Journal of Economic Psychology 59 (2017) 59–86

Contents lists available at ScienceDirect

Journal of Economic Psychology

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

Estimating risky behavior with multiple-item risk measures

Lukas Menkhoff [a,b,c,*], Sahra Sakha [d]

[a] Humboldt University Berlin, Mohrenstrasse 58, 10117 Berlin, Germany
[b] German Institute for Economic Research (DIW Berlin), Mohrenstrasse 58, 10117 Berlin, Germany
[c] Leibniz University Hannover, Germany
[d] Research Center, Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 27 July 2016
Received in revised form 3 February 2017
Accepted 3 February 2017
Available online 10 February 2017

JEL classifications:
D81
C93
O12

PsycINFO classifications:
2229
2260

Keywords:
Risk
Experiments
Household survey
Testing methods
Contexts

ABSTRACT

We compare seven established risk elicitation methods and investigate how robustly they explain eleven kinds of risky behavior with 760 individuals. Risk measures are positively correlated; however, their performance in explaining behavior is heterogeneous and, therefore, difficult to assess ex ante. Greater diversification across risk measures is conducive to closing this knowledge gap. What we find is that performance increases considerably if we combine single-item risk measures to form multiple-item risk measures. Results are improved the more single-item measures they contain, and also if these single-item risk measures use different elicitation methods. Interestingly, survey items perform just as well as incentivized experimental items in explaining risky behavior.

1. Introduction

Analyzing decision making under risk requires a reliable measure of individual risk attitude. Most studies to date have opted for an existing method of eliciting individual risk preferences and take the response to this specific item as "the" individual risk attitude. However, risk attitudes across these measures are often inconsistent and the predictive power of the different methods used is typically low.

Therefore, our study uses seven well-established risk measures with 760 individuals and examines the ability of these measures to explain eleven kinds of risky behavior. When examining explanatory power, we follow Dohmen et al. (2011) and control for a set of individual socio-demographic characteristics (see also, e.g., Barsky, Juster, Kimball, & Shapiro, 1997; Tanaka, Camerer, & Nguyen, 2010). The results obtained demonstrate the limitations of narrow approaches, where isolated relationships between risk measure and behavior may depend on the specific risk measure chosen. While all

* Corresponding author at: German Institute for Economic Research (DIW Berlin), Mohrenstrasse 58, 10117 Berlin, Germany.
E-mail addresses: lmenkhoff@diw.de (L. Menkhoff), sahra.sakha@bundesbank.de (S. Sakha).

http://dx.doi.org/10.1016/j.joep.2017.02.005
0167-4870/

single-item risk measures are able, to a certain extent, to explain risky behavior in our sample, the level of heterogeneity is considerable. While some measures perform better than others, it is unfortunately unclear which risk measure to choose ex ante. As a consequence of our analyses, we propose to diversify across risk measures.

Our main objective is to show that diversification across risk measures improves their ability to predict risk-related behavior. Averaging across single-item risk measures, i.e. creating a "multiple-item risk measure", substantially improves the predictive power of explaining behavior. While the advantage of diversification is known from findings of the forecasting literature (Timmermann, 2006, chap. 4), it questions the idea that there are various risk measures available which are all equally suited to elicit risk attitude. Our evidence, however, suggests that noise in single-item risk measures can be reduced by relying on multiple-item risk measures.

We employ seven well-established methods to elicit risk preferences using incentivized risk tasks: the certainty equivalent task (CE) (see e.g., Abdellaoui, Baillon, Placido, & Wakker, 2011), two choice set tasks (CSL and CS) following Eckel and Grossman (2002, 2008), and an investment choice task (CI) similar to Gneezy and Potters (1997). Alongside these four experimental methodologies (where abbreviations start with "C"), we use three non-incentivized survey items of risk attitude (where abbreviations start with "S"). Two of these stem from the study of Dohmen et al. (2011), i.e. general willingness to take risk (SG) and the willingness to take risk in financial affairs (SF). In addition, we also employ a hypothetical investment question (SI) which has been used by Barsky et al. (1997). We investigate the correlation between these seven risk elicitation tasks at the individual level. Measures of risk attitude are positively correlated, yet in most cases only to a low degree. This indicates remarkable differences between measures, conforming to results found in the literature. Deck, Lee, Reyes, and Rosen (2013) test for domain-specific risk attitudes using multiple risk tasks in a within-subject design (including versions of the choice sets and choice lists) to find that tasks are poorly correlated across elicitation methods. Similarly, Crosetto and Filippin (2016) compare a battery of incentivized and non-incentivized tasks to elicit risk attitudes and find that the estimated risk aversion parameters from the tasks vary greatly. Findings of low consistency between tasks are also found in developing countries. Nielsen, Keil, and Zeller (2013) examine the consistency of risk preferences based on eight hypothetical elicitation methods and a lottery game to smallholder farmers in Vietnam and find - similar to our results – statistically significant but weak correlations between tasks.

In the next step, we average across risk measures which we standardize for this purpose. We find that a simple average across the seven measures has the highest predictive power in our sample as it significantly explains 6 out of 11 kinds of risky behavior, whereas the single-item measures explain on average 2.6 kinds of behavior, ranging from 1 to 4. However, from a practical point of view (i.e. implementing this in the field), it seems rather expensive to collect and combine seven risk measures to form multiple-item risk measures. Hence, we investigate whether predictive ability can be maintained by combining any two risk items. We find that these combinations, indeed, are able to explain more types of risky behavior (on average 3.1) than merely employing a single-item measure (on average 2.6). These results also hold qualitatively if we exclude the willingness to take risk in financial affairs since this is a domain-specific risk measure.

Since predictive power still varies considerably between any two-item risk measures, we are trying to identify general principles for building successful but still relatively simple multiple-item risk measures. We find that including risk items with different elicitation methods make the multiple-item risk measure more reliable and predictive. By contrast, combining items from repeated answers seems less conductive to improving predictive power. Opting for incentivized experiments over easily implementable survey items does not appear to be effective, either. We are aware that our results for specific risk measures might be the consequence of our selection of risk measures, risky behavior, and sample population. We, therefore, hesitate to recommend the inclusion of any specific risk item. However, we do find that multiple-item risk measures outperform single-item risk measures. This leads to the concrete conclusion that researchers attempting to identify risk attitude should consider using two (or better yet: three) risk items with different risk elicitation methods to enhance external validity.

Our research relates to at least three strands of literature: (1) the wealth of studies which examine the within-sample consistency between various risk elicitation methods (e.g., Crosetto & Filippin, 2016; Deck et al., 2013; He, Veronesi, & Engel, 2016; Isaac & James, 2000; Loomes & Pogrebna, 2014). They find degrees of inconsistencies that are difficult to explain within any commonly used model of decision making under risk. (2) Studies assessing the validity of risk measures by predicting risky behavior (e.g., Barsky et al., 1997; Dohmen et al., 2011; Sutter, Kocher, Glätzle-Rützler, & Trautmann, 2013; Vieider et al., 2015). (3) Risk elicitation methods implemented in rural Thailand have been used in our earlier work and were part of a larger household survey: Hardeweg, Menkhoff, and Waibel (2013) replicate the study of Dohmen et al. (2011) and, therefore, use three non-incentivized risk items and two kinds of behavioral outcomes based on data from the second wave of the household survey in 2008. Gloede, Menkhoff, and Waibel (2015) use the "general willingness to take risk item" and the certainty equivalent (CE) task from the third wave of the household survey in 2010 to examine whether the experience of shocks influences individual risk attitude. In a later study, Menkhoff and Sakha (2016) examine changes in risk attitude over time using the certainty equivalent (CE) task. The seven risk items used in this study have been elicited in a separate survey implemented in 2013 and are exclusively analyzed in the current research.

Our paper is organized in seven sections. Section 2 presents the survey data and risk elicitation methods. Section 3 displays the descriptive statistics of our sample and outlines the experiments with correlations between risk measures. Section 4 shows the results on the predictive ability of single-item risk measures. Section 5 outlines the performance of the various multiple-item risk measures. Section 6 introduces various robustness checks and Section 7 concludes.