.5.

4.2 Results

Table 2 summarizes the results of the overconfidence test for US and Swedish drivers' driving safety and skill levels. The table lists three test statistics and their corresponding critical values and p-values. All three different test statistics lead to the same test results for each ability for each country at the 5% significance level. We reject the null hypothesis that US drivers have no overconfidence regarding their driving safety. However, in the other three cases, we do not find statistical evidence to disprove the hypothesis of no overconfidence. For example, from Svenson's (1981) data, we can see that 46% of US

[5] In Svenson's sample, no US drivers placed their driving skills in the lowest decile, and no Swedish drivers placed their safety in the lowest decile; that is, we do not need to test for the corresponding inequality in these two cases.

[6] As $D(j)_i$ takes only a value of zero or one when computing the standard deviation, $\hat{\sigma}_j^{(b)}$ may turn out to be 0 in some bootstrap repetitions. In those cases, we take $\tilde{S}_j^{(b)} = max(\epsilon_b, \hat{\sigma}_j^{(b)})$ where $\epsilon_b = 0.001$. Different values of ϵ_b do not affect the results as long as they are sufficiently small.

Table 2: Test results for null hypothesis of no overconfidence

Version	Reject	Statistics	Crit.val.	p-value	Not-recentered
US safety ($N = 40$)					
MMM	Yes	21.08	14.32	0.03	(9)-(12)
QLR	Yes	13.85	8.22	0.019	(9)-(12)
MAX	Yes	3.698	2.865	0.015	(9)-(12)
US skill ($N = 41$)					
MMM	No	0.653	10.53	0.58	(9)-(12)
QLR	No	0.647	6.615	0.55	(9)-(12)
MAX	No	0.804	2.514	0.54	(9)-(12)
Sweden safety ($N = 35$)					
MMM	No	0	8.075	0.86	(9)-(11)
QLR	No	0	4.669	1.00	(9)-(11)
MAX	No	-1	2.082	1.00	(9)-(11)
Sweden skill ($N = 45$)					
MMM	No	0	3.267	0.82	(10)
QLR	No	0	3.106	0.96	(10)
MAX	No	-2.708	1.773	1.00	(10)

Notes: Results of the Romano et al. (2014) tests. MMM, QLR, and MAX correspond to the three versions of the Romano et al. (2014) tests. Reject gives the test results at the 5% significance level. Statistics gives the value of the test statistics. Crit.val denotes the critical value computed by bootstrap. P-value gives the p-value of the test. Not-recentered gives the list of moment inequalities that are not affected by moment recentering.

drivers placed themselves in the top 20% of the distribution for skill, which is 6% larger than the 40% threshold. However, our tests show that there is no statistical evidence of the presence of overconfidence. To illustrate the effect of moment recentering, Table 2 documents the moments for which we need to set population values equal to zero when we generate the distribution under the null hypothesis. These moment inequalities are not affected by moment recentering. We can see that in most tests, the latter half of the inequalities are selected. In fact, the data indicate that the recorded fractions of the top deciles are less likely to satisfy the inequalities.

Table 3 illustrates the effects of moment recentering. We provide the results of testing for US driving safety and skill levels with (i) eight (or seven) moment inequalities with recentering (the same as the Table 2), (ii) eight (or seven) moment inequalities without recentering, and (iii) four moment inequalities that are selected in Table 2 ((9)-(12)). The critical values in panel (ii) are larger than those in panel (iii) because of the increased number of moment inequalities. Comparing tests (i) and (ii), we find that the tests with test statistics QLR and MAX have smaller critical values when moment recentering is implemented. In particular, for US safety, the QLR test does not reject the null hypothesis at the 5% level without moment recentering.

Table 3: Effects of moment recentering

Version	Reject	Statistics	Crit.val.	p-value	Not-recentered
US safety ($N = 40$)					
(i): Eight moment inequalities with recentering					
MMM	Yes	21.08	14.32	0.03	(9)-(12)
QLR	Yes	13.85	8.22	0.019	(9)-(12)
MAX	Yes	3.698	2.865	0.015	(9)-(12)
(ii): Eight moment inequalities without recentering					
MMM	Yes	21.08	13.83	0.027	
QLR	No	13.85	24.25	0.064	
MAX	Yes	3.698	2.865	0.014	
(iii): Four moment inequalities (9)-(12) without recentering					
MMM	Yes	21.08	13.47	0.027	
QLR	Yes	13.85	8.211	0.017	
MAX	Yes	3.698	2.865	0.014	
US skill ($N = 41$)					
(i): Seven moment inequalities with recentering					
MMM	No	0.653	10.53	0.58	(9)-(12)
QLR	No	0.647	6.615	0.55	(9)-(12)
MAX	No	0.804	2.514	0.54	(9)-(12)
(ii): Seven moment inequalities without recentering					
MMM	No	0.653	10.42	0.75	
QLR	No	0.647	10.44	0.73	
MAX	No	0.804	2.242	0.67	
(iii): Four moment inequalities (9)-(12) without recentering					
MMM	No	0.653	9.863	0.52	
QLR	No	0.647	6.323	0.52	
MAX	No	0.804	2.242	0.52	

Note: Results of the Romano et al. (2014) tests with or without moment recentering. MMM, QLR, and MAX correspond to the three versions of the Romano et al. (2014) tests. Reject gives the test results at the 5% significance level. Statistics gives the value of the test statistics. Crit.val denotes the critical value computed by bootstrap. P-value gives the p-value of the test. Not-recentered gives the list of moment inequalities that are not affected by moment recentering.

5 Re-examination of Benoît et al.'s (2015) data

Benoît et al. (2015) present the results of experiments that are also ranking experiments but provide better data in the sense that the information on the strength of beliefs and actual types is available. In the theory part of Benoît et al. (2015), the authors derive an additional set of inequalities that can be used to test data for the presence of overconfidence. In this section, we demonstrate that our moment inequality approach is useful in re-examining the results from Benoît et al. (2015). Our tests provide evidence of the presence of overconfidence.[7]

5.1 Theoretical framework

We examine the implication of Theorem 2 in Benoît et al. (2015) and provide a set of testable inequalities describing the relationship between subjects' actual ability and their beliefs. Note that Benoît et al. (2015) present three theoretical results that extend Benoît and Dubra (2011) and provide testable implications of no overconfidence. Here, we consider their second result (Theorem 2) only. Their first result is a direct generalization of Theorem 1 in Benoît and Dubra (2011) (it is a slight modification of Theorem 4 in the appendix of Benoît and Dubra (2011)). Note that these restrictions indicated by this first result are satisfied in the data used in Benoît et al. (2015), and our moment inequality approach yields the same results. The authors third result yields an equality restriction and is beyond the scope of the present paper although it would provide the most powerful test among those discussed in Benoît et al. (2015). Note that Benoît et al. (2015) find the evidence for the presence of overconfidence bias from the test based on this equality restriction.

The following theorem, which is Theorem 2 of Benoît et al. (2015), provides a set of inequality restrictions that need to be satisfied in the absence of overconfidence.

Theorem 2 (Theorem 2 in Benoît et al. (2015)). *Suppose that a fraction x of the population believe that there is a probability of at least q that their types are in the top $y < q$ of the population. Let $\tilde{x}$ be the fraction of people who have those beliefs and whose actual type is in the top y of the population. These data can be rationalized if and only if $xq \leq \tilde{x}$.*

This theorem states that, in a large population, at least a fraction q of those who think that they are in the top y with probability at least q must actually be in the top y. The inequality given in this theorem is more restrictive than that in Theorem 1 in Benoît et al. (2015) because $\tilde{x} \leq y$ and, thus, yields a more powerful test. However, the test requires data on actual types. For example, this theorem cannot be used to test the data set in Svenson (1981) because it does not provide the data on actual types. Benoît et al. (2015) designed and conducted experiments that generate data on actual types.

Theorem 2 corresponds to the following moment inequality.

$$E[qD(y,q) - \tilde{D}(y,q)] \leq 0, \tag{13}$$

where $D(y,q)$ is a binary variable indicating whether a subject believes that their type is in the top $y < q$ of the population with a probability of at least q, and $\tilde{D}(y,q)$ denotes whether or not a subject believes that their type is in the top $y < q$ of the population with a probability of at least q.

[7] We use the data set available at http://learnmoore.org/mooredata/OJD/.

Benoît et al. (2015) report the results of two experiments, Experiment I and Experiment II. Below, we discuss these two experiments and their results separately. We briefly discuss the designs; see Benoît et al. (2015) for details.

5.2 Experiment I

In Experiment I, subjects took a quiz composed of 20 questions. The subjects types were defined as the number of questions they answered correctly in the quiz and ranged from 0 to 20. Each subject then made three decisions. Each decision corresponded to the elicitation of their belief that their types were in the top y with a probability of at least q. The three decisions correspond to $(y, q) = (0.5, 0.5), (0.6, 0.5)$, and $(0.5, 0.3)$. Note that the experimental design enables us to specify the value of q, and we do not need to assume median rationalizability, which is made to examine Svenson's data. The number of observations is 129. We observe $\{(D_i(y, q), \tilde{D}_i(y, q) \mid (y, q) \in \{(0.5, 0.5), (0.6, 0.5), (0.5, 0.3)\}\}_{i=1}^{129}$ where $D_i(y, q)$ is a binary variable indicating whether subject i believes that his type is in the top y with a probability of at least q, and $\tilde{D}_i(y, q)$ denotes whether the subject believes the same and their type is actually in the top y.[8] Table 6 in the Appendix, which provides a subset of the information in Table 3 of Benoît et al. (2015), reports a summary of the data.

Benoît et al. (2015) examine each decision separately. We look at all three decisions altogether. This is important because these three decisions are made by the same set of subjects, and they are likely to be correlated. In Benoît et al. (2015), the authors find that the test on the third decision rejects the null hypothesis. We formally test whether it comes from sampling error while taking care of the correlation across inequalities and considering all three simultaneously to avoid the multiple testing problem.

Table 4 summarizes the test results. The null hypothesis is rejected regardless of the choice of test statistic. The p-values are low, and we find strong evidence of the presence of overconfidence.

5.3 Experiment II

Experiment II is similar to Experiment I, but the elicitation of subjects' beliefs uses a different method. There is only one decision with $y = 0.5$, and the subjects themselves specify the probability that their types are in the top half. The quiz is similar to the quiz

[8] Note that the score is a discrete variable, and there are multiple ways of defining percentiles (that is, what the top y means). Here we define $\tilde{D}_i(y, q)$ in the following way; 21, 33, and 18 subjects scored 20, 19, and 18, respectively. For $y = 0.5$, if the score is 19 or 20, $\tilde{D}_i(0.5, q) = D_i(0.5, q)$. If the score is 18, then $\tilde{D}_i(0.5, q) = D_i(0.5, q) \times 5/9$ where $5/9 = 10/18$ comes from the condition that the top 50% should consist of 64 subjects (the total number of subjects is 129) and $64 - 21 - 33 = 10$. Similarly, for $y = 0.3$, $\tilde{D}_i(0.3, q) = D_i(0.3, q)$ for those whose score is 20 and $\tilde{D}_i(0.3, q) = D_i(0.3, q) \times 17/33$. This definition makes the average values of $\tilde{D}_i(y, q)$ equal to those reported in Benoît et al. (2015).

Table 4: Tests for Theorem 2 in Benoît et al. (2015) with data from Experiment I

Version	Reject	Statistics	Crit.val.	p-value	λ^*
$\alpha = 0.05, \beta = \alpha/10, N = 129$					
MMM	Yes	13.414	7.053	0.008	0, 0, 0
QLR	Yes	11.965	4.344	0.001	0, 0, 0
MAX	Yes	3.459	2.035	0.000	0, 0, 0

Notes: Results of the Romano et al. (2014) tests. MMM, QLR, and MAX correspond to the three versions of the Romano et al. (2014) tests. Reject gives the test results at the 5% significance level. Statistics gives the value of the test statistics. Crit.val denotes the critical value computed by bootstrap. P-value gives the p-value of the test. λ^* denotes the values of the re-centering parameters for each of the three inequalities.

Table 5: Tests for Theorem 2 in Benoît et al. (2015) with data from Experiment II

Version	Reject	Statistics	Crit.val.	p-value	Not-recentered
$\alpha = 0.05, \beta = \alpha/10, N = 74$					
MMM	Yes	43.421	21.302	0.018	all
QLR	Yes	6.491	3.640	0.017	all
MAX	Yes	2.360	1.771	0.018	all

Notes: Results of the Romano et al. (2014) tests. MMM, QLR, and MAX correspond to the three versions of the Romano et al. (2014) tests. Reject gives the test results at the 5% significance level. Statistics gives the value of the test statistics. Crit.val denotes the critical value computed by bootstrap. P-value gives the p-value of the test. Not-recentered gives the list of moment inequalities that are not affected by moment recentering.

in Experiment I and has 20 questions. The types are the scores of the quiz. Seventy-four subjects participated.

We test Theorem 2 for the following 19 values of q: q = 0.58, 0.60, 0.66, 0.68, 0.70, 0.72, 0.74, 0.76, 0.78, 0.80, 0.84, 0.86, 0.88, 0.90, 0.92, 0.94, 0.96, 0.98, 1.00. This set of values is also used in Benoît et al. (2015). The original data set contains $(Score_i, q_i)$ for each i where $Score_i$ is subject i's score (type), and q_i is the probability at which subject i believes that their type is in the top half. To test (13), we construct $D_i(0.5, q)$ and $\tilde{D}_i(0.5, q)$ by setting $D_i(0.5, q) = \mathbf{1}(q_i \geq q)$ and $\tilde{D}_i(0.5, q) = \mathbf{1}(q_i \geq q) \times P(i \text{ is in the top half})$.[9] Table 7 in the appendix, which provides a subset of the information in Table C.1 of Benoît et al. (2015), reports a summary of the data.

Table 5 summarizes the results. All tests reject the null hypothesis, and we find evidence of the presence of overconfidence. Note that Benoît et al. (2015) examine each

[9] As in Experiment I, the score is a discrete variable, and there is ambiguity as to how to define the top 50%. Note that the median score is 18. In the data, there are 34 subjects whose score is less than or equal to 17 and, for them, $P(i \text{ is in the top half}) = 0$. For 27 subjects whose scores are more than or equal to 19, $P(i \text{ is in the top half}) = 1$. For 13 subjects whose score is 18, $P(i \text{ is in the top half}) = 3/13$. This definition yields the average values of $\tilde{D}_i(0.5, q)$ corresponding to those reported in Benoît et al. (2015).

of 19 restrictions separately. Here, we consider all 19 inequalities simultaneously and still find the rejection of the null hypothesis. Our result confirms the finding of Benoît et al. (2015) while taking the multiplicity of restrictions into consideration. We also find that none of the moment inequalities is re-centered. Thus, in this data, no restriction is obviously satisfied.

6 Conclusion

In this paper, we show that tests for moment inequalities can be used to examine the presence of overconfidence using data from ranking experiments. The absence of overconfidence does not yield a null hypothesis with a set of equalities. Rather, it yields a set of inequalities if the data are collected by ranking experiments (Benoît and Dubra, 2011). We use state-of-the-art econometric techniques for moment inequalities to test the set of inequalities implied by the absence of overconfidence. Our reexamination of the Svenson (1981) data reveals that although we can confirm the presence of overconfidence among US subjects with respect to driving safety, we cannot reject the null hypothesis of no overconfidence in the other three cases examined. We also reexamine the data of Benoît et al. (2015) and confirm their findings.

The proposed method is useful in two ways. First, it provides a statistical tool to detect the presence of overconfidence. Overconfidence has several policy implications. However, valid statistical evidence of overconfidence is essential for these policy implications to be useful. Second, the proposed statistical procedure would be useful in addressing other problems that may use ranking experiments. This study analyzes data on driving abilites and intelligence; however, overconfidence has important implications in other aspects of human life, such as happiness and health. Researchers can investigate these using ranking experiments (see, for example, Klar and Giladi, 1999, on happiness) and the proposed method for the analysis.

References

Alicke, M. D. and O. Govorun (2005). The better-than-average effect. In M. D. Alicke, D. A. Dunning, and J. I. Krueger (Eds.), *The Self in Social Judgment*, Studies in Self and Identity, Chapter 5, pp. 85–106. Psychology Press.

Allen, R. (2018). Testing moment inequalities: Selection versus recentering. *Economics Letters 162*, 124–126.

Andrews, D. W. and P. J. Barwick (2012). Inference for parameters defined by moment inequalities: A recommended moment selection procedure. *Econometrica 80*(6), 2805–2826.

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics 116*(1), 261–292.

Benoît, J.-P. and J. Dubra (2011). Apparent overconfidence. *Econometrica 79*(5), 1591–1625.

Benoît, J.-P., J. Dubra, and D. A. Moore (2015). Does the better-than-average effect show that people are overconfident?: Two experiments. *Journal of the European Economic Association 13*(2), 293–329.

Camerer, C. and D. Lovallo (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review 89*(1), 306–318.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica 77*(6), 1791–1828.

Daniel, K. and D. Hirshleifer (2015). Overconfident investors, predictable returns, and excessive trading. *Journal of Economic Perspectives 29*(4), 61–88.

Dubra, J. (2004). Optimism and overconfidence in search. *Review of Economic Dynamics 7*(1), 198–218.

Gervais, S. and I. Goldstein (2007). The positive effects of biased self-perceptions in firms. *Review of Finance 11*(3), 453–496.

Grubb, M. D. (2015). Overconfident consumers in the marketplace. *Journal of Economic Perspectives 29*(4), 9–36.

Klar, Y. and E. E. Giladi (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin 25*(5), 586–595.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*. forthcoming.

Malmendier, U. and G. Tate (2015). Behavioral CEOs: The role of managerial overconfidence. *Journal of Economic Perspectives 29*(4), 37–60.

Malmendier, U. and T. Taylor (2015). On the verges of overconfidence. *Journal of Economic Perspectives 29*(4), 3–8.

Mobius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2011). Managing self-confidence: Theory and experimental evidence. National Bureau of Economic Research Working Paper.

Montiel Olea, J. L. and T. Strzalecki (2014). Axiomatization and measurement of quasi-hyperbolic discounting. *The Quarterly Journal of Economics 129*(3), 1449–1499.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Romano, J. P., A. M. Shaikh, and M. Wolf (2014). A practical two-step method for testing moment inequalities. *Econometrica 82*(5), 1979–2002.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica 47*(2), 143–148.

Taylor, S. E. and J. D. Brown (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin 103*(2), 193.

Zuckerman, E. W. and J. T. Jost (2001). What makes you think you're so popular? self-evaluation maintenance and the subjective side of the "friendship paradox". *Social Psychology Quarterly 64*(3), 207–223.

A Appendix: Additional Tables

Table 6: Summary statistics for Experiment I

y	q	x	y/q	xq	$\tilde{x}$
0.5	0.5	0.74	1.00	0.370	0.394
0.6	0.5	0.64	0.83	0.384	0.347
0.5	0.3	0.52	0.60	0.260	0.169

Notes: This table is a reproduction of part of Table 3 in Benoît et al. (2015). The variables are defined as follows: A fraction x of subjects believe that they are in the top y with a probability of at least q, and a fraction $\tilde{x}$ of subjects believe the same and are, in fact, in the top y.

Table 7: Summary statistics for Experiment II

y	q	x	xq	$\tilde{x}$
0.5	0.5	0.905	0.453	0.473
0.5	0.58	0.730	0.423	0.439
0.5	0.60	0.716	0.430	0.425
0.5	0.66	0.608	0.401	0.358
0.5	0.68	0.595	0.401	0.344
0.5	0.70	0.554	0.388	0.320
0.5	0.72	0.432	0.311	0.238
0.5	0.74	0.419	0.310	0.238
0.5	0.76	0.405	0.308	0.228
0.5	0.78	0.365	0.285	0.214
0.5	0.80	0.351	0.281	0.201
0.5	0.84	0.230	0.193	0.150
0.5	0.86	0.203	0.174	0.150
0.5	0.88	0.189	0.166	0.136
0.5	0.90	0.176	0.158	0.126
0.5	0.92	0.081	0.075	0.048
0.5	0.94	0.068	0.064	0.048
0.5	0.96	0.054	0.052	0.037
0.5	0.98	0.041	0.040	0.037
0.5	1	0.027	0.027	0.027

Notes: This table is a reproduction of part of Table C.1 in Benoît et al. (2015). The variables are defined as follows: A fraction x of subjects believe that they are in the top y with a probability of at least q, and a fraction $\tilde{x}$ of subjects believe the same and are, in fact, in the top y.