Economics Letters 242 (2024) 111851

Contents lists available at ScienceDirect

Economics Letters

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

Revalidating a survey instrument for measuring risk preferences☆

Anika Bittner [a], Alexander Haering [b], Timo Heinrich [a,*], Thomas Mayrhofer [c]

[a] Hamburg University of Technology, Am Schwarzenberg-Campus 2, 21073 Hamburg, Germany
[b] RWI – Leibniz Institute for Economic Research and Fresenius University of Applied Sciences, Hohenzollernstraße 1-3, 45128 Essen, Germany
[c] Stralsund University of Applied Sciences and Harvard University, Zur Schwedenschanze 15, 18435 Stralsund, Germany

ARTICLE INFO

JEL Classification:
C81
C91
D81

Keywords:
Survey validation
Experiment
Risk preferences

ABSTRACT

Survey instruments are often a cost-effective alternative to eliciting preferences via incentivized experiments. We revalidate a survey instrument used for measuring risk preferences in the influential Global Preferences Survey. We find that the instrument consisting of a quantitative and a qualitative item predicts risk aversion in a Chinese, an American and a German sample. For out-of-sample predictions, the combination of both items is preferable compared to using only one of the two items.

1. Introduction

Understanding risk preferences is central to analyzing human decision making across various contexts. According to economic theory, the choices of risk-averse people will often differ fundamentally from those who are risk-neutral or risk-loving. In standard economic models, risk-averse people purchase insurance, opt for fixed wage contracts or diversify their investments.

Accordingly, a survey instrument for risk preferences is also included in the influential Global Preferences Survey (GPS) by Falk et al. (2018). The survey measures preferences in representative samples across 76 countries comprising 80,000 respondents. It reveals substantial heterogeneity in preferences across and within countries. For example, risk aversion is relatively high in Western Europe and relatively low in the Middle East and in Africa. Yet, interpreting these differences in survey responses as differences in underlying preferences requires that survey responses are linked to real preferences across the respective sub-samples.

Falk et al. (2023) present the experimental validation of the survey instruments used in the GPS. As they point out, a large body of evidence suggests that their validation, which is based on a German sample, will be sufficient to identify instruments that reliably predict preferences of the wider population. Nevertheless, the authors "…*think that running our validation exercise using different samples would provide valuable results on the usefulness of different preference measures, for example, in other countries or in specific subgroups of populations* […]." (Falk et al., 2023, p. 1943).

We revalidate survey instrument for risk preferences focusing on risk aversion *per se* (and not its strength). Falk et al. (2023) identify the combination of one quantitative item and one qualitative item as the preferred survey instrument. For the GPS they use a staircase measure as the quantitative item and a question about general risk attitude (similar to the question validated by Dohmen et al., 2011) as the qualitative item. Employing a Chinese, an American and a German sample we find that across samples the two items differ in their correlation with experimentally elicited risk aversion. However, the survey instrument consisting of both items predicts risk aversion within all three samples. When using the data by Falk et al. (2023) for out-of-sample predictions, the combination of both items yields the best predictions.

2. Data sets

The first data set, also analyzed by Haering et al. (2020), includes incentivized lottery choices to elicit higher-order risk preferences as

☆ We thank an anonymous reviewer and audiences at the meeting of the German Society for Experimental Economics (2023) and the "Behavioral Economics: Behind the Scenes" (2023) course in Copenhagen for helpful comments. We also thank Christine E. Looser, Haiying Qin, Weijing Le, Chaoliang Yang and Guanzhong Yang for their support in organizing the experiments. Franziska Then and Stephanie Zimmer provided excellent research assistance. Financial support from the BMBF through the IN-EAST School of Advanced Studies (Grant No. 01UC1304) is gratefully acknowledged.

* Corresponding author.
E-mail address: timo.heinrich@tuhh.de (T. Heinrich).

https://doi.org/10.1016/j.econlet.2024.111851
Received 6 June 2024; Received in revised form 1 July 2024; Accepted 2 July 2024

Table 1
Summary statistics.

	Data source	Country	Staircase risk	General risk	Risk averse	*N*
CHN	Haering et al. (2020)	China	11.17	5.64	0.53	140
			(7.40)	(2.01)	(0.50)	
USA	Haering et al. (2020)	USA	10.14	5.54	0.63	129
			(6.08)	(2.06)	(0.49)	
GER	Haering et al. (2020)	Germany	11.27	5.01	0.58	145
			(5.94)	(2.02)	(0.50)	
FBDHS	Falk et al. (2023)	Germany	11.13	4.00	0.45	394
			(6.73)	(2.08)	(0.50)	

The table reports the means and in parentheses the standard deviations. The staircase risk column reports the switching row ranging from 1 to 32. The general risk column reports the answer on the scale from 0 to 10.

developed by Deck and Schlesinger (2014). The lotteries were preceded by a test of understanding and followed by a post-experimental survey that included the staircase measure and the general risk question as used in the GPS.[1] The participants made 38 choices between lottery pairs in the gain domain, one of which was randomly selected for payment.[2] This study will focus only on the seven lottery pairs that allow us to determine whether a participant is (second-order) risk-averse. In each of the lottery pairs, one option is characterized by a lower variance and is thus preferred by risk-averse participants. The experiments were conducted with student samples in economic laboratories at Nankai University in Tianjin (China [CHN]), at CLER at Harvard Business School in Boston (USA) and at the elfe laboratory at the University of Duisburg-Essen in Essen (Germany [GER]).[3]

As the second data set, we consider the German student sample used by Falk et al. (2023) for validation, denoted by FBDHS in the following. They select the items for their survey module by correlating the answers to candidate survey items with risk preferences elicited through two multiple price lists similar to those used by Holt and Laury (2002). They identify the combination of a non-incentivized multiple price list and the general risk question as the preferable survey instrument. However, due to the time constraints of their survey, they measure risk preferences using the staircase measure which is an abbreviated version of the multiple price list, allowing for a more fine-grained assessment of risk preferences while reducing the number of choices from 21 to five.

3. Results

Table 1 presents the summary statistics. It includes responses to the two survey items of interest as well as the classification of being risk-averse or not. In our CHN, USA and GER samples, we classify everyone as risk-averse who makes the risk-averse choice in at least six out of the seven lottery pairs.[4]

Additionally, Table 1 includes the respective statistics for the FBDHS sample. In their sample, the strength of risk preferences is measured using price lists with 21 potential switching rows, where a higher switching row indicates higher risk tolerance. A risk-neutral individual would be indifferent between switching in rows 11 and 12. Accordingly, we classify those with average switching rows below 11 as risk averse.

In a first step, we focus on the within-sample correlations of our two survey items with experimentally elicited preferences. We run logit regressions on being classified as risk-averse with the answers to the two survey items as independent variables. Table 2 presents the estimated average marginal effects. Not surprisingly, in FBDHS the two items correlate with our classification of being risk averse or not.

More importantly, we find significant correlations also in the three other samples: An additional switching row (out of 32) in the staircase risk measure decreases the likelihood of being risk averse by 1.2 percentage points in CHN and by 2.6 percentage points in USA. The relationship is insignificant in GER. A one-point increase (out of 11) when ratings one self`s willingness to take risks decreases the likelihood by 8.8 percentage points in CHN and by 6.9 percentages points in GER. The relationship is insignificant in the USA. As the model holds predictive power in all three samples, we conclude that the survey instrument based on the combination of a quantitative and a qualitative item predicts risk aversion across samples.

In a second step, we consider out-of-sample predictions based on FBDHS. Interpreting differences in survey responses across subsamples as differences in underlying preferences requires that the respective survey instrument can predict preferences out-of-sample. Thus, similar to the procedure employed by Falk et al. (2018) and Falk et al. (2023), we estimate three different logit models without intercepts using normalized independent variables on the FBDHS data set. The three models include either only one of the two items (“staircase risk” or “general risk”) or both (“combined”). We use the estimated coefficients to predict risk aversion in CHN, USA and GER and contrast the estimated likelihood of being risk averse with elicited preferences.

Table 3 presents the resulting mean squared prediction errors. It shows that the preferred model (highlighted in bold) is different in each of the three samples. Across samples, however, the combination of a quantitative and a qualitative item, as advocated by Falk et al. (2023), performs best. Note that when using a more relaxed definition of risk aversion (only five out of seven choices have to be risk averse), the combined model not only predicts best in the CHN and the joint sample but also in the GER sample (see Appendix E).[5]

4. Conclusion

We present a revalidation of the survey instrument for risk preferences used in the influential Global Preferences Survey. We find that the two items forming the survey instrument differ in their correlation with risk aversion across three samples. We provide supporting evidence for the choice by Falk et al. (2023) of combining one qualitative and one quantitative item to construct their survey instrument. In the overall out-of-sample prediction, it outperforms survey instruments relying on only one of the two items.

However, as already indicated by previous studies, the accuracy of

[1] For the exact wordings see Appendix A. Due to a coding error we cannot distinguish between switching rows 9 to 13 in the staircase risk measure, all of which imply risk-averse behavior. Thus, we impute the corresponding choices with the midpoint value of 11.

[2] See Appendix B for the complete set of lotteries, Appendix C for the instructions and Appendix D for screenshots.

[3] See Haering et al. (2020) for details on the procedures which aimed to ensure comparability across data sets.

[4] In Appendix E we present robustness checks for a stricter and a more relaxed classification.

[5] Comparing ROC (receiver operating characteristics) curves as shown in Appendix F confirm these findings. A visual comparison between within-sample and out-of-sample predictions is provided in Appendix G.

Table 2
Regressions on being classified as risk averse.

	(1) CHN	(2) USA	(3) GER	(4) FBDHS
Staircase risk	−0.012**	−0.026***	−0.004	−0.014***
	[−0.022; −0.002]	[−0.039; −0.012]	[−0.019; 0.011]	[−0.022; −0.007]
	(0.005)	(0.007)	(0.007)	(0.004)
General risk	−0.088***	0.008	−0.069***	−0.055***
	[−0.122; −0.053]	[−0.034; 0.050]	[−0.108; −0.031]	[−0.078; −0.031]
	(0.018)	(0.022)	(0.019)	(0.012)
Pseudo R^2	0.168	0.080	0.075	0.097
Observations	140	129	145	394

Dependent variable: classified as being risk averse. Average marginal effects, 95% confidence intervals in square brackets, robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3
Out-of-sample mean squared prediction errors (minimum in bold).

Model	CHN	USA	GER	All
Staircase risk	0.226	**0.225**	0.245	0.232
General risk	0.209	0.263	**0.226**	0.232
Combined	**0.201**	0.242	0.228	**0.223**

the survey instrument varies between populations. Bauer et al. (2020) validate a battery of items similar to those used in the GPS with low-income individuals in Kenya. As in our American sample, they find that the staircase risk measure is significantly correlated with experimentally elicited preferences while the general risk question is not. The general risk question by itself, however, has been previously found to correlate with experimentally elicited risk preferences in Chinese students by Ding et al. (2010) and in individuals in rural Thailand by Hardeweg et al. (2013). Vieider et al. (2015) measure preferences with respect to uncertainty and risk in lab experiments with students in 30 countries. With respect to monetary gains, they find that the general risk question correlates significantly with risk preferences for 19 of these countries. Regarding the USA, Germany and China, we confirm their findings: Vieider et al. (2015) observe significant correlations between elicited risk preferences and the general risk questions for their Chinese and German samples, but not their American sample.

Based on our results and findings from previous studies, we conclude that for out-of-sample predictions, a survey instrument based on the combination of a quantitative and a qualitative risk preference item is preferable compared to using only one of the two items. Furthermore, it may be informative to consider the results of quantitative and a qualitative items separately.

Data availability

Replication files are available at https://osf.io/3a7td/.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.econlet.2024.111851.

References

Bauer, M., Chytilová, J., Miguel, E., 2020. Using survey questions to measure preferences: lessons from an experimental validation in Kenya. Eur. Econ. Rev. 127, 103493.
Deck, C., Schlesinger, H., 2014. Consistency of higher order risk preferences. Econometrica 82 (5), 1913–1943.
Ding, Hartog, J., Sun, Y, 2010. Can We Measure Individual Risk Attitudes in a Survey? Working paper.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: measurement, determinants, and behavioral consequences. J. Eur. Econ. Assoc. 9 (3), 522–550.
Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., Sunde, U., 2018. Global evidence on economic preferences. Quarterly J. Econ. 133 (4), 1645–1692.
Falk, A., Becker, A., Dohmen, T., Huffman, D., Sunde, U., 2023. The preference survey module: a validated instrument for measuring risk, time, and social preferences. Manage Sci. 69 (4), 1935–1950.
Haering, A., Heinrich, T., Mayrhofer, T., 2020. Exploring the consistency of higher order risk preferences. Int. Econ. Rev. 61 (1), 283–320.
Hardeweg, B., Menkhoff, L., Waibel, H., 2013. Experimentally validated survey evidence on individual risk attitudes in rural Thailand. Econ. Dev. Cult. Change 61 (4), 859–888.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review 92 (5), 1644–1655.
Vieider, F.M., Lefebvre, M., Bouchouicha, R., Chmura, T., Hakimov, R., Krawczyk, M., Martinsson, P., 2015. Common components of risk and uncertainty attitudes across contexts and domains: evidence from 30 countries. J. Eur. Econ. Assoc. 13 (3), 421–452.