Journal of Behavioral and Experimental Economics 68 (2017) 18–24

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Estimating the relationship between skill and overconfidence

Jan Feld[a,b,*], Jan Sauermann[b,c,d,e], Andries de Grip[b,e]

[a] School of Economics and Finance, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand
[b] Institute for the Study of Labour (IZA), Schaumburg-Lippe-Strasse 5-9, 53113 Bonn, Germany
[c] Swedish Institute for Social Research (SOFI), Stockholm University, Universitetsvägen 10 F, 106 91 Stockholm, Sweden
[d] Center for Corporate Performance (CCP), Copenhagen Business School, Porcelænshaven 24, 2000 Frederiksberg, Denmark
[e] Research Centre for Education and the Labour Market (ROA), Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

ARTICLE INFO

Article history:
Received 19 June 2016
Revised 9 March 2017
Accepted 10 March 2017
Available online 16 March 2017

Keywords:
Dunning–Kruger effect
Overconfidence
Judgment error
Measurement error
Instrumental variable

ABSTRACT

The Dunning–Kruger effect states that low performers vastly overestimate their performance while high performers more accurately assess their performance. Researchers usually interpret this empirical pattern as evidence that the low skilled are vastly overconfident while the high skilled are more accurate in assessing their skill. However, measurement error alone can lead to a negative relationship between performance and overestimation, even if skill and overconfidence are unrelated. To clarify the role of measurement error, we restate the Dunning–Kruger effect in terms of skill and overconfidence. We show that we can correct for bias caused by measurement error with an instrumental variable approach that uses a second performance as instrument. We then estimate the Dunning–Kruger effect in the context of the exam grade predictions of economics students, using their grade point average as an instrument for their exam grade. Our results show that the unskilled are more overconfident than the skilled. However, as we predict in our methodological discussion, this relationship is significantly weaker than ordinary least squares estimates suggest.

1. Introduction

Kruger and Dunning (1999) argue that the low skilled are typically vastly overconfident while the high skilled assess their skill more accurately. As evidence for this argument, they present an empirical pattern that is now known as the Dunning–Kruger effect: For many different tasks, low performers typically vastly overestimate their performance while high performers more accurately assess their performance (Dunning, 2011).

This evidence, however, is not sufficient, because measurement error alone can cause low performers to overestimate their performance more than high performers. To understand why, we first need to distinguish between skill and overconfidence and their measures performance and overestimation. Performance is the score on a test and overestimation is the difference between the expected and the actual test score. Performance measures skill with some error. We define skill as the ability to perform well on a given test and we can think of measurement error as luck on this test. Overestimation measures overconfidence, the difference between self-assessed and actual skill.[1]

The source of the bias is that researchers typically use the same performance to measure skill and to calculate overestimation. The same measurement error component is therefore part of performance and overestimation. To see how this can contribute to the Dunning–Kruger effect, consider a person with bad luck on a test: Bad luck decreases performance and increases overestimation and thus makes the person *appear* less skilled and more overconfident.

In the methodological part of this paper, we discuss the role of measurement error in the estimation of the Dunning–Kruger effect. We first restate the effect in terms of skill and overconfidence instead of their measures. We then show how measurement error causes an overestimation of the Dunning–Kruger effect and the assumptions under which we can correct for this bias with an instrumental variable (IV) approach.

* Corresponding author.
E-mail addresses: jan.feld@vuw.ac.nz (J. Feld), jan.sauermann@sofi.su.se (J. Sauermann), a.degrip@maastrichtuniversity.nl (A. de Grip).

[1] In their survey of the overconfidence literature, Moore and Healy (2008) define overestimation as overestimation of one's actual performance, overplacement as overestimation of one's performance relative to others, and overprecision as excessive precision in one's beliefs. While these definitions are helpful in distinguishing between the different domains of overconfidence, they are all defined in terms of actual outcomes, which may be affected by measurement error. We thus follow Moore and Healy's definition of overestimation and additionally define overconfidence as the overestimation of one's skill.

http://dx.doi.org/10.1016/j.socec.2017.03.002
2214-8043/

In the empirical part of this paper, we estimate the Dunning–Kruger effect with a sample of economics students who we asked four weeks before the exam to predict their exam performance. In line with the previous literature, we find that students who performed poorly on the exam also vastly overestimated their exam performance. We then estimate the Dunning–Kruger effect with an IV approach, using students' first-year grade point average (GPA) as an instrument for their exam performance. Our results confirm that the effect exists: The low skilled are vastly overconfident and the high skilled are more accurate in assessing their skill. As predicted by our methodological discussion, this effect is, however, significantly smaller than ordinary least squares (OLS) estimates suggest.

Krueger and Mueller (2002) are the first to have pointed out that measurement error can cause bias in the estimation of the Dunning–Kruger effect.[2] They correct for this bias by using two test performances: one to measure skill and one to calculate overestimation. They then regress overestimation calculated with the first performance on the second performance. The advantage of this approach is that it breaks the mechanical relationship between performance and overestimation, because the measurement error parts are now different for both variables. The disadvantage of this approach, however, is that the measurement error of the second performance (the independent variable) may bias the estimates toward zero. Low test–retest correlations of the test performances used by Krueger and Mueller (2002) suggest this measurement error is substantial (the test–retest correlation is 0.17 for their difficult test and 0.56 for their easy test). This could be why Krueger and Mueller do not find evidence of the Dunning–Kruger effect.

In response to Krueger and Mueller (2002), Ehrlinger et al. (2008) estimate the Dunning–Kruger effect using reliability-adjusted OLS. They regress overestimation on performance and then divide the estimated performance coefficient by a measure of the test's reliability. They thus present evidence of the Dunning–Kruger effect. Their approach is, however, problematic, because they still use the same performance as a measure of skill and to calculate overestimation. The performance coefficient of this regression is therefore likely biased and adjusting for test reliability only increases this bias.[3]

2. Dunning–Kruger effect

The setup of Dunning–Kruger effect studies is straightforward. Participants take a test in a given domain (e.g., English grammar, understanding humor, gun safety knowledge) and guess their performance on this test either before or after the test. The main finding is that bottom quartile performers vastly overestimate their performance while top quartile performers more accurately assess their performance.[4] This finding has been widely replicated with different populations and for a number of different tasks (Ehrlinger et al., 2008; Ryvkin et al., 2012; Schlösser et al., 2013).

Kruger and Dunning (1999) interpret this finding as evidence of a negative relationship between skill and overconfidence as opposed to merely an empirical pattern, which can be seen from the title of their paper: "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments." They further argue that differences in metacognitive skills between the low and high skilled drive this relationship. The idea is that the skills necessary to perform well are also those necessary to evaluate one's performance accurately. The low skilled therefore perform badly and lack the metacognitive skills to realize it.[5] As evidence for this explanation, Dunning and Kruger show that a randomly assigned training can increase competence and decrease the overestimation of low performers.

3. Estimating the Dunning–Kruger effect

3.1. Key variables

We define skill broadly as the ability to perform well on a given test. Skill, however, is imperfectly measured by performance, which is partly determined by luck. We use the terms *luck* and *measurement error* interchangeably to refer to all factors besides skill that influence test performance. We therefore define performance p as the sum of skill s^* and a classical measurement error component ε (asterisks indicate unobserved variables and we omit individual subscripts to simplify notation):

$$p \equiv s^* + \varepsilon. \tag{1}$$

Classical measurement error means that ε is a random error term with a mean of zero and independent of all variables included in the regression and the error term.

We define overconfidence as the difference between self-assessed skill and actual skill, that is, $oc^* \equiv s^*_{self-assessed} - s^*$. Overconfidence, however, is imperfectly measured by overestimation, that is, the difference between expected and actual performance:

$$oe \equiv p_{exp} - p. \tag{2}$$

We further assume that people state their self-assessed skill when asked about their expected performance p_{exp}. Expected performance is therefore the sum of a person's actual skill and overconfidence:[6]

$$p_{exp} \equiv s^* + oc^* = s^*_{self-assessed} \tag{3}$$

When we decompose overestimation into its respective elements, shown in Eqs. (1) and (3), we can see that it is equal to overconfidence minus measurement error:

$$oe = (s^* + oc^*) - (s^* + \varepsilon),$$
$$oe = oc^* - \varepsilon. \tag{4}$$

We can see from Eqs. (1) and (4) that the same measurement error component is part of performance and overestimation. This

[2] Krueger and Mueller (2002) argue that the Dunning–Kruger effect may be a statistical artifact caused by regression effects and the better-than-average effect. Their argument is that regression effects would lead to equal overestimation for low performers and underestimation for high performers. However, the fact that people are generally overconfident leads to an increase in the overestimation of the low performers and a decrease in the underestimation of the high performers. These two forces together therefore lead to the high overestimation of the low performers and the accurate performance assessment of the high performers.

[3] See Feld (2014) for a more extensive discussion on the biases of other estimation methods.

[4] When using relative performance measures, high-performing individuals typically slightly underestimate their performance. Kruger and Dunning (1999) explain this with the false consensus effect (Ross, Greene, & House, 1977), which states that people tend to overestimate the degree to which people are similar to them. The high-skilled overestimate the performance of others and therefore slightly underestimate their relative performance.

[5] Krajč and Ortmann (2008) propose an alternative explanation for the Dunning–Kruger effect. They observe that many of the studies showing the Dunning–Kruger effect use students from very selective institutions and argue that the students' skills in these samples follow a J-distribution equivalent to the upper tail of a normal distribution. The authors then show that the Dunning–Kruger pattern can arise even if people make random judgment errors, due to the J-distribution of skills and floor and ceiling effects caused by the test scale. In response, Schlösser et al. (2009) argue that, because student admission is based on many criteria, even in very selective institutions, skill is likely close to normally distributed. They then show that, even in the rare cases where skill follows a J-distribution, the Krajč–Ortmann explanation would only account for a small fraction of the observed Dunning–Kruger effect.

[6] Besides expected skill, a number of other factors might influence a person's expected performance. When expected performance is elicited before the test, as in this paper, these other factors are arguably unrelated to skill and measurement error and therefore do not affect Dunning–Kruger effect estimates.