KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE

OF

ECONOMIC RESEARCH

KYOTO UNIVERSITY

KYOTO, JAPAN

Testing for Overconfidence Statistically: A Moment Inequality Approach[*]

Yanchun Jin[†] and Ryo Okui[‡]

January 2018

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 analyzing 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, they provide a set of inequalities that are consistent with purely rational Bayesian updaters. We propose to apply 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 on safety among the US subjects tested by Svenson. However, the other cases tested do not show evidence of overconfidence.

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 in which data are obtained from "ranking experiments." 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 can confirm

[†]Graduate School of Economics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan. Email: jin.yanchun.74v@st.kyoto-u.ac.jp

[‡]Corresponding author. NYU Shanghai, 1555 Century Avenue, Pudong, Shanghai, China, 200122; and Department of Economics, University of Gothenburg, P.O. Box 640, SE-405 30 Gothenburg, Sweden. Email: okui@nyu.edu

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 placed themselves into the first decile of the distribution for driving safety. Note that there was no overlap of participants between the "safety" experiment and the "skill" experiment.

the presence of overconfidence on driving safety among US subjects tested by Svenson, we cannot reject the null hypothesis of no overconfidence in the other cases tested.

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

However, a recent study by Benoît and Dubra (2011) points out that we should be cautious when interpreting these empirical results. They argue that a typical research design used in the literature on overconfidence—the ranking experiment—may generate data that misleadingly indicate overconfidence even if respondents follow purely Bayesian updating.

In a ranking experiment, participants are asked to place themselves into one of k equally sized intervals according to their beliefs about their skills relative to other members of the groups under question. In the experiment of Svenson (1981) (details of the research design will be presented in Section 4), participants were asked to indicate the decile of the distribution regarding participants' driving abilities to which they belonged. Two abilities were examined in this research: driving safety and driving skills. Svenson conducted this experiment in the US and Sweden, and obtained the data presented in Table 1. Among the 40 US participants tested on driving safety, 22.5% answered that they belonged in the top 10% of participants for driving safety. At first glance, the data give the impression that a substantial degree of overconfidence exists.

Benoît and Dubra (2011) argue that data obtained by a ranking experiment cannot be directly used to investigate the presence of overconfidence. Nonetheless, they derive

---

[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."

a set of inequalities that responses from purely rational Bayesian updaters must satisfy.[2] For example, in Svenson (1981) data, 51.4% of the Swedish subjects thought 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 the US subjects considered that they belonged in the top 20% for skill. This result does indicate overconfidence, according to Benoît and Dubra (2011). However, the threshold above which overconfidence can be inferred is 40%, not 20%. The question that 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, their 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, then 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 testing power. Second, their approach allows us to compute p-values relatively easily. Other approaches that involve moment selection (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.

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 in relation to 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 on 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 research design, and it has been cited frequently as evidence of overconfidence in driving safety.

This paper contributes to the literature on behavioral economics by providing a statistical procedure to test for the presence of overconfidence and by reexamining the data for the seminal paper on this topic (i.e., Svenson, 1981). As previously noted, the body

---

[2]Benoît et al. (2015) further exploit the implications of Benoît and Dubra (2011) and argue that more powerful tests can be obtained if additional information, such as the fraction of subjects who place themselves in a decile and who are actually in that decile, is available.

of research on overconfidence has been growing. It has been argued 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 a firm, when an agent overestimates the marginal product of his/her effort and works harder, this bias makes all agents more productive, including the overconfident one. Dubra (2004) show that unbiased individuals 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, if the latter's biases are large and the individuals are "stubborn." The present paper considers the fundamental question of how we can 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 by Montiel Olea and Strzalecki (2014), who estimate partially identified distributions of the discount rate and hyperbolic parameter.

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 among the US subjects in regard to driving safety but not in the other cases examined by Svenson. Section 5 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 then present the formal statement of the theoretical result.

First, we consider a simple example to provide an understanding of the arguments. Suppose that we are interested in overconfidence in relation to driving skill. Here, the key point is that events that cause people to update their beliefs regarding driving skills are not frequent enough to achieve a convergence of belief for most people. For example, a car accident is arguably rare. If an individual has not experienced a car accident, the individual may consider that his/her driving skills are relatively high, but he/she does not possess a strong degree of certainty regarding this belief. On the other hand, if the individual does experience an accident, then he/she may acquire a strong belief that his/her driving skills are poor. Rational Bayesian updating requires only that the

average of beliefs weighted by their strength be correct. In this example, there are a small number of people who confidently classify their driving skills as poor and a large number of people whose beliefs result in them placing slightly higher probabilities on their skills being good. This results in apparent overconfidence in the sense that although the data indicate that the number of people who consider that they are good drivers is larger than the number of people who actually possess good driving skills, the outcome remains consistent with purely rational Bayesian updating.

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, $\theta \in \Theta$ is 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, at the initial stage, none of the drivers have information about 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_i be the i-th k-cile; i.e., $\Theta_i = \{\theta \in \Theta \mid (i - k)/k \leq p(\theta' < \theta) < i/k\}$ for $i \leq k - 1$ 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 his/her 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 $\Delta^k = \{x = (x_1, \ldots, x_k) \in [0, 1]^k \mid \sum_{i=1}^k x_i = 1\}$. Let

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

S_i is the set of signals conditional on which the median of the posterior is in Θ_i. Let $F(\cdot)$ be the marginal distribution of S so that $F(S_i)$ is the population fraction of participants for whom the medians of the posteriors are in Θ_i. 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 $x_i = F(S_i)$ for $i = 1, \ldots, k$. This means that every driver places him/herself into a certain k-cile if he/she believes that his/her actual skill places him/her in that k-cile or above with a probability of at least $1/2$ and that it also places him/her 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 there is no overconfidence (or underconfidence) affecting beliefs.

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

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

and

$$\sum_{j=1}^{i} x_j < \frac{2}{k}i. \tag{3}$$

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

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

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

$$H_0 : \mu_j \leq 0, \text{ for all } j = 1, \ldots, k, \tag{4}$$

against the alternative:

$$H_1 : \mu_j > 0, \text{ for some } j = 1, \ldots, k. \tag{5}$$

(4) is the hypothesis that all k moment inequalities are satisfied. When $k = 1$—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 notation. 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:

$$\tilde{\Sigma} = \hat{\Sigma} + \max\{\epsilon - \det(\hat{\Omega}), 0\}\hat{D}, \tag{6}$$

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 $j = 1, \ldots, k$, let $\bar{W}_j$ and S_j denote the sample mean and sample standard deviation of $W_{1j}, \ldots, W_{Nj}$, respectively; that is, $\bar{W}_j = \sum_{i=1}^N W_{ij}/N$, $S_j = \left(\sum_{i=1}^N (W_{ij} - \bar{W}_j)^2/N \right)^{1/2}$.

We consider the following three test statistics.

$$T^{\text{MAX}} = \max_{1 \leq j \leq k} \frac{\sqrt{N}\bar{W}_j}{S_j} \tag{7}$$

$$T^{\text{QLR}} = \inf_{t \in \mathbb{R}^k : t \leq 0} (\sqrt{N}\bar{W} - t)'\tilde{\Sigma}^{-1}(\sqrt{N}\bar{W} - t) \tag{8}$$

$$T^{\text{MMM}} = \sum_{j=1}^k \left(\frac{\sqrt{N}\bar{W}_j}{S_j} \right)^2 \mathbf{1}\{\bar{W}_j > 0\} \tag{9}$$

The MAX statistic, (7), is given by the maximum over k t- statistics. The QLR statistic in (8) is of quadratic form and measures the distance to the region that satisfies the inequalities. The MMM statistic in (9) may be regarded as 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 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 $E(W_{j,i})$ 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 regarded as being 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 $S_j^{(b)}$, $j = 1, \ldots, k$, for each bootstrap repetition. Then, we obtain the empirical distribution of $\max_{1 \leq j \leq k}(\sqrt{N}(\bar{W}_j - \bar{W}_j^{(b)})/S_j^{(b)})$.

---

[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 approach—mmoment selection—which drops moment inequalities that are obviously satisfied. See Allen (2018) for the difference between moment recentering and moment selection. He argues that moment recentering provides a more powerful test. We adopt the term "moment recentering" from Allen (2018).

Let $\hat{K}^{-1}(1-\beta)$ be the $(1-\beta)$ quantile of this empirical distribution. The confidence set is:

$$\hat{M}(1-\beta) = \left\{ \mu \in \mathbb{R}^k : \max_{1 \leq j \leq k} \frac{\sqrt{n}(\mu_j - \bar{W}_j)}{S_j} \leq K^{-1}(1-\beta) \right\}, \tag{10}$$

$$= \left\{ \mu \in \mathbb{R}^k : \mu_j \leq \bar{W}_j + \frac{S_j \hat{K}^{-1}(1-\beta)}{\sqrt{N}}, \text{ for all } 1 \leq j \leq k \right\}. \tag{11}$$

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

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:

$$\lambda_j^* = \min \left\{ \bar{W}_j + \frac{S_j \hat{K}^{-1}(1-\beta)}{\sqrt{N}}, 0 \right\}. \tag{12}$$

and $\lambda^* = (\lambda_1^*, \ldots, \lambda_k^*)'$. 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:

$$T^{\text{MAX},(b)} = \max_{1 \leq j \leq k} \frac{\sqrt{N}(\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^*)}{S_j^{(b)}}. \tag{13}$$

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

$$T^{\text{QLR},(b)} = \inf_{t \in \mathbb{R}^k : t \leq 0} (\sqrt{N}(\bar{W}^{(b)} - \bar{W} + \lambda^*) - t)'(\tilde{\Sigma}^{(b)})^{-1}(\sqrt{N}(\bar{W}^{(b)} - \bar{W} + \lambda^*) - t), \tag{14}$$

$$T^{\text{MMM},(b)} = \sum_{j=1}^k \left(\frac{\sqrt{N}(\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^*)}{S_j^{(b)}} \right)^2 \cdot \mathbf{1}\{\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^* > 0\}, \tag{15}$$

where $\tilde{\Sigma}^{(b)}$ is the bootstrap version of $\tilde{\Sigma}$. Note that when $\lambda_j^* < 0$, $\bar{W}_j^{(b)} - \bar{W}_j + \lambda^* = \bar{W}_j^{(b)} + S_j \hat{K}^{-1}(1-\beta)/\sqrt{N}$. Because $\hat{K}^{-1}(1-\beta)$ is the $(1-\beta)$ quantile of the bootstrap distribution of $\max_{1 \leq j \leq k}(\sqrt{N}(\mu_j - \bar{W}_j)/S_j)$, 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 take into account 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 $\{\hat{M}(1-\beta) \nsubseteq \mathbb{R}_-^k\}$ 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 US and Sweden. In each country, he 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 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 to correspond to their position in the distribution of the participants' safety (or skill) levels. Note that there were different participants involved in the two data collections relating 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 consists of the following moment inequalities. Let $D(j)_i$ denote the dummy variable, which takes a value of one if subject i places him/herself in the j-th decile. The set of moment inequalities is:[5]

$$E(D(1)_i - 0.2) < 0, \tag{16}$$

$$E(D(1)_i + D(2)_i - 0.4) < 0, \tag{17}$$

$$E(D(1)_i + D(2)_i + D(3)_i - 0.6) < 0, \tag{18}$$

---

[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.

Table 2: Test results for null hypothesis of no overconfidence

version	reject	statistics	crit.val.	p-value	selected
US safety ($N = 40$)					
MMM	Yes	21.08	14.32	0.03	(20)-(23)
QLR	Yes	13.85	8.22	0.019	(20)-(23)
MAX	Yes	3.698	2.865	0.015	(20)-(23)
US skill ($N = 41$)					
MMM	No	0.653	10.53	0.58	(20)-(23)
QLR	No	0.647	6.615	0.55	(20)-(23)
MAX	No	0.804	2.514	0.54	(20)-(23)
Sweden safety ($N = 35$)					
MMM	No	0	8.075	0.86	(20)-(22)
QLR	No	0	4.669	1.00	(20)-(22)
MAX	No	-1	2.082	1.00	(20)-(22)
Sweden skill ($N = 45$)					
MMM	No	0	3.267	0.82	(21)
QLR	No	0	3.106	0.96	(21)
MAX	No	-2.708	1.773	1.00	(21)

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. "Selected" gives the list of moment inequalities that are not affected by moment recentering.

$$E(D(1)_i + D(2)_i + D(3)_i + D(4)_i - 0.8) < 0, \tag{19}$$

$$E(D(7)_i + D(8)_i + D(9)_i + D(10)_i - 0.8) < 0, \tag{20}$$

$$E(D(8)_i + D(9)_i + D(10)_i - 0.6) < 0, \tag{21}$$

$$E(D(9)_i + D(10)_i - 0.4) < 0, \tag{22}$$

$$E(D(10)_i - 0.2) < 0. \tag{23}$$

The number of bootstrap replications is 5000.[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

---

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

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 data, we can see that 46% of US drivers placed themselves in the top 20% of the distribution for skill, which is 6% more 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 those that 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 the table 2. We can see that the critical values in panel (ii) are larger than those in panel (iii) owing to 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.

5   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 in regard to driving safety, we cannot reject the null hypothesis of no overconfidence in the other three cases examined.

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.

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.

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.

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.

Table 3: Effects of moment recentering

version	reject	statistics	crit.val.	p-value	selected
US safety ($N = 40$)					
(i): Eight moment inequalities with recentering					
MMM	Yes	21.08	14.32	0.03	(20)-(23)
QLR	Yes	13.85	8.22	0.019	(20)-(23)
MAX	Yes	3.698	2.865	0.015	(20)-(23)
(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 (20)-(23) 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	(20)-(23)
QLR	No	0.647	6.615	0.55	(20)-(23)
MAX	No	0.804	2.514	0.54	(20)-(23)
(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 (20)-(23) 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.