CrossMark

The psychometric and empirical properties of measures of risk preferences

Jonathan P. Beauchamp[1] · David Cesarini[2,3] · Magnus Johannesson[4]

Abstract We examine the psychometric and empirical properties of some commonly used survey-based measures of risk preferences in a population-based sample of 11,000 twins. Using a model that provides a general framework for making inferences about the component of measured risk attitudes that is not due to measurement error, we show that measurement-error adjustment leads to substantially larger estimates of the predictive power of risk attitudes, of the size of the gender gap, and of the magnitude of the sibling correlation. Risk attitudes are predictive of investment decisions, entrepreneurship, and drinking and smoking behaviors; are robustly associated with cognitive ability and personality; and our estimates are often larger than those

✉ Jonathan P. Beauchamp
  jonathan.pierre.beauchamp@gmail.com

  David Cesarini
  david.cesarini@nyu.edu

✉ Magnus Johannesson
  Magnus.Johannesson@hhs.se

[1] Department of Economics, University of Toronto, 150 St. George Street, Toronto, ON M5S 3G7,
  Canada

[2] Department of Economics and Center for Experimental Social Science, New York University,
  19 W. 4th Street, New York, NY 10012, USA

[3] Research Institute for Industrial Economics (IFN), Box 55665, 102 15 Stockholm, Sweden

[4] Department of Economics, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm,
  Sweden

in the literature. Our results highlight the importance of adjusting for measurement error across a wide range of empirical settings.

1 Introduction

Preference heterogeneity is a possible explanation for some of the individual-level variation observed in economic behaviors, such as labor supply, saving and consumption decisions, and asset allocation. The fundamental difficulty that arises in testing explanations of individual differences that invoke preference heterogeneity is that preferences are never directly observed. Stigler and Becker (1977) famously argued that economists should assume not only that individual tastes are stable over time, but that tastes are identical across persons. Those who favor this position argue that the problem with preference-based explanations is that defending them is difficult without recourse to tautological, and hence scientifically meaningless, arguments.

In recent years, an alternative methodological approach has gained traction in experimental economics and increasingly also in applied empirical economics. According to this view, researchers should try to obtain empirical measures of the fundamental dimensions of heterogeneity using surveys or experiments. Such direct measurement of preferences, sometimes coupled with the assumption that they are stable functions of some observable states of nature, is a way of disciplining preference-based explanations and thus avoiding the problem of ad hoc theorizing that concerned Stigler and Becker. Proponents of this approach argue for the integration of individual-difference psychology into economics (Almlund et al. 2011; Becker et al. 2012; Ferguson et al. 2011; Borghans et al. 2008) and for a sustained effort to learn more about the properties of the measures of preferences that are commonly used in economic research.

In this paper, we make several contributions to this effort. First, we examine the consequences of measurement error (or other transitory fluctuations) in some frequently used survey-based measures of risk attitudes and show that accounting for them leads to considerably higher estimates of their predictive power and correlations with other variables. To do so, we adopt a uniform latent variable model that provides a general framework for making inferences about the component of measured risk attitudes that is not due to measurement error. Economists have been aware for some time that measurement error may significantly attenuate the relationship with other variables and lead to mistaken inference (c.f., Solon, 1992). In practice, however, much of the work that uses surveys or experimental tasks to elicit preferences – including work that questions the usefulness of measures of economic preferences for predicting economic outcomes – treats behavioral responses as if they yielded error-free measurements of the underlying preferences.

Second, we conduct a detailed examination of the psychometric and empirical properties of some commonly-used measures of risk attitudes using a large, population-based sample of respondents. We document sizable associations between risk attitudes and a host of real-world outcomes and variables, including investment decisions, entrepreneurship, smoking, drinking, gender, cognitive ability, and personality. Our data set is a comprehensive survey of more than 11,000 Swedish twins with rich self-reported data on psychological variables and risky health behaviors, matched to administrative records with information on labor supply and financial portfolio risk. Important for our purposes is that the survey contains five measures of risk attitudes as well as retest data for 500 twins. The first measure is close in spirit to the one developed for the Health and Retirement Survey; it asks people to answer a series of sub-questions about hypothetical gambles over lifetime wealth (Barsky et al. 1997). The next two measures ask, respectively, about risk attitudes in the domain of finance and risk attitudes in general; these measures have been studied by Dohmen et al. (2011), who provide some evidence on their predictive validity. Our final two measures ask about attitudes toward hypothetical gambles over gains and losses.

Our third contribution is to introduce and explore concepts and tools from psychometrics – the field of study concerned with the theory and techniques of psychological measurement – and argue that economists have much to learn from that field in their efforts to understand the properties of the measures of economic preferences. Psychometricians who study personality and cognition make a useful conceptual distinction between reliability, construct validity, and predictive validity. Reliability refers to the consistency of individuals' responses to an instrument across measurement occasions and is a descriptive statistic designed to capture how much measurement error is in a variable; construct validity refers to the degree to which the instrument actually measures the underlying construct it is intended to measure; and predictive validity is the extent to which it correlates with, or predicts, other variables that theory or intuition suggest might relate to the construct purportedly being measured (McArdle and Prescott 1992).

Many of the measures of preferences that have come from economics were designed to directly measure the fundamental dimensions of heterogeneity that feature in economic theory, such as a coefficient of risk aversion. In the language of psychometrics, such instruments have, by their very design, strong construct validity. In contrast, relatively little effort has been devoted to studying the reliabilities and predictive validities of the measures used in economics. The research that does address or touch upon these questions suggests the measures are subject to significant measurement error (Barsky et al. 1997; Gillen et al. 2015; Harrison et al. 2005; Sahm 2012; Kimball et al. 2008; Lönnqvist et al. 2014) and have only limited predictive validity.

The remainder of the paper is structured as follows. We begin by providing an overview of the data set in Section 2. In Section 3, we introduce the uniform and general latent variable model of risk attitudes that we adopt throughout the paper and that accounts for measurement error (or other transitory fluctuations) and for the ordinal nature of the measures of risk attitudes. Then, in Section 4, we estimate the test-retest reliability of the five measures, using data from approximately 500 respondents who responded to the survey on two occasions. Our estimates – which are measures of the stability of risk attitudes over time – vary between about 0.5 to

0.7 for the different measures, showing some stability in responses over time but also substantial measurement error.

Section 5 investigates the predictive validity of the risk measures (focusing on the Risk General, Risk Financial, and Risk HRS variables, which have higher reliability). To do so, we use the GMM estimator proposed by Kimball et al. (2008), which allows consistent estimation of the effect of each individual measure of risk preferences despite the fact that the measures are noisy. We report the first measurement-error-adjusted estimates of the proportion of variation in risky behaviors in the domains of health and personal finance that is explained by the measures of risk attitudes. Our measures of risk attitudes have significant explanatory power for investment decisions, the propensity to run one's own business, and drinking and smoking behaviors. For example, adding risk attitudes to a rich set of covariates in a regression in which portfolio risk is the dependent variable doubles the R^2, though from a low baseline. As well, after adjustment for measurement error, a one-standard-deviation increase in risk attitudes is associated with an increase of approximately 10 percentage points in the probability of having started a business, as well as with two- to four-percentage-point increases in the probability of being an alcohol consumer and in the probability of being a smoker. We thus find strong support for the proposition that persistent differences are present in risk attitudes across people and that these differences translate into statistically significant and economically important differences in economic choices. Also, we find that adjusting for measurement error substantially increases the estimated effects. An important conclusion emerging from our work is that the low R^2's reported in previous work (Barsky et al. 1997; Harrison et al. 2005; Dohmen et al. 2011; Dohmen et al. 2012; Kimball et al. 2008; Sahm 2012) – which are sometimes used to discard preference-based explanations – are partly attributable to the relatively low reliabilities of the risk measures.

Section 6 examines predictors of the measures of risk attitudes and how the coefficient estimates change once we allow for measurement error in these measures. We document large sex differences in risk attitudes after measurement-error correction and find that cognitive ability – measured about four decades before the risk attitudes – is a strong predictor of risk attitudes. In Section 7, we conduct a factor analysis that shows that all five risk variables load significantly on their first common factor, indicating they share a sizable fraction of their variance, although an important part of the variation is variable-specific.

In Section 8, we consider other applications of our framework. We first estimate a sizable and highly significant measurement-error-adjusted correlation between risk attitudes and the personality trait behavioral inhibition. The estimated correlation between behavioral inhibition and risk attitudes increases by about 50% after adjustment for measurement error, and the adjusted estimate of 0.45 is substantially higher than previously reported correlations between risk attitudes and personality traits (Becker et al. 2012; Dohmen et al. 2010; Lönnqvist et al. 2014). We then provide another illustration of the importance of adjusting for measurement error, by reporting rough estimates of the share of the variation of risk attitudes that is attributable to genetic factors. With adjustment for measurement error, the estimates range from 35% to 55% – almost doubling previous estimates in the literature (Cesarini et al.

2009) – compared to 21% to 34% without adjustment for measurement error. We conclude in Section 9.

2 Data

Our data come from four separate sources: the Swedish Twin Registry (STR), Statistics Sweden, the National Insurance Board, and the National Service Administration. We begin by describing the STR-administered SALTY survey, from which we draw our measures of risk preferences (among others). We then provide a brief overview of the variables we use in this paper and present summary statistics for our sample; we provide additional details on the variables and data in the Online Appendix.

2.1 SALTY survey

The STR routinely administers surveys to Swedish twins. We use data from a survey known as SALTY (Screening Across the Lifespan Twin Study: the Younger), which was a collaborative effort between researchers in epidemiology, medicine, and economics, initiated in 2008 and completed in the winter of 2010 (Lichtenstein et al. 2006; Magnusson et al. 2013). SALTY is the first major survey of twins that features entire sections specifically devoted to economic preferences and behaviors. The sampling frame for the SALTY survey was determined using information from an earlier STR survey known as SALT (the Screening Across the Lifespan Twin Study), a phone-based survey in which all Swedish twins born between 1926 and 1958 were invited to participate. The SALT survey attained a response rate of 74%. The SALTY survey was administered to all twin pairs born between 1943 and 1958 except those pairs in which neither twin had elected to participate in SALT.

SALTY was sent to a total of 24,914 Swedish twins and generated a total of 11,743 responses, a response rate of 47.1%; of these respondents, 11,418 (97.2%) gave informed consent to have their responses stored and analyzed. In total, our sample comprises 1,150 monozygotic (MZ) pairs, 1,245 same-sex dizygotic (DZ) pairs, and 1,117 opposite-sex DZ pairs. (Remaining responses were from individuals whose twin siblings were non-respondents.)

We also administered the survey a second time to a subsample of respondents. We determined the subsample by randomly drawing 800 families in which at least one member had responded to the SALTY survey and given informed consent to have the responses stored and analyzed. From each drawn family, we then drew one individual who had previously responded to SALTY, and sent the survey to that individual. Unlike the first round of data collection, participants were promised lottery tickets worth approximately SEK 150 in exchange for their participation.[1] We obtained 500 responses to the second-round survey, of which 496 provided the informed consent necessary to analyze their responses.

---

[1]At the time, one US dollar was worth approximately seven SEK.

2.2 Measuring risk attitudes

To measure attitudes toward risk, we constructed five ordinal variables[2] using responses to the SALTY survey. We denote these variables Risk HRS, Risk General, Risk Financial, Risk Gain, and Risk Loss. All five variables were coded so that a higher ordinal category corresponds to greater risk tolerance.[3]

Risk general and risk financial Risk General is constructed from answers to the following question:

> How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please check a box on the scale, where the value 1 means: "unwilling to take risks" and the value 10 means: "fully prepared to take risks".

Risk Financial is constructed from a similarly phrased question, except the respondents are asked about their willingness to take risks specifically in the financial domain. These questions are available in several waves of the German Socioeconomic Panel. Dohmen et al. (2011, 2012) establish the predictive validity of the Risk General question and use it to study the transmission of risk attitudes from parent to child.[4]

Risk HRS We asked individuals to respond to a series of questions about a guaranteed monthly salary of SEK 25,000 for the rest of their lives or a gamble in which they had a 50-50 chance of earning either SEK 50,000 or SEK X for the rest of their lives. We assigned individuals to one of four ordinal categories based on their answers. This measure is in the spirit of the series of hypothetical gambles used in the Health and Retirement Survey. Barsky et al. (1997) used a similar question and demonstrated its predictive validity for behavior in a number of domains. For additional details, see Sahm (2012) and Kimball et al. (2008).

Risk gain and risk loss Risk Gain asks whether respondents prefer SEK 24,000 for sure or a 25% chance of winning SEK 100,000. Risk Loss asks respondents to choose between a guaranteed loss of SEK 24,000 and a 25% chance of losing SEK 100,000. We coded both variables separately as 1 for the respondents who preferred the gamble and as 0 for the respondents who preferred the sure amount. We adopted

---

[2]Two of these variables are binary.

[3]As we discuss below, our analyses focus on the permanent components of risk attitudes ρ^*, which we assume are continuously distributed latent variables that underlie the ordinal variables' discrete distributions; thus, we do not treat the ordinal risk-attitude variables as having cardinal significance in our analyses.

[4]A minor difference is that our scale ranges from 1 to 10, whereas the original question that Dohmen et al. (2011, 2012) used had 11 response categories, ranging from 0 to 10.

these two questions from the series of hypothetical gambles used by Tversky and Kahneman (1992) to estimate risk attitudes over gains and losses.[5]

2.3 Risky behaviors

Alcohol consumption and smoking Using data from the SALT survey, we constructed two binary variables on smoking and drinking habits. Specifically, we classified as alcohol consumers individuals who responded affirmatively to a question about whether they had drunk strong beer, wine, or liquor more than twice during the last month or who had indicated in a follow-up question that they usually drink strong beer, wine, or liquor at least twice a month. We classified as smokers individuals who indicated they smoke regularly, used to smoke regularly, smoke on and off, or used to smoke on and off.

Equity share and portfolio risk We include two measures of investment behavior. Equity Share is a measure of the share of equity in each respondent's stock of assets, using answers to questions about the value of assets in six different categories in the SALTY survey. These data are only available for a subset of the SALTY respondents, because the questions were removed from the survey in the later waves in an effort to reduce the number of questions.

Portfolio Risk is based on data from the Swedish individualized pension savings accounts introduced in 2000. Under this system, virtually all adult Swedes born after 1938 had to decide how to invest part of their retirement savings and construct an investment portfolio from a menu of several hundred funds. We obtained data from the National Insurance Board on how the individuals in our sample elected to invest their retirement wealth in the year 2000 when the reform was introduced. Our Portfolio Risk variable is the average risk level of the funds owned by an individual, with the risk of each fund measured as the (annualized) standard deviation of the fund's monthly rate of return over the previous years.[6] Cesarini et al. (2010) provide more details.

Own business We also obtained a measure of entrepreneurship from responses to the SALTY question "Have you ever run your own business?" We label this variable "Own Business."

2.4 Other variables

Birth weight Data for our birth weight variable come from delivery archives throughout Sweden (Lichtenstein et al. 2006) or, when archival data are missing,

---

[5]We used considerably higher hypothetical stakes than Tversky and Kahneman (1992), and we included a single binary choice question for each of the two gambles, whereas they used a series of binary choices to estimate the certainty equivalent of the gambles.

[6]Eligible Swedes were provided with information about the risk level of different funds prior to making their investment decisions; they were sent a catalog that color coded funds by risk level and reported the standard deviation of the fund's historical returns.

from a question in the SALT survey asking "What was your birth weight?" The birth weight variable is scaled in kilograms.

Cognitive ability We matched the men in the SALTY sample to conscription data provided by the Swedish National Service Administration. All men in our sample were required by law to participate in military conscription around the age of 18. They enlisted at a point in time when exemptions from military duty were rare: 95% of the male twins in our sample were successfully matched to the information in the military archives. As part of the drafting procedure, recruits had to complete a test of cognitive ability, which consisted of four subtests (logical, verbal, spatial, and technical). We transformed the respondents' test scores to a normally distributed z-score with mean zero and variance one, separately by birth year. Carlstedt (2000) discusses the history of psychometric testing in the Swedish military and provides evidence that the measure of cognitive ability is a good measure of general intelligence (Spearman 1904).

Rotter Locus of Control and behavioral inhibition SALTY respondents filled out two personality scales. To measure beliefs about personal control, we used a 12-item version of the Locus of Control scale (Rotter 1966). The Locus of Control scale classifies individuals along a single dimension capturing the degree to which they feel they control the outcomes of events. Individuals with an internal locus of control feel they control their own destiny, and believe outcomes they realize are the product of their own efforts and skills. Those with an external locus of control believe outcomes are outside their control. To measure behavioral inhibition, the survey included the 16-item Adult Measure of Behavioral Inhibition (AMBI) battery (Gladstone and Parker 2005). Each item is measured on a three-point scale, and the scores on all items are summed to obtain a variable ranging from zero to 32. We code the variable so that individuals with a higher score are more behaviorally disinhibited and outgoing, so that we would expect a positive correlation with risk attitudes. The Behavioral Inhibition variable is a subjective measure designed to capture how an individual responds to novel social situations.

Both our Rotter Locus of Control and Behavioral Inhibition variables are ordinal variables that are based on scales that were developed, tested, and calibrated to have good psychometric properties, and both variables are commonly used in the psychology literature.[7]

Other administrative data We were able to obtain background statistics on education, income, and marital status from administrative records for virtually all respondents.

---

[7]As we explain below, our analyses focus on the permanent components of these ordinal variables, which we assume are continuously distributed latent variables. We do not treat the ordinal variables as having cardinal significance.

Table 1 Summary statistics

	Respondents (R)			Non-Respondents (N-R)			Retest Sample (RS)			p-value	p-value
	Mean	S.D.	#	Mean	S.D.	#	Mean	S.D.	#	R=N-R	R=RS
Birth Year	1949.9	4.6	11,418	1950.4	4.6	13,472	1949.8	4.6	496	< 0.01	0.87
1 if Female	0.54	0.50	11,418	0.48	0.50	13,472	0.55	0.50	496	< 0.01	0.88
Education	11.94	2.72	11,167	11.42	2.64	12,313	12.13	2.80	487	< 0.01	0.13
Income	294	186	11,188	275	211.5	12,361	294	164.5	489	< 0.01	0.98
1 if Married	0.65	0.48	11,191	0.59	0.49	12,367	0.66	0.47	490	< 0.01	0.52
1 if MZ	0.27	0.44	11,418	0.23	0.42	13,471	0.23	0.42	496	< 0.01	0.03
1 if SS DZ	0.36	0.48	11,418	0.37	0.48	13,471	0.38	0.49	496	< 0.01	0.52
1 if OS DZ	0.36	0.48	11,418	0.39	0.49	13,471	0.39	0.49	496	< 0.01	0.17

NOTES: MZ: monozygotic twin. SS DZ: same sex dizygotic twin. OS DZ: opposite sex dizygotic twin. Income is reported in thousands of SEK and is defined as the sum of income earned from wage labor from own business, pension income, and unemployment income compensation. The education variable produced by Statistics Sweden is categorical and the categorical scores are converted into years of education using the population averages in Isacsson (2004). Survey respondents are the individuals who responded to the survey and gave informed consent to have their responses analyzed; non-respondents are the individuals who failed to respond to the survey

2.5 Summary statistics

To ascertain how representative our sample is, we present summary statistics for the background variables in Tables 1 and 2, comparing SALTY respondents to the twins who declined to participate ("Non-Respondents") and to the subset of twins who participated in the retest survey ("Retest Sample"). Comparing the respondents to the retest participants, we find that the differences are typically small and only statistically significant in one case: the share of MZ respondents in the sample.[8] As in other twin studies and most surveys in which resampling is not possible, women are somewhat overrepresented (Lykken et al. 1987) when we compare respondents to the non-respondents. Respondents are also better educated compared with non-respondents, with a difference of approximately 0.5 years of educational attainment. These differences are statistically significant, but the magnitude of the differences is typically modest, rarely exceeding a 10th of a standard deviation.

3 Latent variable model

We model each of the five ordinal measures of risk attitudes with a random-effects ordered probit model. For each measure, we model the permanent component of risk attitudes as a continuous variable ρ_{ij}^*, where $i = 1...N$ indexes the respondents and

---

[8]This difference is a rather mechanical consequence of the fact that our resampling procedure drew participants at the family level. The fact that MZ twins are more likely to be concordant in their decision to participate will depress the share of MZ twins in the retest sample.

Table 2 Summary statistics for respondents and retest sample

	Respondents					Retest Sample					p-value
	Mean	S.D.	Min	Max	#	Mean	S.D.	Min	Max	#	R=RS
Risk General	4.53	2.16	1	10	11,244	4.59	2.11	1	10	492	0.39
Risk Financial	3.51	2.11	1	10	11,279	3.50	2.10	1	10	492	0.93
Risk HRS	0.99	1.05	0	3	10,583	1.07	1.08	0	3	478	0.08
Risk Gain	0.10	0.30	0	1	11,120	0.08	0.27	0	1	487	0.06
Risk Loss	0.67	0.47	0	1	10,829	0.64	0.48	0	1	474	0.13
Alcohol	0.84	0.37	0	1	10,136	0.83	0.38	0	1	442	0.43
Birth Weight	2.67	5.16	0.97	5.00	10,820	2.71	5.12	1.18	4.42	474	0.04
Equity Share	0.03	0.10	0	1	4,328	0.04	0.10	0	0.8	209	0.63
Behav. Inhibition	17.41	4.97	0	32	10,649	17.39	5.01	0	29	466	0.92
Cognitive Ability	0.19	0.94	−2.99	3.23	3,873	0.33	1.01	−2.54	3.05	168	0.08
Locus of Control	6.45	2.14	0	12	9,472	6.50	2.02	1	11	410	0.61
Own Business	0.25	0.43	0	1	11,185	0.22	0.42	0	1	489	0.25
Portfolio Risk	18.99	6.39	0	53	6,622	18.78	6.45	0	41	268	0.53
Smoking	0.55	0.50	0	1	10,819	0.55	0.50	0	1	474	0.65

NOTES: Respondents are defined as all individuals who answered the first wave of the survey. The retest respondents are those who also responded to the second wave of the survey

$j = 1...5$ indexes the variables. For notational convenience, we will only include the subscripts i and j when confusion may otherwise arise. We assume ρ^* depends on the vector of covariates $\mathbf{x}$ in the following way:

$$\rho^* = \mathbf{x}\beta + \varsigma^*, \tag{1}$$

where ς^* is the part of permanent risk attitudes that is orthogonal to the covariates $\mathbf{x}$ and is assumed to be normally distributed with mean zero.

ρ^* is also subject to an additive disturbance m_t, which is independent across measurements,

$$r_t^* \equiv \rho^* + m_t, \tag{2}$$

where $t = 1, 2$ indexes the measurements. (As mentioned earlier, there were two measurements of risk attitudes for a subset of nearly 500 respondents.[9]) In Eq. 2, ρ^* corresponds to the permanent component of risk attitudes (including the effects of the covariates $\mathbf{x}$) and m_t represents the part of measured risk attitudes that is not stable across measurements. For simplicity, we will refer to m_t as "measurement error" and to the estimates of the model as being "measurement-error-adjusted". We follow Hey and Orme (1994) and assume m_t reflects white noise and is normally distributed with mean zero.

---

[9]Having retest data for a larger share of the respondents would increase the precision of our estimates. Nonetheless, the estimators we employ are consistent and any uncertainty due to the limited number of retested respondents will be reflected in the standard errors of our estimates (which are also consistently estimated).

m_t could in principle be further decomposed into a pure measurement error component and a component that captures transitory fluctuations due to changes in observable factors over time[10]. Indeed, it has been shown that personal experiences and time-varying attributes can affect risk attitudes and risk taking (e.g., Malmendier and Nagel 2011; Sahm 2012). We do not attempt such a decomposition of m_t here, in part because previous work has shown that changes in observables can explain only a small share of the transitory variation in measured preferences (Andersen et al. 2008; Josef et al. 2016; Meier and Sprenger 2010; Sahm 2012), and in part because only about a year elapsed on average between the first and the second measurements. Also, though previous work (e.g., Choi et al. 2014; Dave et al. 2010; von Gaudecker et al. 2011) has shown the consistency or variance of decision-making under uncertainty is related to socio-economic characteristics, we do not model m_t as dependent on covariates, as this would add a significant layer of complexity to the model and estimation and is not of direct relevance for the objectives of this paper.

We only directly observe the ordinal variable r_t, which is assumed to be monotonically related to the variable r_t^*:

$$r_t = 0 \text{ if } r_t^* < \tau_1; r_t = 1 \text{ if } \tau_1 < r_t^* < \tau_2; ...; r_t = C \text{ if } r_t^* > \tau_C, \qquad (3)$$

where $C+1$ is the number of response categories of r_t. Throughout, we use an asterisk (*) superscript to denote the latent (unobserved) variables. Thus, although the five risk-attitude variables we observe are ordinal variables, we are ultimately interested in the latent variables whose continuous normal distributions we assume underlie the ordinal variables' discrete distributions.[11] Our framework allows us to interpret the effects of our risk-attitude variables on other variables (and vice versa) in terms of standard deviations of permanent risk attitudes.

To identify and estimate the model, we assume $\sigma_m^2 = 1$.[12] We estimate the model by maximum likelihood.[13] The derivation of the likelihood function is presented in Appendix I.

---

[10]See, e.g., Sahm (2012) for an application in economics and McArdle and Woodcock (1997) for an application in psychology.

[11]Our approach thus differs from Dohmen et al.'s (2011) and Dohmen et al.'s (2012), who use the (standardized) ordinal Risk General and Risk Financial variables in their analyses of the intergenerational transmission of risk attitudes and of the determinants of risky behaviors, thus treating these ordinal variables as having cardinal significance.

[12]As is usual for ordered response models, the above model is not identified without further restrictions, because any affine transformation can be applied to Eq. 1 and to the thresholds τ_c, $c = 1...C$ in Eq. 3 without affecting the substantive implications of the model. Also for that reason, we do not include a constant in the vector of covariates **x**.

[13]We used the Stata program GLLAMM (Rabe-Hesketh et al. 2004) to estimate all models by maximum likelihood, using spherical quadrature with nine integration points. To account for non-independence within twin pairs, we clustered standard errors at the family level. We used the delta method to obtain standard errors for the transformed parameters.

4 Test-retest reliability

An important concept in classical test theory is the test-retest reliability, or reliability coefficient, of an instrument. For a continuous variable y, the test-retest reliability is defined as the ratio of the variance of the (latent) variable of interest to the variance of the observed variable. If $y_t = y_o^* + m_t$, where y_o^* is the permanent component and m_t is a classical measurement error, the test-retest reliability is given by

$$R = \frac{\sigma_{y_o^*}^2}{\sigma_{y_o^*}^2 + \sigma_m^2}.$$

If y is measured on two separate occasions, R can be obtained by calculating the test-retest correlation $R = corr(y_1, y_2)$.[14] One can think of the test-retest correlation as a measure of the fraction of variation that is due to the permanent component of the variable.

For ordinal variables, we can obtain an analogue to R by calculating the polychoric correlation between the variables across measurement occasions, assuming the underlying variable is normally distributed. In the context of the above model, keeping the covariates **x** fixed, the polychoric correlation is given by

$$R_{|\mathbf{x}}^* = corr(r_1^*, r_2^*|\mathbf{x}) = corr(\varsigma^* + m_1, \ \varsigma^* + m_2) = \frac{\sigma_{\varsigma^*}^2}{\sigma_{\varsigma^*}^2 + 1},$$

where the last equality follows from our assumption that $\sigma_m^2 = 1$. We use an asterix (*) superscript to distinguish the retest reliability of the ordinal variables from the retest reliability of the continuous variables, and we use the "|**x**" subscript to indicate that covariates are partialled out from the expression. We focus on ς^* (rather than ρ^*), because doing so allows us to estimate the test-retest reliability of risk attitudes residualized on the sex and birth year covariates, and because the resulting expression is also valid when there are no covariates (since $\varsigma^* = \rho^*$ when there are no covariates). We consistently estimate $R_{|X}^*$ by replacing $\sigma_{\varsigma^*}^2$ with its maximum likelihood (ML) estimate $\hat{\sigma}_{\varsigma^*}^2$.

4.1 Results

Figure 1 shows the test-retest reliabilities and their standard errors for the five risk variables (the corresponding numerical estimates are presented in Table I of Online Appendix V). The light gray bars in the figure show reliabilities for the case without any covariates (in which $\rho^* = \varsigma^*$). The Risk General and Risk Financial questions have the highest estimated reliabilities at 0.63 and 0.67, respectively; the Risk HRS variable has a somewhat lower reliability at 0.59; the Risk Gain and Risk Loss questions

---

[14]Economists may recognize the formula because it corresponds to the "errors in variables" formula derived by Friedman (1957) in his statement of the Permanent Income Hypothesis. The term we here define as measurement error could in principle be decomposed into one component due to changes in observable factors and an idiosyncratic error.
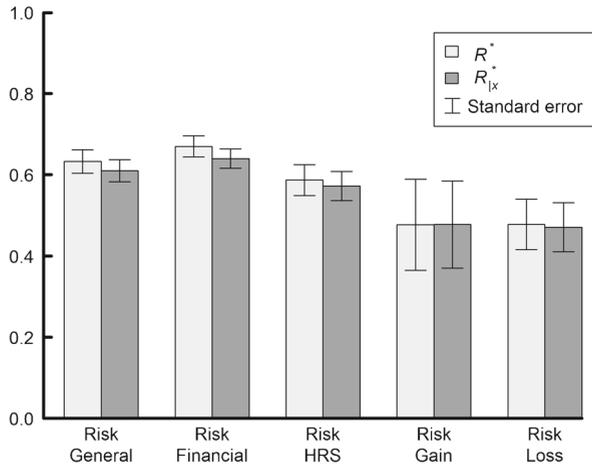
Fig. 1 Estimated test-retest reliabilities, with ($R_{|X}^*$) and without (R^*) the covariates sex and birth year partialled out. (The corresponding numerical estimates are shown in Table I of Online Appendix V)

have the lowest reliabilities at 0.48, but these estimates are less precise. The dark gray bars in the figure show the estimated reliabilities with sex and age partialled out. The test-retest reliabilities fall only marginally, suggesting that most of the systematic individual-level variation in risk is unrelated to sex and birth year.

These reliability estimates are lower than what is typically observed in the psychology literature on personality and cognitive ability. This finding is not surprising given that tests of personality and cognitive ability are based on a larger number of items. However, the reliabilities we report are substantially higher than those in the existing literature on risk attitudes. For example, Kimball et al. (2008) report a retest rank correlation of 0.27 across two waves of the HRS, and estimate a true-to-proxy variance ratio of 6.32 (the proxy being the empirical Bayes prediction of each individual's risk variable), implying a low reliability. These numbers suggest that transitory variance is lower in our sample than in the HRS, and that caution is warranted when disattenuating estimates based on estimated reliabilities from other samples. Lönnqvist et al. (2014) report retest rank correlations of 0.55 for the Risk Financial question and 0.77 for the Risk General question in a small sample of German subjects, which is broadly consistent with the findings of this paper. They also report a substantially lower reliability, a rank correlation of 0.26, for a laboratory-based measure of risk aversion in which subjects' preferences are elicited by measuring attitudes over small-stakes gambles with real monetary payoffs. The most extensive test of the reliability of gambles with real monetary payoffs is the study by Andersen et al. (2008), who test the temporal stability of four lottery tasks adapted from Holt and Laury (2002) in a representative sample of the Danish population.[15] The retest correlations range from 0.34 to 0.58, with a mean for the four tasks of 0.45, which is

---

[15]Other papers that have examined temporal stability of choices over lotteries with monetary payoffs include Horowitz (1992) and Harrison et al. (2005). Neither paper reports retest correlations.

similar to what we find for Risk Gain and Risk Loss, but somewhat lower than what we find for Risk HRS, Risk General, and Risk Financial.[16]

For brevity, most of the remainder of this paper presents and discusses results only for the Risk General, Risk Financial, and Risk HRS variables, which have higher reliability; results for the Risk Gain and Risk Loss variables are presented in the Online Appendix.

5 Predictive validity

If the risk variables are valid proxies for risk attitudes, and preference heterogeneity is an important determinant of individual variation in risky behaviors, we should expect the risk variables to explain a significant part of the cross-sectional variation in risky behaviors. We study five outcome variables: Portfolio Risk, Equity Share, Own Business, Alcohol Consumption, and Smoking. Methodologically, we closely follow Kimball et al. (2008), but we adjust their estimator – which we will refer to as the "KSS estimator" – for clustering, to account for the possible correlation in the error terms within twin pairs. We begin by obtaining posterior estimates – or scores – of the expected value of the risk variable for each individual, where the expectation is taken conditional on the observed data and the estimated parameters. Then we use the KSS estimator to obtain unbiased estimates of the effects of the permanent component of risk attitudes on the outcomes of interest.[17] The KSS estimator is a generalized method of moments (GMM) estimator that accounts for the particular nature of the measurement error present in the posterior scores. We now explain how we compute the posterior scores and use these posterior scores to obtain consistent estimates of the regression coefficients.

5.1 Posterior scores for ρ^*

To obtain the posterior scores, we take the posterior expectation of ρ^* for each respondent, conditional on the ML estimates $\hat{\Theta}$ and the respondent's observed responses r_1 and r_2 (when retest data are available) and covariates $\mathbf{x}$:

$$\hat{\rho}^{EB} = E[\rho^*|\mathbf{r}, \mathbf{x}, \hat{\Theta}]. \tag{4}$$

---

[16]A strict comparison of the reliability of our hypothetical risk questions and that of gambles with real monetary payoffs would require obtaining estimates from the same sample. Studies comparing hypothetical versus incentivized gambles suggest that using financial incentives leads to more risk-averse behavior, and that increasing the size of the financial incentives leads to more risk-averse behavior (Camerer and Hogarth, 1999; Holt and Laury, 2002, 2005). However, because we are not estimating the degree of risk aversion per se, it is not obvious that "hypothetical bias" will affect the reliability and predictive validity of our risk questions. Further, one advantage of using hypothetical questions is that higher stakes could be used, thereby avoiding the Rabin (2000) critique of estimating risk aversion from small-stake gambles. Future work that directly compares the reliability and predictive validity of questions with and without financial incentives would be interesting.

[17]Throughout, we often refer to estimated coefficients as effects, but we note at the outset that a causal interpretation is not necessarily warranted.

The posterior distribution of ς^* given $\mathbf{r}$, $\mathbf{x}$, and $\hat{\Theta}$ is

$$f(\varsigma^*|\mathbf{r}, \mathbf{x}, \hat{\Theta}) = \frac{f(\mathbf{r}, \varsigma^*|\hat{\Theta}, x)}{f(\mathbf{r}|\hat{\Theta}, x)},$$

where $f(\mathbf{r}, \varsigma^*|\hat{\Theta}, x)$ is the joint distribution function of $\mathbf{r}$ and ς^* and $f(\mathbf{r}|\hat{\Theta}, x)$ is the marginal distribution function of $\mathbf{r}$ for a respondent (see Appendix I). From this, we can compute the posterior distribution of $\rho^* = \mathbf{x}\beta + \varsigma$ given $\mathbf{r}$, $\mathbf{x}$, and $\hat{\Theta}$, with respect to which the expectation in Eq. 4 is taken. The ML estimate $\hat{\Theta}$ is plugged directly into Eq. 4 and is thus taken as given in the estimation of $\hat{\rho}^{EB}$; for that reason, this approach is called empirical Bayesian, and the posterior scores thus obtained are referred to as empirical Bayes predictions, factor scores, or posterior means. Skrondal and Rabe-Hesketh (2009) describe this approach in detail.[18]

Because we control for covariates in the next step and because we are not presently interested in the effect of covariates on risk attitudes, we obtain the posterior scores without including any covariates $\mathbf{x}$ in the model. We standardize the posterior scores by dividing them by $\hat{\sigma}_{\rho^*} = \hat{\sigma}_{\varsigma^*}$.[19] Importantly, the standardization is with respect to the standard deviation σ_{ρ^*} of the permanent component of risk attitudes, not the standard deviation of the posterior scores; the resulting posterior scores are thus the posterior expectations of the standardized permanent component of risk attitudes and will have variance smaller than unity. For notational simplicity, we still denote those standardized posterior scores $\hat{\rho}^{EB}$.

5.2 The KSS estimator

To assess the effects of the permanent component of risk attitudes on outcomes of interest, we would like to obtain consistent estimates of δ_ρ in the regression

$$y = \rho^*\delta_\rho + \mathbf{z}\delta_z + \upsilon, \tag{5}$$

where y is the outcome of interest and $\mathbf{z}$ contains the other covariates of the regression. (For expositional convenience, we mostly follow Kimball et al.'s notation). However, ρ^* is not directly observed, and simply replacing ρ^* with $\hat{\rho}^{EB}$ in Eq. 5 would yield biased estimates. To see what generates the bias, notice that $\rho^* = \hat{\rho}^{EB} + u$, where u is the expectation error from Eq. 4; u is different from a classical measurement error, because it is correlated with the true value ρ^* rather than with the observed variable $\hat{\rho}^{EB}$. Thus, simply substituting $\hat{\rho}^{EB} + u$ for ρ^* in Eq. 5 yields

$$y = \hat{\rho}^{EB}\delta_\rho + \mathbf{z}\delta_z + \eta,$$

where $\eta = u\delta_\rho + \upsilon$. The resulting OLS estimates is biased because $E[\mathbf{z}\eta] = \delta_\rho E[\mathbf{z}u] \neq 0$ in general.

---

[18]This approach is implemented in the Stata program GLLAMM referred to earlier.

[19]We also conducted this exercise with the posterior scores $\hat{\varsigma}^{EB} = E[\varsigma^*|\mathbf{r}, \mathbf{x}, \hat{\Theta}]$, with covariates sex and birth year in $\mathbf{x}$; we standardized $\hat{\varsigma}^{EB}$ by dividing it by $\hat{\sigma}_{\varsigma^*}$. In this case, δ_ρ (see below) still gives the effect of a one-standard-deviation increase in true risk attitudes on the outcome of interest, but the standard deviation here is that of true risk attitudes with the effects of $\mathbf{x}$ partialled out. The results were very similar.

Kimball et al. ([2008]) develop a GMM-estimator that consistently estimates δ_ρ by using the previously derived posterior scores $\hat{\rho}^{EB}$. To do so, they assume

$$E[z_k|\rho^*] = \rho^*\gamma_k, \tag{6}$$

where z_k is a covariate in $\mathbf{z}$, $k = 1...K$, and $\gamma_k = E[\rho^{*2}]^{-1}E[\rho^*z_k]$.[20] They then show that these moment conditions hold:

$$E[\hat{\rho}^{EB}\eta] = E[\hat{\rho}^{EB}(y - \hat{\rho}^{EB}\delta_\rho - \mathbf{z}\delta_z)] = 0, \tag{7}$$

$$E[\mathbf{z}'\omega] = E[\mathbf{z}'(y - \lambda\hat{\rho}^{EB}\delta_\rho - \mathbf{z}\delta_z)] = 0, \tag{8}$$

where $\lambda = E[\rho^{*2}]/E[(\hat{\rho}^{EB})^2]$ and $\omega = (\rho^* - \lambda\hat{\rho}^{EB})\delta_\rho + \upsilon$. Kimball et al. ([2008]) derive the corresponding GMM estimator and its asymptotic variance-covariance matrix; they also derive the R^2 implied by the model in Eqs. [5] and [6]. We develop the relevant asymptotics with clustering in the Online Appendix.

Some of the risky outcomes we study are binary variables. If we rewrite ([5]) as a linear probability model,

$$Prob(y = 1|\rho^*, \mathbf{z}) = E[y|\rho^*, \mathbf{z}] = \rho^*\delta_\rho + \mathbf{z}\delta_z,$$

the moment conditions ([7]) and ([8]) still hold. Thus, the KSS estimator and asymptotic variance-covariance matrix are still valid with binary dependent variables, and the coefficients δ give the effects of the covariates on the probability that $y = 1$, as in a linear probability model.

Because of our above standardization, $\hat{\rho}^{EB}$ is scaled in standard deviations of the permanent component of risk attitudes ρ^*. Therefore, the coefficient δ_ρ is the effect of a one-standard-deviation increase in permanent risk attitudes ρ^* on the outcome, holding constant the other covariates, where the standard deviation is that of permanent risk attitudes in the population from which the sample was drawn.

5.3 Model without measurement error

We are interested in comparing the KSS estimates to those that would be obtained without accounting for measurement error, and using only the data from the first measurement. To do so, we set $\sigma_m^2 = 0$ and begin by obtaining posterior estimates, or scores, of expected risk attitudes for each respondent conditional on the respondent's response category, using only the data from the first measurement and with the identifying assumption that $\sigma_{\varsigma^*}^2 = 1$. As before, we construct the posterior scores without partialling out any covariates (so $\rho^* = \varsigma^*$). The posterior scores with no measurement-error adjustment are thus given by

$$\hat{\rho}^{NMA} = E[\rho^*|r_1, \hat{\Theta}], \tag{9}$$

where $\hat{\Theta} = (\hat{\tau}_1, ..., \hat{\tau}_C)'$ and $\hat{\tau}_c = \Phi^{-1}(\Sigma_{r=0}^{r=c}\frac{N_r}{N})$, where Φ^{-1} is the inverse cumulative standard normal distribution, N_r is the number of respondents whose responses

---

[20]Note that $E[\rho^*] = 0$ here because we do not include covariates.

Table 3 Predictive validity of the risk variables: Controlling for cognitive ability, males only

	Portfolio Risk			Equity Share			Own Business			Alcohol			Smoking		
	$\hat{\delta}_p$	R^2	ΔR^2	$\hat{\delta}_p$	R^2	ΔR^2	$\hat{\delta}_p$	R^2	ΔR^2	$\hat{\delta}_p$	R^2	ΔR^2	$\hat{\delta}_p$	R^2	ΔR^2
Risk General															
$\sigma_m^2 = 0$	0.424***	0.024	0.003	0.010***	0.024	0.005	0.079***	—	—	0.018***	—	—	0.018**	—	—
	(0.142)			(0.004)			(0.009)			(0.006)			(0.009)		
$\sigma_m^2 > 0$	0.453***	0.026	0.004	0.010***	0.026	0.007	0.103***	—	—	0.022***	—	—	0.023*	—	—
	(0.182)			(0.005)			(0.011)			(0.008)			(0.012)		
Risk Financial															
$\sigma_m^2 = 0$	0.987***	0.040	0.020	0.020***	0.041	0.023	0.072***	—	—	0.018***	—	—	0.004	—	—
	(0.132)			(0.004)			(0.009)			(0.006)			(0.009)		
$\sigma_m^2 > 0$	1.242***	0.052	0.031	0.027***	0.061	0.043	0.093***	—	—	0.025***	—	—	0.007	—	—
	(0.177)			(0.005)			(0.011)			(0.008)			(0.012)		
Risk HRS															
$\sigma_m^2 = 0$	0.697***	0.029	0.009	0.010**	0.024	0.004	0.091***	—	—	0.007	—	—	−0.009	—	—
	(0.152)			(0.005)			(0.009)			(0.006)			(0.010)		
$\sigma_m^2 > 0$	1.070***	0.042	0.021	0.016**	0.034	0.013	0.145***	—	—	0.013	—	—	−0.009	—	—
	(0.242)			(0.008)			(0.015)			(0.010)			(0.016)		

NOTES: This table reports the KSS estimates of the effects of risk attitudes on the five outcome variables given in the column headings and the KSS R^2 for the model with measurement-error adjustment ($\sigma_m^2 > 0$), as well as the OLS estimates and R^2 for the model without adjustment ($\sigma_m^2 = 0$). 30 regressions were run to generate these results (three risk variables × five outcomes × two). Results are for males only. ΔR^2 is the incremental R^2 from including the risk variable in the regression. R^2 and ΔR^2 are not shown for Own Business, Alcohol, and Smoking, because these are binary dependent variables. All specifications include a constant and controls for birth year, birth weight, log income, marital status, years of education, and cognitive ability

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level
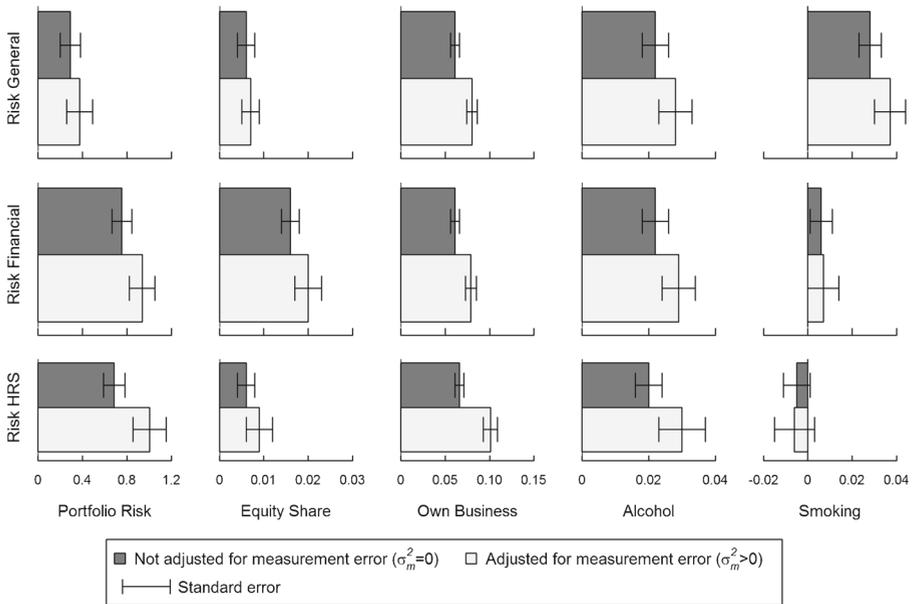
Fig. 2 Predictive validity of the risk variables in the entire sample. The figure shows the KSS estimates of the effects of risk attitudes on the five outcome variables for the model with measurement-error adjustment, and the OLS estimates for the model without adjustment. (The corresponding numerical estimates are shown in Table II of Online Appendix V)

r_1 are in the ordinal category r, and N is the total number of respondents in the sample. It follows that

$$\hat{\rho}^{NMA} = E[\rho^* | \hat{\tau}_r < \rho^* < \hat{\tau}_{r+1}],$$

which is easily obtained numerically because $\rho^* = \varsigma^*$ has a standard normal distribution.

As for the case with measurement-error adjustment, we would like to estimate (5), although here ρ^* refers to the risk variable in the model without measurement error. Following the above reasoning for $\hat{\rho}^{EB}$, we see that estimating (5) with $\hat{\rho}^{NMA}$ instead of ρ^* would yield biased and inconsistent estimates of δ; hence, the KSS estimator should be used to estimate δ. However, we find it more instructive to compare the consistent KSS estimates obtained using the measurement-error-adjusted posterior scores $\hat{\rho}^{EB}$ to the estimates from a procedure without measurement-error adjustment or appropriate econometric handling of the posterior scores, because this latter procedure is closer to what is usually done in the literature on the predictive power of risk attitudes. For that reason, we estimate

$$y = \hat{\rho}^{NMA} \alpha_\rho + \mathbf{z}\alpha_z + v$$

by OLS and cluster the standard errors at the family level.

5.4 Results

Table 3 reports the main results for the sample of male respondents, and Fig. 2 plots the main results for the entire sample (the corresponding numerical estimates are presented in Table II of Online Appendix V).[21] For both samples, we ran 30 regressions (three risk variables × five outcomes × two). Each specification includes the posterior scores for one of the risk variables as well as a vector of additional covariates **z**, which includes a constant, birth weight, birth year, the logarithm of income in 2005, marital status, years of education and, for the sample of male respondents, cognitive ability. For each risk variable, the first and second rows of Table 3 and the top and bottom bars of Fig. 2 show the estimates for the models without (OLS with $\hat{\rho}^{NMA}$, with $\sigma_m^2 = 0$) and with (the KSS estimator with $\hat{\rho}^{EB}$, with $\sigma_m^2 > 0$) measurement-error adjustment, respectively. Risk General is significantly related to all five measures of risky behaviors in the expected direction. (The association between Risk General and smoking behavior is only marginally significant in the sample of male respondents, but is very significant in the entire sample.) The results for the Risk Financial and Risk HRS questions are also strong, but these risk variables are not significantly related to smoking behavior.[22]

As expected, adjustment for measurement error substantially increases the estimated effects. The magnitude of several of the estimated effects is sizable. The coefficients on Own Business are particularly striking: after measurement-error adjustment, a one-standard-deviation increase in risk attitudes on the Risk HRS question is associated with increases of 10 and 14 percentage units in the probability of having started a business in the entire sample and in the male sample, respectively; the corresponding effects for the Risk General and Risk Financial questions range from eight to 10 percentage points. A one-standard-deviation increase in either of the three risk variables is associated with a two- to three-percentage-point increase in the probability of being an alcohol consumer, and a one-standard-deviation increase in Risk General is associated with a four-percentage point increase in the probability of being a smoker, in the entire sample. Also, including risk attitudes in the regressions of Portfolio Risk and Equity Share can more than double the R^2, as can be inferred from the fact that the incremental R^2 of including risk attitudes in the regressions are sometimes more than half the reported R^2. The directions of our effects are in line with those of Barsky et al. (1997) and Dohmen et al. (2011) for comparable behaviors.

---

[21] For all predictive validity regressions, we include all respondents without missing observations. Although the variable Equity Share takes the value of zero for a number of respondents, we use the KSS estimator for that variable as well (a Tobit model would be inappropriate, as such a model does not correct for the measurement error in the posterior scores).

[22] To explore whether these insignificant associations with smoking behavior could be attributable to the way we constructed the smoking variable, we also tried specifications with a different smoking variable which only coded current, and not past, smokers as smokers. The results were generally even less significant than those for our baseline smoking variable. We also tried specifications with a different alcohol-consumption variable, constructed using the responses to a SALT question about excessive alcohol consumption; the results were similar to those for our baseline variable.

6 Predictors of risk attitudes

We turn now to the question of what variables predict risk attitudes. We focus our attention on variables that were determined and measured long before the SALTY survey was administered: birth weight, birth year, cognitive ability, education, and sex. The framework we use allows for consistent estimation of the regression coefficients in the presence of measurement error.

6.1 Model with measurement error

For ease of interpretation, we standardize our estimated coefficients by dividing Eq. 1 by the standard deviation of the permanent component of risk attitudes:

$$\beta_{std} = \frac{\beta}{\sigma_{\rho*}} = \frac{\beta}{\sqrt{\beta'\Sigma\beta + \sigma_{\varsigma*}^2}}. \tag{10}$$

$\sigma_{\rho*}^2 = \beta'\Sigma\beta + \sigma_{\varsigma*}^2$ is the population variance of the latent risk variable ρ^*, and Σ is the population variance-covariance matrix of $\mathbf{x}$. Therefore, β_{std} is the (vector of) effects of one-unit increases in each covariate on the latent risk variable, where the latter is standardized to have a variance of 1. To assess the explanatory power of the covariates, we also introduce a pseudo-R^2:

$$R_{pseudo}^2 = \frac{\beta'\Sigma\beta}{\sigma_{\rho*}^2} = \frac{\beta'\Sigma\beta}{\beta'\Sigma\beta + \sigma_{\varsigma*}^2}. \tag{11}$$

β_{std} and R_{pseudo}^2[23] are consistently estimated by replacing β and $\sigma_{\varsigma*}^2$ in Eqs. 10 and 11 by their respective ML estimates and Σ by $\hat{\Sigma} = X'\mathbf{X}/N$, a consistent estimator of the variance-covariance matrix of $\mathbf{x}$ in the subpopulation from which the sample is drawn ($\mathbf{X}$ is the matrix of the demeaned covariates $\mathbf{x}_i$ of all respondents). We compute standard errors using the delta method, taking $\hat{\Sigma}$ as given.[24]

---

[23]There is an intuitive parallel between R_{pseudo}^2 and the R^2 from an OLS regression analysis: in both cases, the numerator measures the explained sum of squares, the denominator corresponds to the total sum of squares, and thus both are estimates of the share of the variance of the variable of interest that is attributable to variation in the covariates. However, whereas the denominator of the R^2 of regression analysis (the total sum of squares) can be calculated directly from the data and relates to the sample, the denominator of $\hat{R}_{pseudo}^2$ is obtained indirectly from estimated parameters and can only be interpreted as a consistent estimate of the true population value.

$\hat{R}_{pseudo}^2$ in the model without measurement error (the ordered probit model, described below) is identical to the pseudo-R^2 developed by McKelvey and Zavoina (1975) and has been widely used in the sociology and political science literatures. Its sampling properties have been examined by simulation for the case in which the ordinal variable is dichotomous, and its performance compared quite favorably to that of several other pseudo-R^2's (Hagle and Mitchell 1992). The simulations suggest it is a good estimate of the OLS R^2 associated with the continuous variable underlying the dichotomous variable in the simulations. Despite its intuitive appeal and its popularity in other fields, the pseudo-R^2 is rarely used in ordered probit analysis in economics.

[24]Therefore, the standard errors do not reflect the uncertainty from the estimation of Σ and are biased downward.

Table 4 Predictors of risk attitudes: controlling for cognitive ability, males only

	Risk General		Risk Financial		Risk HRS	
	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$
Birth Year	0.015***	0.020***	0.015***	0.019***	0.025***	0.032***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.006)
Birth Weight	0.073**	0.095**	0.044	0.051	0.013	0.021
	(0.034)	(0.044)	(0.034)	(0.043)	(0.036)	(0.047)
Education	0.034***	0.042***	0.053***	0.067***	0.050***	0.066***
	(0.007)	(0.010)	(0.007)	(0.009)	(0.008)	(0.011)
Cognitive	0.119***	0.158***	0.154***	0.197***	0.232***	0.299***
Ability	(0.022)	(0.027)	(0.022)	(0.028)	(0.023)	(0.040)
R_{pseudo}^2	0.040	0.067	0.070	0.115	0.112	0.188
n_1	3,605	3,605	3,616	3,616	3,482	3,482
n_2	0	158	0	158	0	150
$ln(L)$	−7,563.68	−7,854.48	−7,413.62	−7,700.46	−4,347.91	−4,523.23

NOTES: The estimates in the first column under each variable name are not adjusted for measurement error ($\sigma_m^2 = 0$); the estimates in the second column are adjusted ($\sigma_m^2 > 0$). Results are for males only. All specifications include covariates for birth year, birth weight, educational attainment, and cognitive ability. Standard errors (in parentheses) do not reflect the uncertainty from the estimation of Σ and are thus downward biased

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level

6.2 Model without measurement error

For the model with no measurement error ($\sigma_m^2 = 0$) and with data from the first measurement only, the model described in Eqs. 1, 2, and 3 becomes the commonly used ordered probit model. To identify the model, we assume that $\sigma_{\varsigma^*}^2 = 1$. The definitions of β_{std}, R_{pseudo}^2 and their estimators $\hat{\beta}_{std}$ and $\hat{R}_{pseudo}^2$ still apply.

6.3 Results

Table 4 reports the results for the sample of male respondents and Fig. 3 plots the results for the entire sample (the corresponding numerical estimates are shown in Table III of Online Appendix V). All specifications include the covariates birth year, birth weight, years of education; specifications for the sample of male respondents also include the covariate cognitive ability (which is not available for females), and specifications for the entire sample also include the sex covariates. All covariates were determined and measured many years before the risk variables were measured in the SALTY survey. We find that years of education is associated with a greater willingness to take risks and that birth weight is positively associated with Risk General. Table 4 also reports the $\hat{R}_{pseudo}^2$ from each model. The covariates explain a non-negligible share of the variation in the risk variables, especially after adjustment for measurement error.
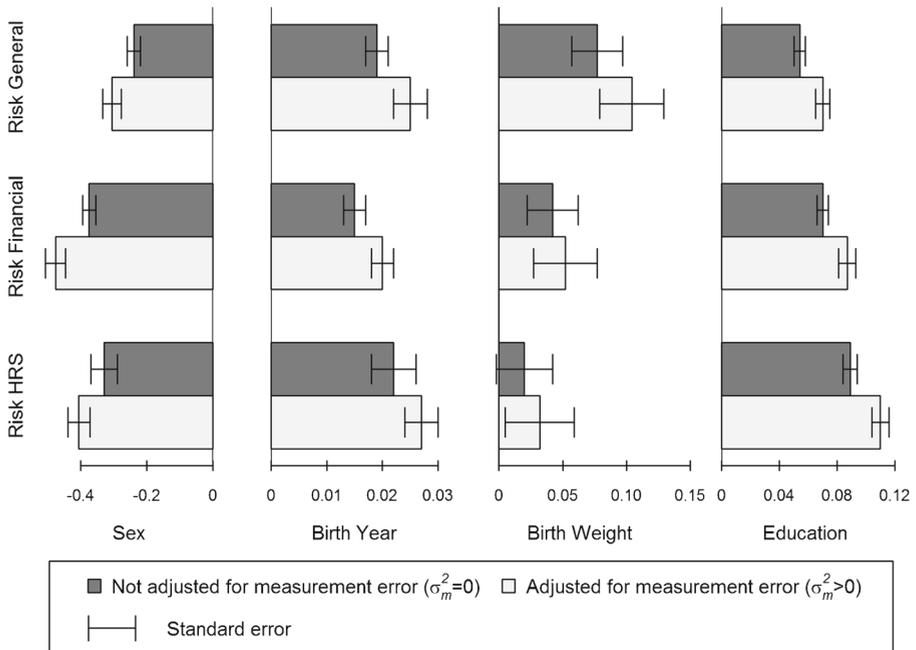
Fig. 3 Predictors of risk attitudes in the entire sample. The figure shows the estimates of the effects of sex, birth year, birth weight, and education on permanent risk attitudes, for the models with and without measurement-error adjustment. (The corresponding numerical estimates are shown in Table III of Online Appendix V)

We replicate previously reported associations between attitudes toward risk and cognitive ability (Benjamin et al. 2013; Burks et al. 2009; Dave et al. 2010; Dohmen et al. 2010; Frederick 2005) in the sample of male respondents. Our study departs from these previous works in that our model corrects for measurement error, and our measure of cognitive ability was taken when the male respondents in our sample were about 18 years old. Because our respondents were born between 1943 and 1958, the average respondent would have taken the test of cognitive ability approximately four decades before the administration of the SALTY survey. Viewed in this light, the fact that our measurement-error-adjusted estimates suggest a one-standard-deviation increase in cognitive ability is associated with up to a 0.3-standard-deviation increase in risk attitudes is remarkable. The association between IQ and risk attitudes is also consistent with the recent findings of Grinblatt et al. (2011), who found that stock market participation increases with higher IQ. Further, the association is very significant for all risk variables, including Risk General and Risk Financial (which involve simple self-rated scales), thus suggesting the association is not only or primarily driven by the fact that measurement noise may be higher for individuals with lower cognitive ability (Andersson et al. 2013).

The estimates for the entire sample suggest there are some striking differences between males and females, which appear to be most pronounced for the Risk Financial and Risk HRS questions. Based on those questions, the females in our sample

have risk attitudes that are nearly half a standard deviation lower than those of men, holding the other covariates constant. The difference is smaller but still sizable for the Risk General question.

These findings are consistent with an emerging literature documenting differences in the measured preferences of males and females (e.g., Croson and Gneezy, 2009).[25] The difference between males and females may be important from a policy point of view and it may account for part of the gender wage-gap: Bonin et al. (2007) find that individuals who are more willing to take risks sort into occupations with more earnings risks and higher average earnings.

7 Multivariate factor model

Having examined how each of the five latent risk-attitude variables relate to other variables, we now ask how they relate to one another. If these were continuous variables without measurement error, two natural next steps would be to compute the pairwise correlations between the variables and to extract their first principal component. Here, however, the variables are ordinal and measurement error is present; we thus expand our model to a multivariate setting in a way that allows the computation of measurement-error-adjusted pairwise correlations and the identification of a variable akin to the first principal component adjusted for measurement error. We label this variable the common factor.[26]

7.1 The multivariate factor model

To expand our model to a multivariate setting, we assume the permanent components of risk attitudes depend on the common factor f^* as follows:

$$\rho_j^* = \mathbf{x}\beta_j + \varsigma_j^* = \mathbf{x}\beta_j + \lambda_j f^* + \varepsilon_j^*, \, (j = 1...J), \tag{12}$$

where the subscript j indexes the different risk variables we include in the multivariate model, λ_j is the factor loading of ρ_j^* on the common factor f^*, and ε_j^* is the unexplained part of permanent risk preference ρ_j^* in Eq. 12 – that is, the part that is consistently measured across tests but is not correlated with the other variables and thus is not captured by $\mathbf{x}\beta_j$ or $\lambda_j f^*$.

For each risk variable j, we maintain the j-subscripted analogues of Eqs. 2 and 3 from our univariate model. The resulting model can be summarized as follows:

$$r_{jt}^* = \rho_j^* + m_{jt}; \quad \rho_j^* = \mathbf{x}\beta_j + \lambda_j f^* + \varepsilon_j^*; \quad r_{jt} = 0 \text{ if } r_{jt}^* < \tau_{j1}; \tag{13}$$
$$r_{jt} = 1 \text{ if } \tau_{j1} < r_{jt}^* < \tau_{j2}; \, ... \, ; \, r_{jt} = C_j \text{ if } r_{jt}^* > \tau_{jC}; \quad (j = 1...J).$$

---

[25] For an exception to this result, see Harrison et al. (2007), who did not find a significant gender difference in risk attitudes.

[26] Though we use an analogy to principal component analysis for expositional convenience, our multivariate model falls under the purview of factor analysis (Bartholomew and Knott 1999).

We assume f^*, ε_j^*, and m_{jt} are normally distributed with mean zero. To identify the model, we assume further and without loss of generality that f^* and m_{jt} have unit variance. For simplicity and to reduce the computational burden, we did not consider models with covariates **x** for this exercise.

Two quantities are useful for summarizing the results of our estimation of Eq. 13. The first is the pseudo-R^2 describing the fraction of variance in permanent risk attitudes $\rho_j^* = \varsigma_j^*$ that variation in the common factor f^* explains. It is given by

$$R^2_{pseudoMV,j} = \frac{\lambda_j^2}{\lambda_j^2 + \sigma_{\varepsilon_j}^2}. \tag{14}$$

The second is the correlation between the common factor and each permanent risk-attitude variable, which is given by

$$corr(\rho_j^*, f^*) = \frac{\lambda_j}{\sqrt{\lambda_j^2 + \sigma_{\varepsilon_j}^2}}. \tag{15}$$

We obtain consistent estimates of these two expressions by substituting the estimates $\hat{\lambda}_j^2$ and $\hat{\sigma}_{\varepsilon_j}^2$ from the estimation of the multivariate model (12).[27] In the absence of measurement error ($\sigma_m^2 = 0$) and when only data from the first measurement are used, we make the usual identifying assumption that $\sigma_{\varepsilon_j}^2 = 1$ ($j = 1..J$), and expressions (14) and (15) still apply.

The Online Appendix describes how to obtain measurement-error-adjusted estimates of the pairwise correlation between any two variables, and reports estimates of all the pairwise correlations between the risk variables.

7.2 Results

Table 5 reports estimates of $R^2_{pseudoMV,j}$ and $corr(\rho_j^*, f^*)$ for the multivariate factor model with all five risk-attitude variables, with and without measurement-error adjustment. The measurement-error-adjusted estimates are much larger than the unadjusted estimates. Notably, the correlation between the common factor and the measurement-error-adjusted Risk Financial variable is 0.977, indicating the common factor is almost identical to that variable; the correlation between the common factor and the Risk General variable is also very high at 0.862. In part, the high correlations are a consequence of the fact that the Risk Financial and the Risk General variables are similar and highly correlated: as is well known in factor and principal component analysis, when similar variables are included, the first factor or principal component tends to load heavily on these variables and to explain their common variance. To circumvent this issue, we estimated two other multivariate factor models: one without the Risk General variable and the other without the Risk Financial variable. The common factors in both models are not dominated by a single variable and are more balanced. In all cases, the explanatory power of the common factor is high, demonstrating again the presence of much common variation between the different risk variables.

---

[27]To make the estimation computationally manageable, we collapsed the Risk General and Risk Financial variables into six and seven categories, respectively.

Table 5 Multivariate factor models

		Model with All Variables		Risk Finance Omitted		Risk General Omitted	
		$\sigma_m^2 = 0$	$\sigma_m^2 > 0$	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$
Correlation with f^*	Risk General	0.716 (0.008)	0.862 (0.020)	0.499 (0.022)	0.612 (0.025)	–	–
	Risk Financial	0.835 (0.005)	0.977 (0.016)	–	–	0.665 (0.028)	0.794 (0.031)
	Risk HRS	0.505 (0.011)	0.632 (0.024)	0.662 (0.030)	0.826 (0.044)	0.687 (0.028)	0.847 (0.041)
	Risk Gain	0.289 (0.018)	0.415 (0.046)	0.395 (0.023)	0.558 (0.064)	0.385 (0.021)	0.546 (0.058)
	Risk Loss	0.180 (0.014)	0.255 (0.025)	0.248 (0.019)	0.349 (0.033)	0.216 (0.016)	0.305 (0.029)
R^2 of f^*	Risk General	0.512 (0.012)	0.742 (0.035)	0.249 (0.022)	0.375 (0.030)	–	–
	Risk Financial	0.697 (0.008)	0.955 (0.030)	–	–	0.442 (0.037)	0.630 (0.050)
	Risk HRS	0.255 (0.011)	0.399 (0.030)	0.438 (0.039)	0.683 (0.073)	0.472 (0.038)	0.717 (0.070)
	Risk Gain	0.083 (0.010)	0.172 (0.039)	0.156 (0.018)	0.311 (0.071)	0.148 (0.016)	0.298 (0.064)
	Risk Loss	0.032 (0.005)	0.065 (0.013)	0.062 (0.009)	0.122 (0.023)	0.047 (0.007)	0.093 (0.018)
$ln(L)$		–61,104.80	–63,517.12	–42,749.57	–44,408.73	–43,565.62	–45,265.02

NOTES: The estimates in the first column under each variable name are not adjusted for measurement error ($\sigma_m^2 = 0$); the estimates in the second column are adjusted ($\sigma_m^2 > 0$). All estimates are significant at the 1% level

Because four of the variables concern risk-taking in the financial domain, they naturally share a considerable part of their variance. However, it was less obvious that the Risk General question would be so highly correlated with the four other measures. This result is in line with what Dohmen et al. (2011) find. Also, in agreement with our earlier results, Barsky et al. (1997) found that their HRS measure of financial risk-taking was correlated with both smoking and drinking and Dohmen et al. (2011) found that financial and health risk-taking were correlated. These findings certainly do not rule out the possibility that risk attitudes involve some domain-specificity, because the correlations between the risk measures in the different domains are always far from unity. Several studies from psychology also argue that risk-taking behavior is domain specific (Hanoch et al. 2006; Weber et al. 2002). Nonetheless, taken together these results suggest the existence of a common factor accounts for risk-taking across domains, consistent with Einav et al. (2012). Domain-specific utility parameters related to risky activities such as smoking and bungee jumping are likely to be important for these activities, as Hanoch et al. (2006) show, but this is consistent with the existence of a general risk factor.

7.3 Extensions: Posterior scores for f^*

As above for ρ^*, we can obtain posterior scores for f^* by taking the posterior expectation of f^* conditional on the observed responses $\mathbf{r}_j$ ($j = 1..J$), the covariates $\mathbf{x}$, and the ML estimates $\hat{\Theta}$:

$$\hat{f}^{EB} = E[f^*|\mathbf{r}_1, ..., \mathbf{r}_J, \mathbf{x}, \hat{\Theta}]. \tag{16}$$

As for $\hat{\rho}^{EB}$, substituting $\hat{f}^{EB}$ for f^* on the right-hand side of an OLS regression will yield biased estimates. Unfortunately, the econometrics of using $\hat{f}^{EB}$ as a covariate is not simple, because the KSS method is not valid with $\hat{f}^{EB}$.[28] A thorough investigation of the properties of the posterior scores $\hat{f}^{EB}$ is beyond the scope of this paper, but we believe such posterior scores and the accompanying multivariate factor models (such as the one presented above) are promising tools to summarize the information on economic preferences that is increasingly being collected in large surveys.

8 Other applications

In this section, we provide two concrete illustrations of how our framework can be used to provide measurement-error-adjusted estimates of the relationship between an individual's risk attitudes and other variables. In our first application, we examine the relationship between risk attitudes and personality. In our second application,

---

[28]To see, notice that $cov(\hat{f}^{EB}, \varepsilon_j^*) \neq 0$ if $\lambda_j \neq 0$ (for any $j \, \epsilon\{1...J\}$), and suppose $cov(\varepsilon_j^*, \upsilon) \neq 0$; i.e. the variable-specific factor ε_j^* in Eq. 12 affects y independently of f^* and $\mathbf{z}$ in Eq. 5. In that case, which cannot be ruled out, $cov(\hat{f}^{EB}, \upsilon) \neq 0$, and so the first moment of the KSS estimator (7) does not hold. This issue does not arise with $\hat{\rho}^{EB}$, because the term m_t in Eq. 2 is a pure measurement disturbance and is not correlated with anything.

Table 6 Correlations between risk attitudes and personality

	Behavioral Inhibition		Rotter Locus of Control	
	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$
Risk General	0.297*** (0.010)	0.448*** (0.018)	0.071*** (0.011)	0.119*** (0.019)
Risk Financial	0.222*** (0.010)	0.331*** (0.017)	0.100*** (0.011)	0.162*** (0.019)
Risk HRS	0.206*** (0.011)	0.338*** (0.024)	0.121*** (0.012)	0.207*** (0.021)

NOTES: The estimates in the first column under each variable name are not adjusted for measurement error ($\sigma_m^2 = 0$); the estimates in the second column are adjusted ($\sigma_m^2 > 0$)

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level

we report estimates of the sibling correlation in risk attitudes with and without adjustment for measurement error. We also perform a standard behavior genetic decomposition of the variance in permanent risk attitudes.

8.1 Correlates of risk attitudes

Our analyses to date considered variables that naturally belong on either the right-hand or left-hand side of regressions with risk attitudes. The SALTY survey contained two personality variables that, because they were measured contemporaneously with our risk-attitude variables, do not clearly belong on either side of such regressions: Behavioral Inhibition and the Locus of Control variables. As with the risk-attitude variables, we use our framework together with retest data for several hundred respondents to estimate pairwise correlations between each of our risk-attitude variables and the two personality variables; our analyses focus on the permanent components of the ordinal Rotter Locus of Control and Behavioral Inhibition variables, which we assume are continuously distributed latent variables.[29] The procedure employed to produce these estimates is described in more detail in the Online Appendix.[30]

Table 6 reports estimates of the pairwise correlations between each of the five risk variables and the two personality measures, both for a model that controls for measurement error and for a model without measurement-error adjustment. The correlations between the risk variables and Behavioral Inhibition are large and precisely estimated. The estimated correlations increase by about 50% after adjustment for measurement error; the strongest observed correlation after adjustment for measurement error is with Risk General and is 0.45, much higher than what has been reported for the "Big Five" and risk-taking (Becker et al. 2012; Dohmen et al. 2010; Lönnqvist et al. 2014). The high correlation between Behavioral Inhibition and risk attitudes has much intuitive appeal, because the behavioral inhibition system (Gray 1982) inhibits

---

[29]By contrast, previous works on the relationship between risk attitudes and personality measures often use the ordinal variables directly in analyses (e.g., Becker et al. 2012). Our approach has the advantage of not treating the ordinal Rotter Locus of Control and Behavioral Inhibition variables as having cardinal significance.

[30]To make the estimation computationally manageable, we collapsed the Risk General, Risk Finance, Behavioral Inhibition, and Locus of Control variables to 6, 7, 11, and 12 categories, respectively.

behavior that may lead to painful outcomes, punishment, non-reward, and novelty (Carver and White 1994). The measurement-error-adjusted correlations between Rotter Locus of Control and the risk variables are lower (they range from 0.12 to 0.21) but are all highly significant.

8.2 Behavior genetic decomposition of risk attitudes

A voluminous literature in economics attempts to measure and interpret sibling correlations in skills, preferences, and economic outcomes (Carver and White 1994). Sibling correlations provide a crude omnibus measure of the extent to which family background, broadly construed, can account for variation in economic behaviors and outcomes. Because the SALTY respondents are twins, we use our data to obtain measurement-error-adjusted estimates of the DZ and MZ sibling correlations (r_{DZ} and r_{MZ}) for the risk variables. Under some assumptions that we describe below, these correlations can be used to obtain rough estimates of the fraction of variance accounted for by genetic factors and by non-genetic factors that siblings share.[31]

The standard variance decomposition that most behavior genetic studies use is known as the ACE model. The ACE model decomposes outcome variances into an additive genetic factor (A^*), common environment (C^*), and individual environment (E^*).[32] The model requires a number of identifying assumptions, and its estimates are subject to a number of interpretational caveats; for a discussion, see Beauchamp et al. (2011). Let y denote the trait of interest, standardized to have unit variance; let A^*, C^*, and E^* denote the latent additive genetic, common environment, and individual environment factors, respectively; and let

$$y = A^* + C^* + E^*.$$

We assume the variance components are mutually independent, that DZ twins share their common environment to the same degree as MZ twins (the "equal-environment assumption"), that all genetic effects are linear and additive, and that no assortative mating occurs. Under these assumptions, the covariance of the additive genetic factor in DZ twins is 0.5, and the two moments $r_{DZ} = \frac{1}{2}\sigma_{A^*}^2 + \sigma_{C^*}^2$ and $r_{MZ} = \sigma_{A^*}^2 + \sigma_{C^*}^2$ can be used to estimate the variance components. Besides the conventional ACE models, we also report estimates of the ADE model whenever the unconstrained maximum likelihood estimates of the $\sigma_{C^*}^2$ is negative.[33,34] In an ADE model, the variance of C^* is restricted to equal zero and a dominant genetic factor (D^*) is included to capture effects that are not linear in the number of alleles (the alternative forms of the DNA sequences at a specific location in the genome). The Online Appendix provides additional details on the four factors and on our empirical framework.

---

[31] Kimball et al. (2009) is the only paper of which we are aware that attempts to disattenuate parent-child and sibling correlations in risk-taking while adjusting for measurement error.

[32] More sophisticated models exist that account for sex effects and higher-order genetic effects and that distinguish between common family environment and sibling-specific environment, among others.

[33] $\sigma_{C^*}^2$ is negative when $\rho_{MZ} > 2 \cdot \rho_{DZ}$; this may indicate the dominant genetic factor is important.

[34] Our standard errors do not take the model selection uncertainty into account, nor are they adjusted for the constraint that the variance components cannot be negative, which is imposed in the estimation.

Table 7 Twin correlations and behavior genetic variance decomposition

	Risk General		Risk Financial		Risk HRS	
	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$	$\sigma_m^2 = 0$	$\sigma_m^2 > 0$
$corr_{MZ}$	0.354***	0.513***	0.305***	0.433***	0.424***	0.698***
s.e.	(0.031)	(0.053)	(0.031)	(0.049)	(0.036)	(0.074)
# pairs	1,128	1,128	1,134	1,134	1,000	1,000
$corr_{DZ}$	0.141***	0.216***	0.176***	0.249***	0.192***	0.311***
s.e.	(0.028)	(0.042)	(0.030)	(0.045)	(0.037)	(0.064)
# pairs	1,200	1,200	1,212	1,212	1,083	1,083
$\hat{\sigma}_{A*}^2$	0.212*	0.353**	0.260***	0.369***	0.344**	0.545**
	(0.118)	(0.169)	(0.087)	(0.128)	(0.154)	(0.251)
$\hat{\sigma}_{C*}^2$	–	–	0.046	0.065	–	–
	–	–	(0.068)	(0.099)	–	–
$\hat{\sigma}_{D*}^2$	0.143	0.160	–	–	0.080	0.152
	(0.130)	(0.185)	–	–	(0.166)	(0.267)
$\hat{\sigma}_{E*}^2$	0.646***	0.487***	0.695***	0.567***	0.576***	0.302***
	(0.031)	(0.053)	(0.031)	(0.049)	(0.036)	(0.074)

NOTES: The estimates in the first column under each variable name are not adjusted for measurement error ($\sigma_m^2 = 0$); the estimates in the second column are adjusted ($\sigma_m^2 > 0$). The lower panel shows the implied estimates for the variance components in the ACE model, whenever they are all positive; if the estimate of one of the variance components is negative, we report the estimates for the variance components in the ADE model instead

* Significant at 10% level; ** significant at 5% level; *** significant at 1% level

Table 7 reports the results for the risk variables for the specification with sex and birth year partialled out for the models with and without measurement-error adjustment. Taken together, our measurement-error-adjusted estimates of the narrow-sense heritability (i.e., the share of the variance explained by additive genetic factors, σ_{A*}^2) of the permanent component of risk attitudes range from 35% to 55%, almost doubling previous estimates in the literature (Cesarini et al. 2009), and much higher than the unadjusted estimates (which range from 21% to 34%). These estimates are within the range of the consensus estimates in the literature on the heritability of the "Big Five" factors of personality and a bit lower than the consensus estimates for intelligence (Bouchard and McGue 2003). Indeed, we conjecture that once measurement error is controlled for, the heritability of most economic attitudes will approach that of the "Big Five" in personality research.

The results reported here suggest the heritability of economic preferences may be considerably higher than that previously implied by the emerging literature that uses behavioral genetic techniques to study the genetic and environmental sources of variation in economic risk preferences. As a result, preference-based channels of intergenerational transmission of economic status may be quantitatively more important than presently believed. The emerging literature on the empirical properties of measured risk attitudes is thus reminiscent of the literature on the intergenerational

transmission of inequality, where the first papers reported modest estimates of the relationships between the economic status of parents and their children (e.g., Becker and Tomes, 1986), which subsequent papers revised upward as more econometrically sophisticated tools were developed for the measurement of economic status and as the quality of available data improved (Mazumder 2005; Solon 1992).

9 Conclusion

We examined a large, population-based sample of 11,000 twins with data on risk attitudes and important behavioral outcomes and made several contributions to the effort to learn more about some measures of risk preferences that are commonly used in economic research.

Using retest data for 500 respondents, we demonstrated that accounting for measurement error (or other transitory fluctuations) leads to substantially larger estimates of the risk measures' predictive power and correlations with other variables. For instance, correcting for measurement error raises the estimate of the effect of a one-standard-deviation increase in the Risk HRS variable on the probability of having started a business from 9.1% to 14.5%. More generally, it considerably increases the incremental R^2 of measures of risk attitudes in regressions of risky behaviors, the predictive power of various covariates on risk attitudes, as well as the magnitude of the estimates of sibling correlations.

We also report sizable estimates of the measured risk attitudes' predictive power for risky behaviors and of their associations with other variables. We find that our preferred measures of risk attitudes are strongly and consistently associated with retirement investment decisions, the equity share of assets, and the decision to run one's own business. Adding the Risk Financial or the Risk HRS measures of risk preferences to a rich set of covariates in regressions of Portfolio Risk and Equity Share, and accounting appropriately for measurement error, can more than double the R^2 (though from a low baseline). The risk variables also predict health-related behaviors such as consuming alcohol and smoking, though the patterns appear to be less robust. Further, we report large sex differences in risk attitudes and we find that an IQ measure taken four decades before the survey is a strong predictor of risk attitudes. We document a novel relationship between the personality-trait behavioral inhibition and risk attitudes; that relationship is considerably stronger than the relationships between risk attitudes and the "Big Five" personality traits previously reported in the literature.

Although a number of papers attempt to explicitly model and adjust for measurement error (Andersen et al. 2008; Harrison et al. 2005; Kimball et al. 2008; Sahm 2012), most empirical work still ignores this important issue. Our results underscore the importance of adjusting for measurement error in a number of settings and have implications for the design and interpretation of empirical research seeking to account for heterogeneity in risk aversion. For example, risk aversion is often a confound when testing auction theory, and researchers often include a measure of risk aversion similar to the ones considered here as a control variable when testing the theory (see the review in Kagel and Levin, 1995). Our results imply risk aversion can remain a large confound if researchers fail to properly adjust for measurement error.

This paper speaks to an important methodological debate about the appropriate role of preference-based explanations in economics and suggests the relatively low fraction of variation in risky behaviors explained by risk variables in previous work (e.g., Barsky et al. 1997) is in part a consequence of measurement error. Accordingly, caution is warranted when interpreting results in which survey-based measures of economic preferences explain only a tiny share of variation in risky behaviors but where measurement error is not adjusted for: these results do not necessarily imply preference heterogeneity can be safely ignored. Our findings thus reinforce Kimball et al.'s (2008) conclusion that carefully controlling for preference heterogeneity in empirical work is important.

Our results are also relevant for ongoing efforts to integrate individual-difference psychology with economics (Ferguson et al. 2011; Almlund et al. 2011). We see this paper as but a first step to introduce tools and concepts from the field of psychometrics in economics. Psychometricians have long thought about how to design optimal survey instruments to measure psychological constructs and about how to develop and refine the theoretical approaches underlying the measurements, and we believe economists have much to learn from them as they attempt to develop a better understanding of the properties of measures of economic preferences. In addition, our results are relevant for the question of how the primitive constructs in economics are empirically related to those of individual-difference psychology. Work to date has found these relationships to be fairly weak (Becker et al. 2012) and has consequently concluded economic preferences should be considered complements to personality measures; however, our measurement-error-adjusted estimates of the associations between risk attitudes and cognitive ability and behavioral inhibition suggest otherwise, and research on the links between preferences and personality could still produce interesting insights.

Overall, our paper contributes to the emerging literature attempting to measure whether, how, and why risk preferences differ across people, and how observable choices and outcomes correlate with those risk preferences. Designing more reliable and predictive risk measures will be a very productive area for future research.

Appendix I - Derivation of the likelihood function of the latent variable model

To derive the likelihood function for our latent variable model, let $\Theta = (\tau_1, ..., \tau_C, \beta', \sigma_{\varsigma^*})'$ be the vector of parameters to be estimated, let $\tau_o = -\infty$ and $\tau_{C+1} = +\infty$, and let $\mathbf{r} = (r_1, r_2)'$, and consider first the distribution function of $\mathbf{r}$

for any respondent, conditional on ς^*:

$$f(\mathbf{r}|\varsigma^*, \Theta, \mathbf{x}) = \prod_{c=0}^{c=C} p_c(\varsigma^*, \Theta, \mathbf{x})^{1\{r_1=c\}+1\{r_2=c\}} \quad \text{(for } r_1, r_2 \in \{0...C\}),$$

$$p_c(\varsigma^*, \Theta, \mathbf{x}) = \Phi(\tau_{c+1} - \mathbf{x}\beta - \varsigma^*) - \Phi(\tau_c - \mathbf{x}\beta - \varsigma^*),$$

where $\Phi(\cdot)$ is the cumulative normal distribution function.[35] The joint distribution function of $\mathbf{r}$ and ς^* for the respondent is given by

$$f(\mathbf{r}, \varsigma^*|\Theta, \mathbf{x}) = f(\mathbf{r}|\varsigma^*, \Theta, \mathbf{x}) \cdot \phi\left(\frac{\varsigma^*}{\sigma_{\varsigma^*}}\right),$$

where $\phi(\cdot)$ denotes the standard normal distribution function and $\phi\left(\frac{\varsigma^*}{\sigma_{\varsigma^*}}\right)$ is the marginal distribution function of ς^*. The marginal distribution function of $\mathbf{r}$ for the respondent is thus:

$$f(\mathbf{r}|\Theta, \mathbf{x}) = \int_{\mathbb{R}} f(\mathbf{r}, \varsigma^*|\Theta, \mathbf{x})d\varsigma^*,$$

where $\mathbb{R}$ denotes the real numbers. Finally, we obtain the likelihood function for all the respondents:

$$L(\Theta|\mathbf{R},\mathbf{X}) = f(\mathbf{R}|\Theta, \mathbf{X}) = \prod_{i=1}^{N} f(\mathbf{r}_i|\Theta, \mathbf{x}_i),$$

where $\mathbf{R} = (\mathbf{r}_1, ..., \mathbf{r}_N)$ and $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_N)$.

References

Almlund, M., Duckworth, A.L., Heckman, J.J., & Kautz, T. (2011). Personality psychology and economics. In Hanushek, E., Machin, S., & Woessman, L. (Eds.) Handbook of the Economics of Education (Vol. 4, Chapter 1 1–181). Elsevier B.V.

Andersen, S., Harrison, G.W., Lau, M.I., & Rutström, E. (2008). Lost in state space: Are preferences stable? *International Economic Review*, *49*, 1091–1112.

Andersson, O., Tyran, J.-R., Wengström, E., & Holm, H.J. (2013). Risk aversion relates to cognitive ability: Fact or fiction? Working Paper Department of Economics, Lund University, No 2013:9.

Barsky, R.B., Juster, F.T., Kimball, M.S., & Shapiro, M.D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *Quarterly Journal of Economics*, *112*(2), 537–579.

Bartholomew, D.J., & Knott, M. (1999). *Latent Variable Models and Factor Analysis* (Second ed.). Wiley.

Beauchamp, J., Cesarini, D., Johannesson, M., van der Loos, M.J.H.M., Koellinger, P.D., Groenen, P.J.F., Fowler, J.H., Rosenquist, J.N., Thurik, A.R., & Christakis, N.A. (2011). Molecular genetics and economics. *Journal of Economic Perspectives*, *25*(4), 57–82.

Becker, A., Deckers, T., Dohmen, T., Falk, A., & Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annual Review of Economics*, *4*, 453–478.

Becker, G.S., & Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics*, *4*(3), S1–S39.

Benjamin, D.J., Brown, S.A., & Shapiro, J.M. (2013). Who is 'Behavioral'? Cognitive ability and anomalous preferences. *Journal of the European Economic Association*, *11*(6), 1231–1255.

---

[35]Retest data is only available for some respondents. We ignore this subtlety in the foregoing derivations.

Bonin, H., Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics*, *14*(6), 926–937.

Borghans, L., Duckworth, A.L., Heckman, J.J., & ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972–1059.

Bouchard, T.J., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology*, *54*(1), 4–45.

Burks, S.V., Carpenter, J.P., Goette, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, *106*(19), 7745–7750.

Camerer, C.F., & Hogarth, R.M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*(1-3), 7–42.

Carlstedt, B. (2000). Cognitive abilities - aspects of structure, process and measurement. Ph.D. thesis, University of Gothenburg.

Carver, C.S., & White, T.L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The bis/bas scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333.

Cesarini, D., Dawes, C.T., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Genetic variation in preferences for giving and risk taking. *Quarterly Journal of Economics*, *124*(2), 809–842.

Cesarini, D., Johannesson, M., Lichtenstein, P., Sandewall, Ö., & Wallace, B. (2010). Genetic variation in financial decision making. *Journal of Finance*, *65*(5), 1725–1754.

Choi, S., Kariv, S., Müller, W., & Silverman, D. (2014). Who is (more) rational? *American Economic Review*, *104*(6), 1518–1550.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–474.

Dave, C., Eckel, C.C., Johnson, C.A., & Rojas, C. (2010). Eliciting risk preferences: When is simple better? *American Economic Review*, *41*(3), 219–243.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes. *Review of Economic Studies*, *79*, 645–677.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G.G. (2011). Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550.

Dohmen, T., Falk, A.F., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, *100*(3), 1238–1260.

Einav, L., Finkelstein, A., Pascu, I., & Cullen, M.R. (2012). How general are risk preferences? Choices under uncertainty in different domains. *American Economic Review*, *102*(6), 2606–2638.

Ferguson, E., Heckman, J.J., & Corr, P.J. (2011). Personality and economics: Overview and proposed framework. *Personality and Individual Differences (Special Issue: Personality and Economics)*, *51*(3), 201–209.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Friedman, M.A. (1957). *Theory of the consumption function*. Princeton University Press.

Gillen, B., Snowberg, E., & Yariv, L. (2015). Experimenting with measurement error: Techniques with applications to the Caltech cohort study. Working paper, California Institute of Technology.

Gladstone, G.L., & Parker, G.B. (2005). Measuring a behaviorally inhibited temperament style: Development and initial validation of new self-report measures. *Psychiatry Research*, *135*, 133–143.

Gray, J.A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. Clarendon Press/Oxford University Press.

Grinblatt, M., Keloharju, M., & Linnainmaa, J. (2011). IQ and stock market participation. *Journal of Finance*, *66*(6), 2121–2164.

Hagle, T.M., & Mitchell, G.E. (1992). Goodness-of-fit measures for probit and logit. *American Journal of Political Science*, *36*(3), 762–784.

Hanoch, Y., Johnson, J.G., & Wilke, A. (2006). Domain specificity in experimental measures and participant recruitment: An application to risk-taking behavior. *Psychological Science*, *17*(4), 300–304.

Harrison, G.W., Johnson, E., McInnes, M.M., & Rutström, E.E. (2005). Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters*, *1*, 31–35.

Harrison, G.W., Lau, M.I., & Rutström, E.E. (2007). Estimating risk attitudes in Denmark: A field experiment. *Scandinavian Journal of Economics*, *109*(2), 341–368.

Hey, J.D., & Orme, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, *62*(6), 1291–1326.

Holt, C.A., & Laury, S.K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Holt, C.A., & Laury, S.K. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, *95*(3), 902–904.

Horowitz, J.K. (1992). A test of intertemporal consistency. *Journal of Economic Behavior and Organization*, *17*, 171–182.

Josef, A.K., Richter, D., Samanez-Larkin, G.R., Wagner, G.G., Hertwig, R., & Mata, R. (2016). Stability and change in risk-taking propensity across the adult life span. *Journal of Personality and Social Psychology*, *111*(3), 430–450.

Kagel, J., & Levin, D. (1995). Auctions: A survey of experimental research. In Kagel, J.H., & Roth, A.E. (Eds.) Handbook of Experimental Economics (pp. 501–587). Princeton University Press.

Kimball, M.S., Sahm, C.R., & Shapiro, M.D. (2008). Imputing risk tolerance from survey responses. *Journal of the American Statistical Association*, *103*, 1028–1038.

Kimball, M.S., Sahm, C.R., & Shapiro, M.D. (2009). Risk preferences in the PSID: Individual imputations and family covariation. *American Economic Review Papers and Proceedings*, *99*(2), 363–368.

Lichtenstein, P., Sullivan, P.F., Cnattingius, S., Gatz, M., Johansson, S., Carlström, E., Björk, C., Svartengren, M., Volk, A., Klareskog, L., de Faire, U., Schalling, M., Palmgren, J., & Pedersen, N.L. (2006). The Swedish twin registry in the third millennium: An update. *Twin Research and Human Genetics*, *9*(6), 875–882.

Lönnqvist, J.-E., Verkasalo, M., Walkowitz, G., & Wichhardt, P.C. (2014). Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. Working Paper 1905, Kiel Institute for the World Economy.

Lykken, D.T., McGue, M.K., & Tellegen, A. (1987). Recruitment bias in twin research: The rule of two-thirds reconsidered. *Behavior Genetics*, *17*(4), 343–362.

Magnusson, P., Almqvist, C., Rahman, I., Ganna, A., Viktorin, A., Walum, H., Halldner, L., Lundström, S., Ullen, F., Langström, N., Larsson, H., Nyman, A., Gumpert, C.H., Rastam, M., Anckarsäter, H., Cnattingius, S., Johannesson, M., Ingelsson, E., Klareskog, L., de Faire, U., Pedersen, N., & Lichtenstein, P. (2013). The Swedish twin registry: Establishment and other recent developments. *Twin Research and Human Genetics*, *16*(1), 317–329.

Malmendier, U., & Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics*, *126*(1), 373–416.

Mazumder, B. (2005). The apple falls even closer to the tree than we thought: New and revised estimates of the intergenerational transmission of earnings. In Bowles, S., Gintis, H., & Groves, M.O. (Eds.) Unequal chances: Family background and economic success. Princeton University Press.

McArdle, J.J., & Prescott, C.A. (1992). Age based construct validation using structural equation modeling. *Experimental Aging Research*, *18*(3), 87–115.

McArdle, J.J., & Woodcock, R.W. (1997). Expanding test-retest designs to include developmental timelag components. *Psychological Methods*, *2*(4), 403–435.

McKelvey, R.D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, *4*(1), 103–120.

Meier, S., & Sprenger, C.D. (2010). Temporal stability of time preferences. *The Review of Economics and Statistics*, *97*(2), 273–286.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika*, *69*, 167–190.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, *68*(5), 1281–1292.

Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General & Applied*, *80*(1), 1–28.

Sahm, C.R. (2012). How much does risk tolerance change? *Quarterly Journal of Finance*, *2*(4), 1250020.

Skrondal, A., & Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society*, *172*(3), 659–687.

Solon, G. (1992). Intergenerational income mobility in the United States. *The American Economic Review*, *82*(3), 393–408.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Stigler, G.J., & Becker, G. (1977). De gustibus non est disputandum. *American Economic Review*, *67*(2), 76–90.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

von Gaudecker, H.-M., van Soest, A., & Wengström, E. (2011). Heterogeneity in risky choice behaviour in a broad population. *American Economic Review*, *101*(2), 1–33.

Weber, E.U., Blais, A.-R., & Betz, N.E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263–290.