Erasmus School of Economics

Erasmus University Rotterdam

3 July 2021

Bachelor Thesis: Econometrics & Operations Research

Testing for student overplacement when ranking data is not available

Tobias van Lith, 502829

Supervisor:
dr. Robin Lumsdaine

Second assessor:
dr. Wendun Wang

Abstract

This research aims to find the answer to the following question: *"Are students overconfident regarding their grades?"*. The paper uses data from the Erasmus School of Economics from official results and surveys, filled in by second-year Economics students. These data are not ranking data. Because there is no standard definition of overplacement without ranking data, this paper introduces the concept of indirect overplacement. This way, overplacement can be studied without ranking data. To test for indirect overplacement, this paper uses state-of-the-art statistical methods to test for moment inequalities. This paper also considers characteristics such as age, gender, average Bachelor 1 grade and number of study hours, to find out if there is any difference between students. The research finds no signs of overconfidence in general, as well as for different characteristics.

Table of contents

1 Introduction

Many students recognise the following proclamation: “I really thought I did much better!” Sometimes students think they would have a better grade than they eventually get. This could be a sign of overconfidence. Overconfidence can be blamed for several situations, such as lawsuits or wars. The perseverance of ineffective lawsuits can be the result of plaintiffs and defendants believing that they deserve more and are overconfident that they are right. If countries, nations or terrorist groups believe they are stronger or have a better army than others, that could provide an additional incentive for battles and wars (Thompson and Loewenstein, 1992; Johnson et al., 2004). Inaccurate estimation of their grades could lead to students experiencing such ‘catastrophic’ implications, such as dropping out, lack of focus or loss of motivation (Wüst and Beck, 2018).

The main goal of this research is to examine whether overconfidence is present among students. Additionally, this paper compares the differences in overconfidence by gender, age, number of study hours and average grade of Bachelor 1. This research does not aim to find the causes of overconfidence or to explain consequences of overconfidence, if present. The paper will, however, provide some general thoughts on what this overconfidence could imply.

Overconfidence is an umbrella term for three types of psychological biases: “1. Overestimation of one’s actual performance, 2. overplacement of one’s performance relative to others and 3. overprecision of the unfounded certainty in the accuracy of one’s beliefs” (Moore and Healy, 2008; Moore and Schatz, 2017). This paper studies overconfidence as overplacement, but without ranking data. The main idea is to transform the available data into ranking data and then test for overplacement. Because there is no standard definition of overplacement without ranking data, I introduce the concept of indirect overplacement. This way, overplacement can be studied without ranking data. This is one of the main contributions of the paper.

This paper aims to answer the following research questions:

1. Do students indirectly over- or underplace themselves among other students?

2. Is there a difference between female and male students?

3. Is there a difference between students with a higher-than-average or lower-than-average Bachelor 1 grade?

4. Does age matter for indirect overplacement?

5. Does the number of study hours influence indirect overplacement?

This paper examines overplacement for students by comparing the expected grade against their actual grade for a second-year statistics course. Since the actual grade is not known before the survey[1], the average grade for Bachelor 1 is used as a proxy for the actual grade. The main variables are the actual grade and expected grade of the statistics course, the average grade for Bachelor 1 and high school performance. Other important variables include the age, gender, high school grades and number of study hours of participants. The data have been gathered by the Erasmus School of Economics from surveys and official results over a period from 2003 to 2012, filled in by second-year Economics and IBEB[2] students.

The methodology of the research consists of two parts. The first part is the transformation of the data to obtain ranking data. This is explained in the data section. The second part consists of the methodology from Jin and Okui (2020) to test for overplacement statistically.

The research finds no sign of overplacement. Also for different subsamples, divided by age, gender, etc, no signs for overplacement are found.

The main contribution of this research is the development of a test for overplacement without using ranking data and the introduction of a new concept called "indirect overplacement". Furthermore, this paper contributes to the existing literature by considering a number of characteristics such as age, gender, etc.

The remainder of this research is structured as follows. First, a brief summary of existing literature is given. Second, the paper discusses the data and transformations of the data. The third section details the statistical methodology to test for indirect overplacement. The fourth section contains the empirical results. The fifth section discusses the implications of the results and contains some conclusions. Last, I discuss several ideas for further research.

[1] The distribution should be known before the survey that provides the data for this research, more information in the data section

[2] International Bachelor of Economics and Business Economics

2 Literature

2.1 General overconfidence

This section provides a brief summary of the existing literature on overconfidence in general, its relevance in education and states my hypotheses regarding indirect overplacement. Overconfidence is an umbrella term for three types of psychological biases: "1. Overestimation of one's actual performance, 2. overplacement of one's performance relative to others and 3. overprecision in the unfounded certainty in the accuracy of one's beliefs" (Moore and Healy, 2008; Moore and Schatz, 2017). For example, overconfidence is overestimation if a student expects to have a higher grade than the student actually receives, overplacement if a student expects his or her grade to be higher than other students and overprecision if a student (mistakenly) believes that he or she accurately predicts his or her actual grade. The phenomenon of overconfidence is studied for a wide variety of topics, such as driving skills Svenson (1981), political elections Radzevick and Moore (2009) and stocks/securities markets Daniel et al. (2005).

2.2 Overconfidence in classroom

A common way to study overconfidence is by comparing how one thinks one carried out a certain task and how well one actually did it (Adams and Adams, 1960). In this paper, as well as in the following literature, these certain tasks are exams, where a student's performance is based on the exam grade. The level of over- or underconfidence differs between easy and difficult tasks. Moore and Healy (2008) find that, on difficult tasks such as exams, people tend to overestimate their own performance but underplace themselves among others. On the other hand, on easy tasks, people tend to underestimate their own performance but overplace themselves among others. These findings show that it is important to make the distinction between overestimation and overplacement, since some of the previous literature failed to do so.

Students tend to make an overconfident estimation of their future exam grades. Nowell and Alston (2007) find that most students overestimate their exam grades. This overestimation is present at all points in the term, but accuracy of their estimates improved throughout the term (Beyer, 1999). In macro-economic classes, students have the tendency to be overconfident in evaluating their performance (Grimes, 2002). Some research, such as Boud and Falchikov (1989), however, do not come to the same conclusion that students are overconfident in assessing their performance. The authors find no indication of either over- or underconfidence. There are even a few signs of underconfidence in the classroom, especially for female and well-performing students (Wüst and Beck, 2018).

The level of overconfidence, or underconfidence, differs per type of student. Nowell and Alston (2007) and Beyer (1999) state that male students are more overconfident than female students in economics courses. The same result applies for statistical courses (Magnus and Peresetsky, 2018). Grimes (2002) found that expressed overconfidence reduces as students get older.

Furthermore, students with a lower overall academic performance are more overconfident than better performing students (Grimes, 2002; Nowell and Alston, 2007). To give one example, Stinson and Zhao (2008) find that students on average overestimate their grades by five points (of a score between 0 and 100). This increases to the equivalent of a letter grade (ten points) for freshmen students with low cumulative grade point averages.

Nowell and Alston (2007) found that overconfidence is greater for economics students who study longer hours. Students who studied more hours expected higher grades. The more experienced a student becomes, the greater his or her confidence level in receiving a high grade. This is confirmed by Grimes (2002), who noted that students tend to be more overconfident when they have previous experience in a certain field of study.

2.3 Hypotheses for indirect overplacement

Indirect overplacement is a newly introduced concept. Therefore, it is hard to formulate a hypothesis, since existing literature is based on existing definitions (overestimation, overplacement and overprecision). Nevertheless, I try to formulate a hypothesis for the indirect overplacement of students. Existing literature suggests the following hypotheses:

1. It can both be argued that there is either indirect over- or underplacement among students.
2. Male students show more indirect overplacement than female students.
3. Students with a higher-than-average grade show less indirect overplacement than students with a lower-than-average grade.
4. A higher age results in a decrease in indirect overplacement.

5. Students that study more than average show less indirect overplacement.

3 Data

The data are gathered by the Erasmus School of Economics from surveys and official results over a period from 2003 to 2012. These data are used in the course Academic Skills of the Bachelor 'Econometrics and Operations Research' and is merged by Erik Kole[3]. Second-year Economics and IBEB students annually fill out a survey with 36 questions regarding their grades, interests and personal information. This leads to a dataset with 4872 participants and 47 variables. The main variables are the expected grade for Applied Statistics 2 (AS2), the actual grade for AS2, the average grade of Bachelor 1 and high school performance (mathematics, economics and English grade). Other important variables include age, gender and number of hours a student studies per week.

The dataset shows a number of missing and aberrant observations. The missing observations are either denoted by NA or unrealistic values, such as 99 for a grade between 1 and 10. For the primary analysis[4], the approach to remove missing observations is as follows: if one or more main variables are missing for a certain participant, then the whole participant is removed from the data set. Following this procedure, a total of 2640 observations remain. These contain complete information for all main variables. The remaining dataset is from now on called the general dataset.

For the secondary analysis, the approach above is repeated, but also for variables such as age, gender and number of study hours per week. After this procedure, 2438 observations remain with complete information for all used variables. This approach is chosen because the primary analysis needs as many observations as possible. Furthermore, the removal of all missing observations for the secondary analysis is chosen because this way, all secondary analyses are made with the same data set. The remaining dataset is from now on called the subsample dataset.

[3] I would like to thank dr. Erik Kole for providing this data for my thesis and additional information on the survey.

[4] discussed in subsection 3.1

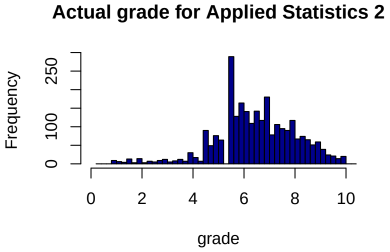

Figure 1: Histogram actual grade for AS2

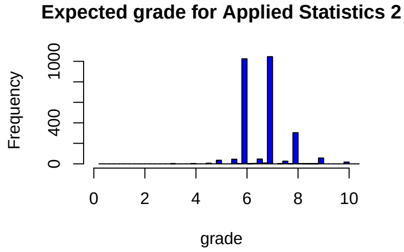

Figure 2: Histogram expected grade for AS2

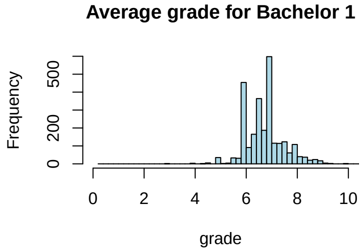

Figure 3: Histogram average Bachelor 1 grade

The statistics of the average grade of Bachelor 1, the actual grade for Applied Statistics 2 and its expected grade are based on the general dataset. The actual grades for AS2 are not taken from the annual survey, but from the official grades from students willing to share their grade. The grade is rounded up to one decimal point. Figure 1 shows that the distribution of the actual grade is widespread, with scores from the minimum of 1 to the maximum of 10. The distribution peaks at 5.5, the cut-off point for a sufficient grade. Furthermore, there are no observations equal to the values of 5.4 and 5.3. This could provide an indication that teachers round up these values to a sufficient grade of 5.5. The mean grade for AS2 is 6.51 with a standard deviation of 1.58.

The expected grade for Applied Statistics 2 was asked in the survey on a scale from 1 to 10. The survey was always taken at the beginning of the year, thus before the final exam. Therefore, this expected grade can not be interpreted as an informed expectation, but more as a goal or desire. The histogram in Figure 2 displays the distribution of the expected grade, that shows peaks for the integers, especially 6, 7 and 8. This can be explained by rounding and focal points in survey responses, as described in Kleinjans and Soest (2014). Other observations are barely visible. The mean expected grade for AS2 is 6.72 with a standard deviation of 0.84.

The survey asked the following for the average grade for Bachelor 1: *"What was your average exam grade obtained on the courses of the first year? (Give you answer only in numbers, so not in letters, and give your answer with one decimal.)"* Students could not look up their average grade, so students had to guess their average Bachelor 1 grade, resulting in peaks at 6, 6.5 and 7. This can again be explained by rounding and focal points in survey responses, as described in Kleinjans and Soest (2014). The mean average grade for Bachelor 1 is 6.78 with a standard deviation of 0.74. Since this mean is almost equal to the mean of the expected grades, a mapping is performed (described in the data section). Then, the mean is equal to the lower mean of the actual grades, which is important to test for possible overplacement.

Other main variables include age, gender, number of hours a student studies per week and high school performance in mathematics, economics and English. The following statistics are based on the subsample dataset. The median age of the participants is 20 years old, which seems reasonable for second year students. The participants are 70.6% male and 29.4% female, which is quite similar to the actual percentages of 67% and 33%[5]. The average high school grades for mathematics, economics and English are respectively 7.22, 7.47 and 6.91. The average student studies 12.6 hours per week.

3.1 Data transformation

Before using the data to assess overplacement, the data are transformed in order to get ranking data. The first step is finding a proxy for the actual grades and transforming its distribution with mappings such that it resembles the distribution of the actual grades. The second step consists of using this distribution to make ranking data.

The average of Bachelor 1 is used as a proxy for the actual grade for Applied Statistics 2 because it is known at the time of the survey. This is because the expected grade and the variable used for the ranking should be available at the time of the survey. Then, all information is available at the time of the survey and then it can be compared. Since the proxy (average grade of Bachelor 1) contains only numbers with one decimal point, a ranking with clear cut-off points (later called bounds of deciles) can not be made. Therefore, uniformly distributed random noise up to 0.05 is added to this variable, positive for a higher-than-average high school grade and negative for a lower-than-average high school grade, where the high school grade is the average of the mathematics, economics and English grade. I choose a maximum noise of 0.05 such that a grade can not increase or decrease to another decimal point.

Also, I introduce three mappings such that the distribution of the average Bachelor 1 grade looks more similar to the actual grades. The first mapping is to adjust the mean of the proxy such that it is equal to the mean of the actual grades. This is done by

[5] https://www.studiekeuze123.nl/opleidingen/1064-economie-en-bedrijfseconomie-erasmus-universiteit-rotterdam-wo-bachelor gives these percentages

subtracting all observations with the difference of the mean of the actual grades and the mean of the proxy:

$$B_i = B_i - \bar{B} + \bar{A}$$

where B_i is the i^{th} observation of the average Bachelor 1 grade, $i = 1, ..., N$, and $\bar{B} = \frac{1}{N}\sum_{i=1}^{N} B_i$ is its sample mean. $\bar{A} = \frac{1}{N}\sum_{i=1}^{N} A_i$ is the sample mean of the actual grades A_i , $i = 1, ..., N$. The result is that all observations are decreased by 0.28, which is the difference between $\bar{B}$ and $\bar{A}$.

The second mapping is to redistribute the observations at the peaks (6.0, 6.5 and 7.0) to either ends of the grading scale. The average Bachelor 1 grade has most observations around the peaks and not much observations at the low or very high grades. So some observations are redistributed from those peaks to grades around the maximum and minimum. This is done by using uniformly distributed random noise, positive for a higher-than-average high school grade and negative for a lower-than-average high school grade on intervals defined below. Since the mean is shifted in the first mapping, the second mapping takes into consideration that the peaks occur at different values by redistributing from the "new" peaks, namely 5.72, 6.22 and 6.72 (a decrease of 0.28). The observations equal to 5.72 are redistributed on the interval [1.0 , 5.72] for a lower-than-average high school grade and on the interval [5.72 , 7.0] for a higher-than-average high school grade. For observations equal to 6.22, the intervals are respectively [5.3 , 6.22] and [6.22 , 8.0]. For observations equal to 6.72, the intervals are [5.3 , 6.72] and [6.72 , $\max(A_i)$] where $\max(A_i)$ is the maximum actual grade for AS2.

The third mapping is to get more values to 5.5. Teachers often round 5.3 and 5.4 to 5.5 such that these students have a sufficient grade. That is why, in the second mapping, some peaks are redistributed up to 5.3, such that there are more values of 5.3 and 5.4. This way, the transformed data has more values for 5.5, which is a property of the actual grade for AS2. The transformed data form the proxy for the actual grade of Applied Statistics 2.

After the data are transformed, a ranking can be made. First, the transformed proxy is sorted from the lowest value to the highest value. Then, the data are divided into ten equal deciles. Each decile has two values bounding its interval to make sure that each decile contains $1/10^{\text{th}}$ of the sample. Finally, in order to get the fractions of students who are placed in the k^{th} decile according to their expectations, a student is placed in the k^{th} if his or her expected grade is between the bounds of the k^{th} decile. Students do not place themselves in a certain decile, but are 'indirectly' placed according to their expected grade.

A number of different sets of data are used to conduct the research. First, the general dataset is used for the primary analysis with the whole sample. The primary analysis includes three separate analyses. The first analysis makes the ranking with the proxy and the mappings. The second analysis makes the ranking with the proxy, but without the mappings to see whether the mappings have any effect. The third analysis makes the ranking with the actual grade to observe whether the average grade of Bachelor 1 is a good proxy.

Second, the subsample dataset is used for the secondary analysis. This analysis aims to find out if there are differences in indirect overplacement between types of students; male and female students, younger and older students, etc. Therefore, the subsample dataset is divided in two parts according to a certain division rule. For gender, the sample is naturally split by male/female. For age, the sample is split between students under 20 and students equal to or above 20. For the average grade of Bachelor 1, the students are split in students with a higher-than-average and lower-than-average Bachelor 1 grade. For the number of study hours, the mean number of study hours splits the sample.

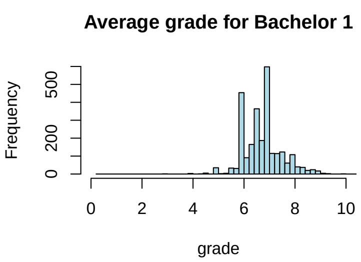

Figure 4: Histogram average Bachelor 1 grade

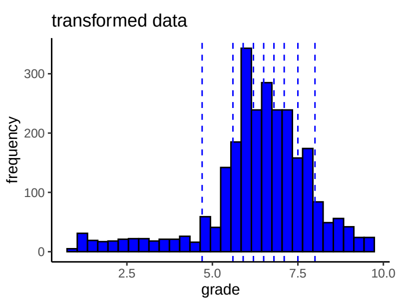

Figure 5: Histogram transformed average Bachelor 1 grade

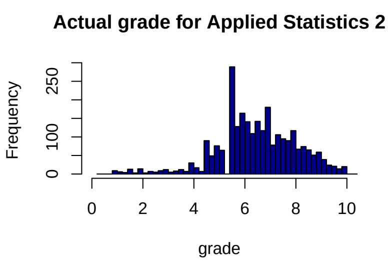

Figure 6: Histogram actual grade for AS2

Figure 5 shows the distribution of the transformed proxy for the general dataset. The decile bounds, represented by the dashed lines, are based on this distribution. The decile bounds in this example are 4.71, 5.62, 5.93, 6.26, 6.57, 6.84, 7.11, 7.56 and 8.02. The figure illustrates where the decile bounds are. Note that the distribution, and therefore the bounds, for the other analyses differ from this distribution. Figure 4 shows the distribution of the average Bachelor 1 grade before the mappings to show how the mappings changed the distribution. Figure 6 shows the distribution of the actual grades for AS2, the distribution that is approximated by the transformed data.

The ranking data, formed by the application of the mappings on the proxy and ranking procedure, look as follows:

Deciles	1	2	3	4	5	6	7	8	9	10	N
Main	0.4	2.9	0.2	38.8	0.2	2.1	39.6	1.0	0.1	14.7	2640

Table 1: Percentages for primary analysis

Notes: This is a table of the percentages for the primary analysis. The table indicates the percentage of participants placed in the k^{th} decile according to their expectations. N denotes the number of observations.

Table 1 shows the percentage of participants placed in the k^{th} decile according to their expectations for the primary analysis. For example, 14.7% of the students believe that they are in the top 10%, as indicated by their expected grade. The percentages of the other analyses can be found in Appendix A.

4 Methodology

After the transformation of the data such that it became ranking data, this section describes the moment inequalities and the statistical methods to test for indirect overplacement.

4.1 Moment inequalities

Along with the theory of apparent overconfidence of Benoît and Dubra (2011)[6], the following moment inequalities are chosen (after eliminating unnecessary items) :

[6] For more information regarding this theory, I refer to the paper of Benoît and Dubra (2011)

$$x_1 \leq 0.2 \tag{1}$$

$$x_1 + x_2 \leq 0.4 \tag{2}$$

$$x_1 + x_2 + x_3 \leq 0.6 \tag{3}$$

$$x_1 + x_2 + x_3 + x_4 \leq 0.8 \tag{4}$$

$$x_7 + x_8 + x_9 + x_{10} \leq 0.8 \tag{5}$$

$$x_8 + x_9 + x_{10} \leq 0.6 \tag{6}$$

$$x_9 + x_{10} \leq 0.4 \tag{7}$$

$$x_{10} \leq 0.2 \tag{8}$$

where x_k is the fraction of subjects who are placed in the k^{th} decile. These inequalities are the same as in the paper of Jin and Okui (2020). The theory of apparent overconfidence is based on ranking data. Even though the data are strictly not ranking data, I assume it to be. Therefore, the theory of apparent overconfidence can be applied to form the moment inequalities.

4.2 Testing procedure for moment inequalities

This section provides the testing procedure by Romano et al. (2014) in a general setting with some supplementary explanation for the moment inequalities in this paper. The notations and definitions in the methodology are directly adopted from the paper of Jin and Okui (2020), where the testing procedure of Romano et al. (2014) is used to test for overplacement of driving skill and safety.

Let $W_i = (W_{i1}, W_{i2}, ..., W_{iJ})', i = 1, ..., N$ be an i.i.d. sequence of random vectors with mean $\mu \in \mathbb{R}^J$. Let μ_j be the j^{th} component of μ, such that $\mu_j = E(W_{ij})$ for $1 \leq j \leq J$. This paper examines the following testing problem with the null hypothesis that all moment inequalities are satisfied

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

against the alternative hypothesis that one or more inequalities are not satisfied

$$H_0 : \mu_j > 0, \quad \text{for some } j = 1, ..., J.$$

In this paper, for example, W_{i2} corresponds to $D(1)_i + D(2)_i - 0.4$, where $D(j)_i$ denotes the dummy variable, which is equal to 1 if participant i is in the j^{th} decile. Then, $\mu_2 = E(D(1)_i + D(2)_i - 0.4) = x_1 + x_2 - 0.4$.

If there is only one moment inequality, a standard t-test may be used. If there are more moment inequalities, the procedure is not that straightforward. Let $\bar{W} = \frac{1}{N}\sum_{i=1}^{N} W_i$ be the sample average of W_i and let $\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(W_i - \bar{W})(W_i - \bar{W})'$ be the sample variance covariance matrix of W_i. This paper uses a regularized version $\tilde{\Sigma}$ that is always invertible:

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

with $\hat{D} = \text{diag}(\hat{\Sigma})$, $\hat{\Omega} = \hat{D}^{-1/2}\hat{\Sigma}\hat{D}^{-1/2}$ and $\epsilon = 0.012$. ϵ ensures that $\tilde{\Sigma}$ is invertible, as shown in Romano et al. (2014). Let $\bar{W}_j = \frac{1}{N}\sum_{i=1}^{N} W_{ij}$ and $\hat{\sigma}_j = \left(\frac{1}{N}\sum_{i=1}^{N}(W_{ij} - \bar{W}_j)^2\right)^{1/2}$ be the sample mean and variance of $W_{1j}, ..., W_{Nj}$ for $j = 1, ..., J$. Now consider the following three test statistics:

$$T_{\text{MAX}} = \max_{1\leq j\leq J} \frac{\sqrt{N}\bar{W}_j}{\hat{\sigma}_j}$$

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

$$T_{\text{MMM}} = \sum_{j=1}^{J} \left(\frac{\sqrt{N}\bar{W}_j}{\hat{\sigma}_j}\right)^2 \mathbf{I}\{\bar{W}_j > 0\}$$

The MAX statistic is the maximum over J t-statistics. The QLR (quasi likelihood ratio) statistic measures the distance to the region that satisfies the inequalities. The MMM (modified method of moments) statistic is an adjusted version of QLR. The difference is that the MMM statistic ignores the correlation across elements of $\bar{W}$.

The critical values are computed by bootstrap with moment recentering. This procedure is similar to the least favorable distribution approach, where the critical values are based on the distribution that satisfies $E(W_i) = 0$. However, allowing some of the moments to be negative significantly improves the power of the test. This moment recentering procedure is used during the computation of the bootstrap equivalent of the distribution, when the inequalities are clearly satisfied. This moment recentering approach, formulated by Romano et al. (2014), consists of two steps. First, the approach constructs an upper confidence bound for the moments by bootstrap. Second, the critical values are computed with moment recentering in the bootstrap distributions.

First, a confidence region for the moments is constructed. The main goal of this confidence region is to find the upper confidence bound. A confidence interval is constructed for each $E(W_{ij})$ at a confidence level of $(1-\beta)$, with α being the nominal size of the test and $\beta = \alpha/10$ in this application. To obtain this confidence region, bootstrapping is used. In each bootstrap iteration, N observations are taken from the sample with repetition. Let $W_i^{(b)}$, $i = 1, ..., N$ be a sample in the b^{th} bootstrap iteration, where $b = 1, ..., B$ and B is the number of bootstrap iterations. $\bar{W}_j^{(b)}$ and $\hat{\sigma}_j^{(b)}$, $j = 1, ..., J$, are computed for each bootstrap iteration. Take $\hat{\sigma}_j^{(b)} = \max\{\epsilon_b, \hat{\sigma}_j^{(b)}\}$ where $\epsilon_b = 0.001$ because $\hat{\sigma}_j^{(b)}$ can be equal to zero in some bootstrap iterations. If ϵ is small enough, it does not influence the outcomes. The MAX statistic operates as an example in the remainder of the paper. We obtain the following empirical distribution

$$\max_{1\leq j\leq J} \frac{\sqrt{N}(\bar{W}_j^{(b)} - \bar{W}_j)}{\hat{\sigma}_j^{(b)}}. \tag{9}$$

Let $\hat{L}^{-1}(1-\beta)$ be the $(1-\beta)$ quantile of this empirical distribution. Consequently, the upper confidence bound is formed for each μ_j by $\bar{W}_j + \hat{\sigma}_j \hat{L}^{-1}(1-\beta)/\sqrt{N}$.

The second step consists of the moment recentering procedure and with that adjusting the bootstrap distribution. Then, the critical values are computed with this adjusted bootstrap distribution. Here follows the explanation of the moment recentering procedure. Essentially, the means of the moment inequalities are adjusted if the upper confidence bound is below zero. Let $\lambda_j^* = \min\left\{\bar{W}_j + \hat{\sigma}_j \hat{L}^{-1}(1-\beta)/\sqrt{N}, 0\right\}$ and $\lambda^* = (\lambda_1^*, ..., \lambda_J^*)'$. The mean value μ_j under the bootstrap distribution, $\bar{W}_j$, is substituted with $\bar{W}_j - \lambda_j^*$. This results in the empirical distribution of

$$T_{MAX,(b)} = \max_{1\leq j\leq J} \frac{\sqrt{N}(\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^*)}{\hat{\sigma}_j^{(b)}}. \tag{10}$$

Similarly, to complete the test statistics, the empirical distributions of T_{QLR} and T_{MMM} are

$$T_{QLR,(b)} = \inf_{t\in\mathbb{R}^J: t\leq 0} (\sqrt{N}(\bar{W}^{(b)} - \bar{W} + \lambda^*) - t)'\tilde{\Sigma}^{-1}(\sqrt{N}(\bar{W}^{(b)} - \bar{W} + \lambda^*) - t) \tag{11}$$

$$T_{MMM,(b)} = \sum_{j=1}^{J} \left(\frac{\sqrt{N}(\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^*)}{\hat{\sigma}_j^{(b)}} \right)^2 \mathbf{I}\{(\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^*) > 0\} \tag{12}$$

where $\tilde{\Sigma}^{(b)}$ is the bootstrap version of $\tilde{\Sigma}$. If the upper confidence bound is greater than zero, then $\lambda_j^* = 0$ and the bootstrap distribution is the same as in equation (9), when we continue with the example of the MAX statistic. If $\lambda_j^* < 0$, then $\bar{W}_j^{(b)} - \bar{W}_j + \lambda_j^*$ is equal to $\bar{W}_j^{(b)} + \hat{\sigma}_j \hat{L}^{-1}(1-\beta)/\sqrt{N}$. Because $\hat{L}^{-1}(1-\beta)$ is the $(1-\beta)$ quantile of the bootstrap distribution of $\max_{1 \leq j \leq J} \frac{\sqrt{N}(\mu_j - \bar{W}_j)}{\hat{\sigma}_j}$, it is not likely that the moment with $\lambda_j^* < 0$ affects the bootstrap distribution of a test statistic. This finishes the procedure of moment recentering and the adjusted bootstrap distribution remains.

Before the critical values are computed with the adjusted bootstrap distributions, the effect of moment recentering needs to be considered. The probability of making an error in the recentering procedure is β. Therefore, instead of the $(1-\alpha)$ quantile of the adjusted bootstrap distribution, the $(1-\alpha+\beta)$ quantile should be used because this quantile is a bit higher to counteract the possibility of making a mistake in the recentering procedure.

Let the test statistic T be either T_{MAX}, T_{QLR} or T_{MMM} and let $\hat{c}(1-\alpha+\beta)$ be the corresponding critical value. The test rejects the null hypothesis at size α if the upper confidence bound $\hat{L}^{-1}(1-\beta) > 0$ and $T > \hat{c}(1-\alpha+\beta)$.

5 Results

Table 2 summarizes the results of the overconfidence test for students. "Transformed" uses the data with the proxy (average grade for Bachelor 1) and the mappings such that the proxy resembles the actual grades. "Not transformed" uses the data with the proxy, but without the mappings. This way, it can be checked whether the mappings on the proxy have any effect. "Actual" uses the data with the actual grades, so no proxy and no mappings. Although it is not feasible to use[7], it can still be useful to compare the differences with "Transformed".

The table shows the MAX, QLR and MMM statistics along with their critical values and p-values. For "Transformed", the null hypothesis of no overplacement is not rejected on a 5% significance level for all three test statistics. In other words, students do not (indirectly) over- or underplace themselves among other students in general. The three test statistics give the same results for "Not transformed" and "Actual". A noticeable result is that the QLR and MMM statistics are equal to zero in all three cases. These statistics

[7]the distribution of the variable used to make the ranking should be known at the time of the survey, see data section for more information

denote the distance to the region that satisfies the moment inequalities. Therefore, if the test statistics are equal to zero, then the test statistic is within this region and the null hypothesis is not rejected.

Version	**Reject**	**Statistics**	**Crit. value**	***p*-value**	**Not recent.**
Transformed	N=2640				
MAX	No	-1.21	1.64	0.97	8
QLR	No	0	2.65	0.50	8
MMM	No	0	3.11	0.50	8
Not transformed	N=2640				
MAX	No	-1.21	1.63	0.97	8
QLR	No	0	2.69	0.75	8
MMM	No	0	3.42	0.50	8
Actual	N=2640				
MAX	No	-3.38	1.27	1	-
QLR	No	0	1.59	0.64	-
MMM	No	0	1.61	0.18	-

Table 2: Results primary analysis

Notes: This is a table of the results of the primary analysis. Version is the type of statistic (MAX, QLR and MMM). "Reject" denotes whether the test for overplacement is rejected at the 5% significance level. "Statistics" gives the value of the test statistics. "Crit. value" denotes the critical value that is computed by bootstrap. It is the $(1-\alpha+\beta)$ quantile of the adjusted bootstrap distribution, as explained in the methodology. The p-value is naturally given by "p-value". "Not recent." gives the moment inequalities that are not affected by moment recentering. This is the case when $\lambda_j^* = 0$ for inequality j. N is the number of observations. Table 3 has the same outline and should be similarly interpreted.

Version	Reject	Statistics	Crit. value	p-value	Not recent.
Male	N=1722				
MAX	No	-0.28	1.87	0.882	6,8
QLR	No	0	3.61	0.68	6,8
MMM	No	0	6.60	0.75	6,8
Female	N=716				
MAX	No	-4.2	0.45	1	-
QLR	No	0	0.19	0.12	-
MMM	No	0	0.20	0.07	-
Version	**Reject**	**Statistics**	**Crit. value**	p**-value**	**Not recent.**
$<$ ***20***	N=1046				
MAX	No	-4.58	0.14	0.94	-
QLR	No	0	0.02	0.19	-
MMM	No	0	0.02	0.06	-
$\geq$ ***20***	N=1392				
MAX	No	-0.84	1.92	1	6,8
QLR	No	0	4.09	1	6,8
MMM	No	0	5.99	0.69	6,8
Version	**Reject**	**Statistics**	**Crit. value**	p**-value**	**Not recent.**
Higher ave.	N=1276				
MAX	No	-3.04	1.28	1	-
QLR	No	0	1.63	0.367	-
MMM	No	0	1.76	0.22	-
Lower ave.	N=1162				
MAX	No	0.28	1.70	0.43	6,7
QLR	No	0.08	3.81	0.45	6,7
MMM	No	0.08	6.17	0.45	6,7

Version	Reject	Statistics	Crit. value	p-value	Not recent.
Study hard	N=968				
MAX	No	-3.12	1.65	1	-
QLR	No	0	2.69	0.94	-
MMM	No	0	2.72	0.29	-
Study little	N=1470				
MAX	No	-0.84	1.97	1	6,8
QLR	No	0	3.82	0.69	6,8
MMM	No	0	6.00	0.79	6,8

Table 3: Results secondary analysis

Notes: This table has the same outline as Table 2 and should be similarly interpreted.

Table 3 gives the results of the secondary analysis for a number of subsamples. "Male" represents the subsample of male students. "Female" is the subsample of female students. "< 20" forms the subsample of students under 20 and "≥ 20" is the subsample of students above or equal to 21. "Higher ave." represents the subsample of students with a higher-than-average Bachelor 1 grade. "Lower ave." shows the subsample of students with a lower-than-average Bachelor 1 grade. "Study hard" represents the subsample of students who study more than average and "Study little" the subsample of students who study less than average. For all subsamples, the null hypothesis of no overplacement is not rejected on a 5% significance level. For example, if we only look at male students, these students do not show overplacement among other male students. The same result applies for female students as well as for students that are younger than 20 and students that are older or equal to 20. It also applies for students with a higher-than-average and lower-than-average Bachelor 1 grade. Also students who study more than average and less than average do not show overplacement among others.

The procedure is implemented in R 4.0.4. The code for testing the moment inequalities is directly used from the research of Jin and Okui (2020). The number of bootstrap iterations is equal to 5000.

6 Conclusion

This paper researched whether overplacement is present among students. Additionally, this paper compared the differences in overplacement by gender, age, number of study hours and average grade of Bachelor 1. This paper uses data from second-year Economics students from the Erasmus University. These data were transformed to ranking data and indirect overplacement was tested with state-of-the-art statistical methods. The research found no signs of indirect overplacement. Also, for students with different characteristics (age, gender, average grade Bachelor 1 and number of study hours) no clear overplacement was found. The main contribution of this research was the development of a test for overplacement without using ranking data and the introduction of a new concept called "indirect overplacement". Furthermore, this paper contributed to the existing literature by considering a number of characteristics (age, gender, ...). Ideas for further research are discussed in the next section.

7 Discussion

In this section, I provide some ideas for further research and discuss the differences between the results in this paper and existing literature. Regarding the data manipulation, there are many possibilities to improve the available data. This paper chose to delete a participant when one variable is not available. Other approaches could have been sample matching (Stuart, 2010), a Bayesian probability matching approach (Datta et al., 2000) or to include dummy variables. The dummy variable for a certain participant takes a value of 1 if a variable is available in the data, and 0 otherwise. With this technique, the number of observations can be maximised without making changes to the data. The reason that this paper did not use the dummy variable approach was to have a consistent dataset in the secondary analysis such that the subsamples could be split from the same dataset.

Regarding the methods, I could have used the moment inequality approach without recentering, compare a number of different bootstrap replications or the moment selection approach of Chernozhukov et al. (2019) and Allen (2018).

Further research could include examining the (catastrophical) implications of overconfidence, such as dropout rates, lack of focus or laziness (Wüst and Beck, 2018). On the other hand, research could also focus on positive implications of overconfidence, for example the positive effect of an overconfident and optimistic manager on the value of a firm (Gervais et al., 2002). Since the available data include the number of hours a student studies per week, further research could examine whether there is a causal effect of overconfidence on over- or understudying or vice versa. This would require additional data. For example, a first survey may ask the number of hours studied per week and expectations at the beginning of the year. Then, a second survey, at the end of the year, asks the number of hours studied per week. Then, it is possible to see whether there was a change in the number of study hours (Wüst and Beck, 2018). The data also contain values for stress levels and how students feel in general. Currently, burn-outs or depressions among students are becoming more frequent and thus increasingly important (Lipson et al., 2019). Therefore, it might seem interesting to research this.

This paper introduces a new way to research overplacement when ranking data is not available. It may be used for researching overplacement in other applications than student behavior. In other words, this procedure increases the range of possibilities to examine overplacement.

Using the distribution of the actual grade of AS2 for the mappings raises some robustness issues. As mentioned in the data section, the distribution of the variable that is used to make the ranking must be known at the time of the survey. This is not the case when the actual AS2 grade is used to make the ranking (the grade is only known after the exam, so also after the survey). This could result in correlation, dependence or even entirely wrong results. When the average of Bachelor 1 is taken as a proxy and the actual grade is used to transform the distribution, the distribution of the actual grade is still used to make the ranking. This could bring some issues.

The data in this paper is, at first, not ranking data. Because of the assumption that it is real ranking data, even though it is constructed from non-ranking data, the theory of apparent overconfidence of Benoît and Dubra (2011) can be applied. If this assumption was lifted, the moment inequalities would be less strict. Then, perhaps, the likelihood of finding overplacement is much higher. Basically, it could be the case that the moment inequalities are too strict which leads to the result of no overplacement. Both these robustness issues could possibly explain the difference in the results of this paper and existing literature.

References

P. A. Adams and J. K. Adams. Confidence in the recognition and reproduction of words difficult to spell. *The American Journal of Psychology*, 73(4):544–552, 1960.

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

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

S. Beyer. Gender differences in the accuracy of grade expectancies and evaluations. *Sex Roles*, 41(3):279–296, 1999.

D. Boud and N. Falchikov. Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher education*, 18(5):529–549, 1989.

V. Chernozhukov, D. Chetverikov, and K. Kato. Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies*, 86 (5):1867–1900, 2019.

K. Daniel, D. Hirshleifer, and A. Subrahmanyam. Investor psychology and security market under-and overreaction. *Advances in Behavioral Finance, Volume II*, pages 460–501, 2005.

G. S. Datta, M. Ghosh, and R. Mukerjee. Some new results on probability matching priors. *Calcutta Statistical Association Bulletin*, 50(3-4):179–192, 2000.

S. Gervais, J. Heaton, and T. Odean. The positive role of overconfidence and optimism in investment policy. 2002.

P. W. Grimes. The overconfident principles of economics student: An examination of a metacognitive skill. *The Journal of Economic Education*, 33(1):15–30, 2002.

Y. Jin and R. Okui. Testing for overconfidence statistically: A moment inequality approach. *Journal of Applied Econometrics*, 35(7):879–892, 2020.

D. D. Johnson et al. *Overconfidence and war*. Harvard University Press, Cambridge, Massachusetts, United States of America, 2004.

K. J. Kleinjans and A. V. Soest. Rounding, focal point answers and nonresponse to subjective probability questions. *Journal of Applied Econometrics*, 29(4):567–585, 2014.

S. K. Lipson, E. G. Lattie, and D. Eisenberg. Increased rates of mental health service utilization by us college students: 10-year population-level trends (2007–2017). *Psychiatric services*, 70(1):60–63, 2019.

J. R. Magnus and A. A. Peresetsky. Grade expectations: Rationality and overconfidence. *Frontiers in psychology*, 8:2346, 2018.

D. A. Moore and P. J. Healy. The trouble with overconfidence. *Psychological review*, 115 (2):502, 2008.

D. A. Moore and D. Schatz. The three faces of overconfidence. *Social and Personality Psychology Compass*, 11(8):e12331, 2017.

C. Nowell and R. M. Alston. I thought I got an A! Overconfidence across the economics curriculum. *The Journal of Economic Education*, 38(2):131–142, 2007.

J. R. Radzevick and D. A. Moore. Competing to be certain (but wrong): Social pressure and overprecision in judgment. *Academy of Management Proceedings*, 2009(1):1–6, 2009.

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

T. A. Stinson and X. Zhao. Unmet expectations: why is there such a difference between student expectations and classroom performance? *Journal of College Teaching & Learning (TLC)*, 5(7), 2008.

E. A. Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1, 2010.

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

L. Thompson and G. Loewenstein. Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes*, 51(2):176–197, 1992.

K. Wüst and H. Beck. “I thought I did much better”—overconfidence in university exams. *Decision Sciences Journal of Innovative Education*, 16(4):310–333, 2018.

Appendices

A Percentages

Deciles	1	2	3	4	5	6	7	8	9	10
Actual	0.4	1.3	1.8	38.9	0.2	2.1	39.7	1.0	11.7	2.9
Non-transformed	3.5	38.9	0.2	1.8	0.3	0.0	39.6	0.0	1.1	14.6
male	0.4	2.2	0.4	36.9	0.2	1.7	0.5	41.3	1.0	15.4
female	0.4	4.6	41.8	0.1	2.1	0.0	36.5	0.8	11.0	2.7
young	2.3	1.2	35.3	1.5	0.1	42.1	0.2	1.2	13.2	2.9
old	0.0	1.4	1.9	0.2	40.7	0.1	2.2	38.7	0.7	14.1
smart	2.1	27.4	1.4	0.1	0.3	45.1	0.1	1.3	17.2	5.0
dumb	0.0	2.3	2.6	0.2	50.5	0.0	0.3	2.2	34.5	7.4
study hard	0.2	2.5	36.1	0.2	2.0	0.5	39.4	1.2	13.3	4.6
study little	0.5	3.3	0.2	40.0	0.2	1.7	0.3	40.2	0.8	12.8

Table 4: Percentages for the following analyses: the remaining two analyses from the primary analysis and all of the analyses from the secondary analysis

B List of abbreviations

1. IBEB: International Bachelor of Economics and Business Economics

2. AS2: Applied Statistics 2