Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Performance of Time-Preference and Risk-Preference Measures in Surveys

Joshua Tasoff, Wenjie Zhang

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–25
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

The Performance of Time-Preference and Risk-Preference Measures in Surveys

Joshua Tasoff,[a] Wenjie Zhang[a]

[a] Department of Economics, Claremont Graduate University, Claremont, California 91711

Contact: joshua.tasoff@cgu.edu, https://orcid.org/0000-0003-0956-8359 (JT); wenjiezhang01@gmail.com (WZ)

Received: October 2, 2019
Revised: April 3, 2020; October 19, 2020
Accepted: November 8, 2020
Published Online in Articles in Advance: March 11, 2021

https://doi.org/10.1287/mnsc.2020.3939

Abstract. Time preferences and risk preferences play an important role in a wide range of behavior, including financial decisions, entrepreneurship, and the proper incentivizing of agents. Numerous methods have been developed to measure these preferences hypothetically in surveys, but they have yielded inconsistent results. We analyze a panel data set in which subjects have collectively answered more than 400 surveys including 15 time-preference and 36 risk-preference elicitations. We evaluate the performance of these measures using the criteria of (1) ability to predict economically important behavior and (2) distinctness from other observables. We find substantial heterogeneity in the predictiveness of the measures. The best performing measure for time-preference is a titration method, in which a sequence of adaptive binary-choice questions narrows in on a subject's indifference point, and for risk-preference it is a self-report measure of risk aversion. Using factor analysis, we find that time preferences are well explained by a single factor, but risk preferences load on multiple factors. However, the first factor loads almost entirely on self-reported risk-preference measures, and this factor explains much of the variation. The evidence can help inform researchers about which elicitation methods to include in their surveys.

History: Accepted by Yan Chen, decision analysis.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2020.3939.

Keywords: time preferences • risk preferences • surveys • elicitation method

1. Introduction

According to economic theory, time preferences and risk preferences play a profound role in nearly all aspects of decision making. They determine one's financial behavior such as one's saving rate, asset allocation, insurance decisions, human capital investment, career choice, and one's health behaviors such as smoking, exercise, and overeating. In the domain of business, time and risk preferences affect entrepreneurship and they affect managerial decisions including capital expenditure and strategic operations. Proper measurement of these latent parameters is important for testing economic theory and for structuring incentives. In principle, a better understanding of a population's time and risk preferences could be used to inform the design of compensation packages to more efficiently align the behavior of employees and managers with firm interests. Well-designed compensation packages that leverage information about the population's time and risk preferences may also be more effective at attracting talent.

Consequently, researchers have innovated a myriad of methods to measure these latent preferences. However, stark inconsistencies have been observed in cross-method comparisons. Subject rank is not preserved in risk-aversion measures across different elicitation procedures (Berg et al. 2005, Crosetto and Filippin 2016). Weber et al. (2002) find that risk preferences appear to vary widely based on the domain. Different elicitation methods have yielded different population distributions of both latent risk and time preference parameters.[1] Test-retest reliabilities have ranged the gamut with very low correlations ($r = 0$) to very high ($r = 0.75$) depending on sample and method (Chuang and Schechter 2015). The performance of these measures in predicting behavior has also varied greatly, with some papers finding no associations with behaviors that theory predicts should be associated (Chabris et al. 2008) and other papers demonstrating robust associations (Goda et al. 2019).

An economist who needs to design a survey may wonder which method should be used to elicit time and risk preferences. In this paper, we evaluate common time- and risk-preference elicitations used in surveys applying the criteria of predictability and distinctness.[2] We bring considerable data to bear on this question. The American Life Panel contains several thousand individuals who are representative of the U.S. population. Participants have collectively participated in more than 400 surveys. The surveys are designed by third-party researchers—many of whom are economists—to answer a wide range of questions. Dozens of time- and risk-preference elicitations, using diverse methods, have been conducted on this population. Moreover, there is an impressive depth of information available on some participants including health behaviors, financial behaviors, and measures of cognitive ability, in addition to standard background characteristics (i.e., income, race, gender, education, marital status, and job status).

With these data, we compare the predictiveness of 15 time-preference measures (4 of which are for real stakes) spanning 8 methods, and 36 risk-preference measures (2 of which are for real stakes) spanning 16 methods, on 19 real-world financial and health outcomes. Our comparison controls for many individual characteristics that are often neglected, such as cognitive ability and financial literacy. This is arguably the most comprehensive *horse race* between methods yet conducted.

Additionally, a good measure of a latent characteristic should be distinct from other characteristics. For example, a good measure of cognitive ability should be distinct from education. It is possible that measures of time preference and risk preference are highly correlated with each other. If true, it would reduce the empirical value of measuring both of them. If they are both highly correlated with other easily observed individual characteristics, it may be more efficient to use those characteristics as proxies rather than measuring either time preferences or risk preferences. We test the distinctness of time-preference and risk-preference measures using exploratory factor analysis. Our data set has numerous characteristics for each individual, including multiple cognitive measures. Specifically, we test whether a single factor emerges that loads primarily on time-preference measures with little loading on other characteristics, and we do likewise for risk preferences. If so, this would indicate that (1) the measures that load onto the factor are measuring the same latent characteristic and that (2) this characteristic is distinct from the characteristics associated with the other factors. As far as we are aware, this is the first test for the distinctness of time- and risk-preference measures from each other and other commonly observed individual characteristics.

The analysis yields several important findings. First, we observe heterogeneity in the performance of time-preference and risk-preference measures in predicting the 19 financial and health outcomes. Some measures are never statistically significant while others are statistically significant for numerous outcomes. Second, we find that time-preference measures based on titration methods, in which subjects are asked a sequence of binary choice questions to narrow in on an indifference point, tend to be strong predictors of behavior. Third, we find only mixed evidence that supports the predictiveness of any economic risk-aversion measures, including titration-based measures. In contrast, measures based on self-reports predict behavior quite effectively. This is consistent with the findings of Dohmen et al. (2011), who find that simply asking people how willing they are to take risks in general best predicts choice over lotteries.

Our analysis on the measures' distinctness yields additional findings. Fourth, exploratory factor analysis shows that the titration time-preference measures load together heavily on a single factor with lower loadings coming from a time-preference multiple-price list (MPL) measure and Fair Isaac Corporation score (FICO). This is consistent with the claim that titration time-preference measures consistently reflect the same latent characteristic, and this characteristic is distinct from risk preferences, cognitive ability, and income. Collectively, other time-preference measures do not heavily load on any of the first five factors. Fifth, risk-preference measures from self-reports heavily load on a single factor that does not load heavily on anything else. The evidence suggests that the risk-preference self-reports are measuring the same latent characteristic, and this characteristic is distinct from time preferences, FICO, cognitive ability, income, and education and whatever the economic risk-aversion elicitations are measuring. Finally, as an additional test of the independence of time and risk preferences from other individual characteristics, we regress economic outcomes on the preference measures while systematically omitting control sets. As we omit control sets, the coefficients remain remarkably unaltered. This suggests that the effect of the preference measures on economic outcomes is mostly unmediated by the control variables. It also reveals that omitted-variable bias is not a major concern for the control sets that we test.

Although we believe the data and analysis have considerable strengths, with more measures and outcomes than any previous study, the results deserve a cautious interpretation because there are several important limitations. First, most of the elicitations ran without incentives and all surveys were conducted online. Laboratory elicitations with incentives and a more controlled environment may lead to different results. Second, specific implementation details (i.e., the instructions, questions, alternatives) varied across the surveys. For example, one elicitation may have asked people whether they prefer \$100 now or \$110 a year from now, whereas another elicitation may have asked people whether they prefer \$500 today or \$525 in a month. This variation in implementation adds noise to the data that may lead to overstating the number of factors in our factor analysis. Third, these specific implementation details may have a large effect on the measures' predictability, and this should be taken into consideration when assessing external validity. As a specific example, the convex time-budget (CTB) method uses 45 questions to estimate time-preference measures according to the original implementation (Andreoni and Sprenger 2012), yet the measures in our data set used only 12 questions. Fourth, the sample is not identical across all surveys. Differences in sample may partially affect estimated coefficients. Fifth, there may be other elicitation methods of interest that are missing in this data set.

These limitations result from the fact that the data are not experimental but observational. Each individual survey may have been conducted in a controlled manner but there was no coordination over the 400+ surveys that make up the American Life Panel (ALP) database. However, each survey was designed and fielded by a researcher who invested considerable resources in time and money to field it. These are the elicitations that economists chose to use in their surveys because they thought it would produce useful data. In a way, economists have thus voted with their time and money to determine which methods are subjected to analysis in our study.

2. Background and Related Literature

2.1. Comparing Measures for Prediction

We are aware of two studies that have compared the predictiveness of different time-preference measures. Burks et al. (2012) elicit three different measures of time preference in a large sample of truck-driver trainees and combine it with administrative data. They first elicit preferences with multiple MPLs, overidentifying β and δ, and then use ordinary least squares (OLS) to find the best-fitting risk parameters. Their second measure, meant to be a measure of impulsivity, elicits the participants with a big red button. Each click of the button shortened a mandatory waiting period but reduced earnings. The third measure was a set of self-reports addressing tendency to procrastinate, conscientiousness, restlessness, and similar attributes. They regress their outcomes, smoking, credit score, and body mass index (BMI), on the three measures and find that the MPL measure better predicts than the impulsivity measure and the self-reports. Andreoni et al. (2015) compare the CTB to the double multiple-price list (DMPL) in the laboratory with a college sample. Their subjects undergo both elicitations in addition to a third task in which they elicit the value of \$25 received at a later date. They find that the CTB measure predicts as well out-of-sample in the DMPL data as the DMPL does in-sample. The CTB measure predicts better than the DMPL in the valuation of the delayed \$25. They conclude that the CTB performs better than the DMPL. Our study greatly expands the research in this area by including many more measures (15 in total), several uncompared elicitation methods, such as titration, real effort, and lifetime—in addition to the MPL, CTB, and self-reports—and uses many more real-world outcomes (19 in total compared with 6 in the two studies combined). Furthermore, lack of adequate controls may be a concern in any prediction task. We have a more comprehensive control set than either study. Burks et al. (2012) has good controls for cognitive ability but has only coarser socio-economic controls. Andreoni et al. (2015) do not use any controls.[3]

The risk-preference literature is relatively larger. There are now several review papers (Charness et al. 2013, Mata et al. 2018, Hertwig et al. 2019) and many original studies that compare different risk measures to each other (Berg et al. 2005, Dave et al. 2010, Crosetto and Filippin 2016, Pedroni et al. 2017). Many of these papers examine between-measure correlation and examine the measures' temporal stabilities. However, we are aware of only four papers that compare the predictiveness of different risk measures for economically meaningful outcomes. Dohmen et al. (2011) collect data from a large-scale survey in which participants self-reported their risk tolerance generally and in the specific domains of driving, finance, sports/leisure, career, and health. They found that the general measure predicts choice in an MPL risk elicitation for real stakes. They also find that the measures predict investment in stocks, activity in sports, self-employment, and smoking (with the exception that the sports measure was not predictive of self-employment or smoking and the financial measure was not predictive of smoking). The best predictor was the risk-taking in general question.

Galizzi et al. (2016) use the general, finance, and health self-reports from Dohmen et al. (2011), in addition to the Holt-Laury measure (MPL) and Eckel and Grossman (2008) measure. They use the measures as regressors to predict smoking, BMI, junk food consumption, fruit and vegetable consumption, heavy drinking, savings, and having a personal pension fund in a representative sample of UK households. None of the measures reliably predict smoking status, consumption of junk food, likelihood of regularly saving, or the time horizons of savings. The two behavior measures predict BMI, Holt-Laury predicts the regular consumption of fruit and vegetables, and Eckel-Grossman predict having a private pension fund. The general self-report predicts savings, and the health self-report predicts heavy alcohol drinking. They make no judgments on which measure predicts best given that different measures predict different outcomes.

Beauchamp et al. (2017) elicit three risk-preference measures in a large survey of twins. They use self-reported general risk tolerance and self-reported risk tolerance in the domain of finance, along with a third measure that we refer to in this paper as a titration elicitation. They find that all three measures predict portfolio risk, equity share, and whether the participant owns a business. The self-report general measure and self-report financial measure both predict alcohol use. Only the general measure predicts smoking. They make no judgments on which measure predicts best as again they find that different measures better predict different outcomes.

Similarly, Andreoni and Kuhn (2019) compare four different real-stakes measures in the laboratory: the Holt-Laury task (MPL), certainty equivalent (MPL), the probability equivalent (MPL), and a modified CRB (mCRB). Using five outcomes—trust others, do you gamble, plan ahead, buys warranties, and self-reported risk tolerance—they find that only the Holt-Laury measure predicts whether one trusts others, and only the self-report predicts gambling. However, when they correct for measurement error, they find that all the measures are predictive. Like Galizzi et al. (2016) and Beauchamp et al. (2017), they make no judgments on which measure predicts best overall, because no one measure seems to strictly dominate the others across all outcomes.

With the exception of these four papers, there have been no studies that compare risk measures in terms of predictive validity. There is no other study that has our variety of elicitation methods: we horse-race 36 measures, which is double the 18 measures in the other four studies combined. There is no other study that has our breadth of control variables: only Beauchamp et al. (2017) controls for cognitive ability, a highly relevant omitted variable, and none of the studies control for financial literacy, which has been shown to be highly predictive of many financial outcomes (see Lusardi and Mitchell 2014 for a review). There is no other study that has our breadth of real-world outcome variables: 19 in total, which exceeds the existing 18 in the four other studies combined.

2.2. Distinctness and Common Factors

Time preference and risk preference are modeled by economists as disconnected attributes that are independent of each other. A person's patience need not have any mechanical relationship to a person's risk preference and vice versa. The canonical model of additively separable utility over time with quasi-hyperbolic discounting is presented here (Laibson 1997), in which β represents the short-run discount factor and δ represents the long-run discount factor, where time is indexed by t, $\vec{c}$ is the vector of consumption over the lifetime, and $u(\cdot)$ is the instantaneous utility function.

$$V_t(\vec{c}) \equiv u(c_t) + \beta \sum_{\tau=t+1}^{T} \delta^{\tau-t} u(c_\tau) \tag{1}$$

Risk preferences are represented by the curvature of the instantaneous utility function, with a concave transformation of $u(\cdot)$ representing more risk aversion. Here the values of β, δ, and the curvature of $u(\cdot)$ can all vary independently. However, because time preferences are typically elicited with "money earlier or later" questions, the inferred discount factor may not represent the person's true discount factor. Consider the choice between \$1 now at time t or \$$x$ later at time $t' > t$. If people consume their income upon receipt, indifference is represented by the Equation (2), where p_t represents the price of consumption at time t.

$$u\left(c_t + \frac{1}{p_t}\right) = \beta\delta u\left(c_{t'} + \frac{x}{p_{t'}}\right) \tag{2}$$

If utility is locally linear, for example if 1 and x are very small relative to consumption, then the equation can be approximated as

$$\frac{1}{p_t} u'(c_t) \approx \beta\delta \frac{x}{p_{t'}} u'(c_{t'}). \tag{3}$$

Researchers typically also assume prices and consumption are not changing over time, $p_t = p_{t'}$ and $c_t = c_{t'}$, allowing the discount factor to be calculated as

$$\beta\delta \approx \frac{1}{x}. \tag{4}$$

All these assumptions are tenuous at best. It defies both theory and common sense that people consume all income on receipt, in which case, Equation (2) has misspecified the relevant periods. According to expected-utility theory, people should be approximately linear for small monetary sums, but other theories disagree (e.g., prospect theory), and there is no assurance that this holds in practice. For small delays, inflation may be a relatively minor concern, but there is little reason to believe that consumption will be the same over time. From a theoretical standpoint, perfect consumption smoothing will generically not be the solution to an optimal lifecycle consumption problem. The degree of smoothing depends on the interest rate, the discount factors, and the curvature of the utility function. From an empirical standpoint, consumption can vary widely over the lifecycle, with a person consuming, for example, instant ramen during the college years to later indulge in glorified hipster ramen in midcareer.

These criticisms have two major implications for our analysis. First, the time preference measures are capturing an empirical discount rate that will only correspond to the theoretical discount rate under strong assumptions. Whether this empirical discount rate is predictive of behavior is an open question. Second, there is a mechanical relationship between risk-preferences and the time-preference measures. More concave utility functions are equivalent to more risk aversion, and in the lifecycle-consumption model, lead to more consumption smoothing. As illustrated by Equation (2), the degree of consumption smoothing affects the measured discount rate; that is, if an economist uses Equation (4) when consumption over time varies ($c_t \neq c_{t'}$), thereby producing different marginal utilities on the two sides of Equation (3), the economist will get a biased answer. This means that there could be an empirical association between the time-preference and risk-preference measures produced as an artifact of the elicitation methods.[4]

In addition to a mechanical relationship induced as an artifact of "money earlier or later" elicitations, there may be other reasons that the measures are related. Time-preference measures may be manifestations of risk preferences or other commonly measured attributes such as cognitive ability, education, or liquidity, and vice versa for risk preferences. This paper addresses the question of whether the measures are distinctly representing time preference and risk preference respectively and not other observables. We do so by running factor analysis on all measures simultaneously along with our control variables (demographics, cognitive ability, etc.). Distinctness implies that time-preference measures load onto a factor, risk-preference measures load onto a different factor, and these two factors do not have high loadings with the control variables.

A contemporaneous paper (Chapman et al. 2018) conducts a similar exercise. In a sample of 1,000 people, they elicit 21 behaviors including one time-preference measure and nine different measures related to risk and uncertainty and conduct principal component analysis. They find that risk measures load onto three different components. Their one time-preference measure does not load onto any component by more than 0.28, a component associated with punishing others in a trust game. Their approach is similar to ours, but our papers differ in terms of their application. Chapman et al. (2018) conducts their analysis across a wide number of measures meant to capture disparate aspects of preference (e.g., generosity, punishment, risk, ambiguity, overconfidence). In contrast, we conduct our analysis across a wide number of measures meant to capture the same preference, either time or risk. Our setting is suited to answer the question of whether the many measures are capturing the same latent preference and not how different preferences relate to each other.

The question of whether risk preferences are convergently valid, loading onto a single factor, has been explored before. Dohmen et al. (2011) show that 60% of the variation in self-reported risk attitudes across various domains can be explained by the first principal component. Frey et al. (2017) conduct factor analysis on self-reported risk attitudes, revealed-preference measures, and the frequency of risky behaviors. They find that a single factor adequately explains the variation in attitudes and risky behaviors but does not explain the variation in revealed-preference measures. Pedroni et al. (2017) elicit six risk-preference measures in a large sample: Holt-Laury (MPL), balloon analogue risk task (BART), a series of MPL tasks, a series of binary adaptive lotteries (titration), the Columbia card task, and the marbles task. The correlations between the tasks are fairly low, peaking at 0.34, with a median correlation of 0.15. In a review article, Mata et al. (2018) conclude that self-reported risk preferences tend to correlate well,k but behavior-based measures tend to have low convergent validity. Beauchamp et al. (2017) have a similar finding. They have two self-report measures (general and financial), and three behavioral measures (job titration, a gamble in the domain of gains, and a gamble in the domain of losses). They find the primary factor and correlate it with their five measures. The self-reported financial measure correlated very highly with the factor, and the self-reported general measure followed close behind; however, the three behavioral measures were considerably less correlated. Notably, no papers to date have applied exploratory factor analysis to multiple measures of time preference.[5] This is another contribution of this paper.

The work of Stango et al. (2017a) is adjacent to our research question but deserves mention here. They measure numerous behavioral biases and conduct exploratory factor analysis. Among the biases, they include present bias as measured by the CTB, present bias as indicated by a choice between snacks, and risky gambles in the domain of losses. They retain three factors, unfortunately none of which have clear interpretations.

3. Data

The American Life Panel is a representative sample of Americans, maintained by the RAND Corporation, used to run surveys primarily for research economists. As of the writing of this paper, more than 400 surveys have been conducted on subsamples of the panel. Although the data are publicly accessible, the data sets are not integrated. Researchers are free to design surveys how they choose. Aside from baseline characteristics, the same variable could be measured in different ways across the surveys. It is not unusual for researchers to incorporate a few variables from earlier surveys to bolster their analyses. However, with the exception of Hung et al. (2009), who study financial literacy, we are not aware of any papers that combine ALP data sets to look at broader patterns across the surveys.

3.1. Variables and Sample

At the time that we started this project, the American Life Panel had slightly more than 400 surveys. Each survey had approximately 200 questions on average. We went over all the variable descriptions, approximately 80,000 survey questions, and selected *all* questions that involved a time-preference elicitation, a risk-preference elicitation, or had a real-world outcome that is relevant to time or risk preference as suggested by economic theory.

We use 24 surveys covering 7,223 individuals. Very few subjects have taken all relevant surveys; hence, the overlap samples, which depend on which variables are needed for the analyses, are considerably smaller. Table A.2 in the Online Appendix shows the summary statistics of control and outcome variables. Table A.1 in the online appendix lists the papers for which the variables were created.

We use different samples depending on the analysis. The full sample includes all surveys with a time-preference elicitation or risk-preference elicitation that has $N > 150$. We include subjects who participate in at least one of the surveys. Sample size is 7,054 as reported in Table A.2. We dropped 169 individuals who were missing demographic information.

3.2. Methods of Elicitation

Table 1 lists all the measures of time preference we use in our analysis. We categorize time-preference elicitations into four kinds: MPL, titration, CTB, and lifetime. An MPL is a list of binary choices placed into two vertical columns: money earlier or later (Andersen et al., 2006). One of the columns incrementally improves or worsens, whereas the other column incrementally worsens or improves, respectively (it may also remain constant). The point at which one switches columns indicates the approximate indifference point. The titration method presents a binary choice between money earlier or later. The second binary choice is conditional on the first decision. The alternatives adjust upward or downward

Table 1. Methods of Elicitation for Time Preference

Methods	Surveys	Action	Time	Stakes	Start date	*N*
MPL	14	Pay	Now, 1 year later	210/250/300/400/600	11/6/2007	1,169
	248	Pay	Now, 1 year later	210/250/300/400/600	3/23/2012	4,705
	169	Receive	1 month, 7 month later	307.5/315.0/322.5/330/337/ 345/352.5/360/367.5/375	1/24/2011	756
Titration	11	Pay	Now, a year later	210/220/250/300	6/11/2007	1,108
	10	Receive	Now, 1 year later	100/105/110/115/120/125	5/7/2007	1,154
			10 year, 11 years	100/105/110/115/120/125	5/7/2007	1,154
	276	Receive lottery	Now, 1 year later	1100/1250/1650	8/15/2012	2,163
			1 year, 2 years later	1100/1250/1650	8/15/2012	2,163
	121	Receive check	Now, 10 days later	110%	3/30/2010	317
	125	Receive payment	Now, 10 days later	101/102/103/105/107/110	5/10/2010	105
			3 months, 3month+10 days	101/102/103/105/107/110	5/10/2010	105
	126	Receive payment	Now, 10 days later	101/102/103/105/107/110	5/19/2010	80
			3 months, 3month+10 days later	101/102/103/105/107/110	5/19/2010	80
	399	Receive payment	Now, 12 months later	See Goda et al. (2019)	10/17/2014	1,300
			12 months, 24 months later	See Goda et al. (2019)	10/17/2014	1,300
	189	Receive lottery	Now, 1 year later	1100/1250/1650	6/17/2011	2,999
			1 year, 2 years later	1100/1250/1650	6/17/2011	2,999
CTB	212	Divide $500	Now, 1 month later	500	8/25/2011	3,928
			1 year, 1 year+1 month later	500	8/25/2011	3,928
	263	Divide $500		500	12/25/2012	1,069
Real Effort	***263***	Complete surveys	5 days, 35 days later	15/18/21/24/27 minutes	3/6/2013	1,107
			90 days, 120 days later	15/18/21/24/27 minutes	3/6/2013	1,107
Lifetime	14	Divide spending	work vs. retirement	0.41/0.5/0.53/0.57/0.61/0.67	11/6/2007	1,171
			work vs. retirement	0.41/0.5/0.53/0.57/0.61/0.67	11/6/2007	1,171
	33	Divide spending	work vs. retirement	0.41/0.45/0.5/0.53/0.57/0.61	8/8/2008	877

Notes. Real-stake surveys are in bold: 169, 121, and 263. All the other surveys are hypothetical. Stake units for surveys 14, 33, and 121 are proportional. All the other surveys' units are in dollars. In the titration ball method, subjects choose from a list of two boxes with visually displayed balls of different colors, which represent risk. One box represents certain gains while the other represents uncertain gains.

to narrow in on the indifference point. The CTB method is an elicitation whereby the subject chooses a point on a downward sloping line in the money earlier versus money later dimensions (see Andreoni and Sprenger 2012 for a comprehensive description). Several of these CTBs are presented, and then time preferences are structurally estimated from the choices. Finally, the lifetime measure presents a hypothetical scenario in which the subject must choose how to spend their money over the lifecycle.

The second column states which ALP survey is the source. The third column states the mode. Most surveys are about when to *receive* money. However, there are three surveys, 11, 14, and 248, in which subjects decide when to *pay* money. The fourth column presents the time frames used. The fifth column presents the hypothetical stakes used. Surveys 169, 121, and 263 are for real monetary stakes. The start date of the surveys ranges from May 2007 to October 2014.

Table 2 lists all the risk preference we use in our analysis. We categorize risk preference elicitations into three kinds: titration, Eckel and Grossman, and MPL. Titration and MPL for risk preference are very similar to those for time preference. Instead of choosing between money earlier or later, subjects choose between lotteries. In our data, one of the two options is always a certain value. The Eckel and Grossman method has subjects choose one lottery from a menu of six lotteries. The lotteries exhibit a risk-return tradeoff. Lotteries with higher expected value are riskier. All surveys use hypothetical stakes except surveys 263 and 399. The start date of the surveys ranges from July 2007 to October 2014.

3.3. Identifying Time and Risk Preferences

For time preferences, we use the data to construct a measure of $\beta\delta$ using Equation (4) of the hyperbolic discounting model (Laibson 1997). This discount factor applies to now-versus-later decisions. The data are more complete for this comparison relative to comparisons between "money later or more money even later" decisions. As mentioned in Section 2.2, in order for this identification to be valid, several strong and arguably implausible assumptions are made: income is consumed at the time of receipt, consumption and prices do not change over time, and utility is locally linear.[6]

We are not aware of proposed solutions to the first three issues, but an alternative approach is offered by Andersen et al. (2008) that relaxes the linearity assumption. They structurally estimate the curvature of the utility function and baseline consumption level. Although there are advantages to this approach, we opt to use the ratio of the payments for several reasons. First, we cannot use a structural approach for every time-preference measure because the necessary data are not available. Andersen et al. (2008) use risk-elicitation questions to identify the curvature, but there are many surveys that conduct a time-preference elicitation that do not conduct a risk-preference elicitation.[7] Second, this approach necessarily confounds the quality of a time-preference measure with the quality of the risk-preference data used to construct it. We wish to compare time-preference elicitations on their own merit. Third, the quality of a structural approach depends heavily on the chosen functional form, which then requires a thorough investigation of different functional forms. This introduces an additional degree of freedom further complicating comparisons. Therefore, we opt for the simple and transparent measure of using the ratio of money now to money later that makes the person indifferent. The one exception is the CTB method, which has structural estimation of its parameters as an explicitly necessary step in the construction of the measure.[8] Another interpretation of our approach is that we are testing the performance of the empirical discount factor.

We standardize all the distributions and then perform our analysis on the standardized measures. The standardized distributions preserve rank and cardinality. According to theory, the elicited values should contain useful cardinal information about latent parameters and the standardization of the distributions retains this information. Cardinal differences between two values can be measured in terms of standard deviations. Point values of the discount factors are not preserved. Although recovering point values is of importance for informing theory and making predictions in new environments, we are less concerned with point values in this paper because they are not necessary for estimating predictiveness or distinctness.

For risk preferences, we take a slightly different approach. Typically, researchers apply a constant absolute risk aversion (CARA) or constant relative risk aversion utility function and estimate the corresponding coefficient of risk aversion. For comparability we want all measures in a single metric. Unfortunately, some elicitations ask about lotteries over gains in money and others ask about lotteries in terms of proportions of one's wealth. The first kind of elicitation identifies a CARA parameter, and the latter identifies a CRRA parameter. Without knowing a person's wealth, we cannot convert between the two, so transforming all measures into a consistent coefficient of risk aversion is not possible. Rather than calculate a coefficient, we simply put individuals into ordered categories. All the methods place individuals into a finite number of ordered bins. Our measure is simply which bin the person is placed in. We then standardize the measure.

Table 2. Methods of Elicitation for Risk Preference

Methods	Surveys	Content	Cetain payoff	Lottery	Start date	*N*
Titration	2	Job	Current level	Double vs. −10%, −20%, −33%, −50%, −75%	7/9/2007	1,652
	2	Job	Current level	+20% vs. −5%, −10%, −15%	7/9/2007	1,652
	167	Job	Current level	Double vs. −10%, −20%, −33%, −50%, −75%	3/2/2011	734
	197_h	Job	Current level	Double vs. −1%, −5%, −10%, −20%, −33%, −50%, −75%, −90%, −99%	10/21/2011	1,726
	197_a	Job	+50% increase	Double vs. +1%, +10%, +20%, +30%, +40%, +50%	10/21/2011	1,726
	197_i	Insurance	1/4 chance of −1/3 to buy insurance	1%, 5%, 10%, 15%, 20%, 25%, 30% (of income for insurance)	10/21/2011	1,726
	50	Invest for inheritance	1 million	Double vs. −10%, −20%, −33%, −50%, −75%	9/10/2009	2,283
	186	Invest for inheritance	1 million	Double vs. −10%, −20%, −33%, −50%, −75%	5/20/2011	423
	189	Invest for inheritance	1 million	Double vs. −10%, −20%, −33%, −50%, −75%	6/17/2011	3,061
	243	Ball	Series1: 0.3-0.7 40 vs. 10	0.1–0.9: 68/75/83/93/106/125/150/185/220/300/400/600/1000/1700 vs. 5	10/17/2014	3,290
			Series2: 0.9-0.1 40 vs. 30	0.7–0.3: 54/56/58/60/62/65/68/72/77/83/90/100/110/130 vs. 5	3/6/2013	3,290
	338	Ball	Same as survey 243	Same as survey 243	3/20/2012	235
	339	Ball	Same as survey 243	Same as survey 243	3/20/2012	229
	118	Retirement savings	Half the income	Current vs. 1/8, 1/4, 1/3, 2/5, 9/20	8/15/2013	2,920
Eckel and Grossman	***399***	Six lotteries		\$5–\$5, \$7–\$4, \$9–\$3, \$11–\$2, \$13–\$1, \$15–\$0	8/15/2013	1,877
	263	Six lotteries		\$28–\$28, \$36–\$24, \$44–\$20, \$52–\$16, \$60–\$12, \$70–\$2	4/30/2010	1,107
MPL	210	Roll a die	\$1,000	\$0 vs. \$1,600, \$2,000, \$2,400, \$2,800, \$3,200, \$3,600	8/9/2011	3,786
Self-report	12, 133, etc.	Generally	—	—	10/07/2007	—
	130, 167, etc.	Financial matters	—	—	8/27/2010	—
	130, 167, etc.	Daily activities	—	—	8/27/2010	—
	197, 179	Others	—	—	5/16/2011	—

Notes. Real stakes are used in surveys 263 and 399. All the other surveys are hypothetical. General self-rating risk elicitation is used in surveys 12, 48, 133, 169, 189, 197, and 284. Self-ratings on financial matters are used in surveys 130, 167, 179, 186, 193, 197, 260, 342, and 349. Self-rating on daily activities is used in surveys 130, 167, and 260. Self-rating on other topics is used in surveys 179 and 197.

3.4. Measuring Differences Between Subsamples

Because each subject responds to some but not all surveys, the sample is not consistent across all our analyses. A potential concern is selection bias: that variation in the performance of a measure reflects not the validity of elicitation method but variation in sample unobservables. In the online appendix, Figure C.1–C.5 display box plots of the distributions of income, age, education, and gender across the surveys. About 5 of 36 studies stand out as having lower average incomes, about 3 have noticeably younger age distributions, 5 have lower education distributions, and the ethnic makeup and gender distributions vary somewhat across the surveys. Nonetheless, the figures make apparent that there is much greater variation within than across samples, evident by the fact that the 95% confidence intervals are similar for nearly all the samples.

In Section 5, we vary the control sets to test the extent of omitted-variable bias when we omit available controls. The supposition is that if omitted-variable bias appears limited when omitting observables from the regression, then it is more likely that omitted-variable bias is limited when omitting unobservables from the regression.

4. Prediction

In this section, we attempt to address two questions. First, do the preference measures predict real-life outcomes in line with theory? Second, if they do, which measures are the strongest predictors. To address the first question, we regress numerous outcome variables on measures of time preference and risk preference controlling for a rich set of observables.

Given that there are 15 measures of time preference, 36 measures of risk preference, and 19 outcome variables, there are combinatorially many models that could be estimated. We display the results of *all* model regressions for each measure and each outcome in Online Appendix C. Table C.1 contains the results of the regressions of the 19 outcome variables on the 15 time-preference measures, and Table C.2 contains the regressions of 19 outcome variables on 36 risk-preference measures. Each cell in the table reports the coefficient, robust standard error, p value, and N. Because we test multiple hypotheses on a large scale in this analysis, we also report false-discovery rate (FDR) q-values, which are interpreted as p values that correct for multiple-hypothesis testing.[9] Regressions control for 17 income categories, 6 age categories, indicator variable for female, 16 education categories, and 5 race categories. Time-preference regressions contain a control for risk preferences equal to the average of all the economic (i.e., excluding the self-report) risk-preference measures in the overlap sample, and risk-preference regressions contain a control for time preference analogously.[10]

We also include controls for two variables that are highly relevant yet absent in most analyses that link life outcomes to time and risk preferences: cognitive ability and financial literacy (Goda et al. 2019 is a notable exception). Cognitive ability is an important control for its obvious pervasive role in shaping people's life outcomes (Heckman et al. 2006), but it is rarely included in survey data. Financial literacy has been shown to predict retirement planning, stock market participation, and portfolio diversification (Lusardi and Mitchell 2007, van Rooij et al. 2011, Gaudecker 2015) and is plausibly an important input for other financial outcomes. Cognitive ability is measured in survey 399 using a subset of questions from the public-domain assessment tool, the International Cognitive Ability Resource (ICAR) (Condon and Revelle 2014), using two pattern-matching questions, two mathematical questions, and one mental rotation question. Financial literacy is based on the "Big Three" questions developed by Lusardi and Mitchell (see Lusardi and Mitchell 2014 for a review). Financial literacy was measured in multiple surveys using the exact same three questions. Unsurprisingly, performance on this financial-literacy quiz improves on average as a subject acquires more experience with it (surveys often report the correct answers at the end of the survey). We therefore use a subject's first recorded financial-literacy score.

The 285 regressions in Table C.1 and 684 regressions reported in Table C.2 are difficult to encapsulate in a few sentences. Table 3 summarizes the results for time preferences by outcomes. The columns display the fraction of the coefficients that are at various levels of significance: $p \le 0.05$, $p \le 0.01$, $q \le 0.05$, and $q \le 0.01$. The last of the four sets a high bar as it applies a low critical value while also controlling for the more than 1,000 multiple-hypothesis tests. The table shows that even after correcting for the more than 1,000 hypothesis tests in this paper, many coefficients remain significant at $q < 0.05$, and some even at $q < 0.01$, providing strong evidence that some time preference measures do indeed predict economically relevant outcomes. The outcomes best predicted are owning stock, paying off one's debts in full, checking savings amount, and retirement savings. The measures fail to predict fruit and vegetable consumption, and very poorly predict health, lifespan expectations, having any retirement savings, alcohol consumption, exercise, cereal fiber, going to the doctor routinely, smoking, and medical insurance.

Table 4 shows the analogous summary of the 684 risk-preference regressions. Again, some outcomes are well predicted by the measures. In particular, health, having any stock, and life expectancy are

Table 3. Predictability of Outcomes Using Time-Preference Measures

	Proportion of time-preference coefficients that are significant			
	$p \leq 0.05$	$p \leq 0.01$	$q \leq 0.05$	$q \leq 0.01$
Alcohol	0.20	0.00	0.00	0.00
Any stock	0.40	0.40	0.33	0.27
Bankruptcy	0.33	0.20	0.20	0.00
Cereal fiber	0.07	0.00	0.00	0.00
Checking savings amount	0.47	0.33	0.27	0.13
Fruit vege	0.00	0.00	0.00	0.00
Go to doctor routinely	0.07	0.00	0.00	0.00
Has retirement savings	0.07	0.00	0.00	0.00
Health	0.20	0.07	0.07	0.00
Live longer 85	0.07	0.00	0.00	0.00
Low dose of aspirin	0.13	0.13	0.07	0.00
Medical insurance	0.07	0.00	0.00	0.00
Moderate exercise time	0.20	0.00	0.00	0.00
Pay bill full	0.40	0.20	0.20	0.00
Pay off debt	0.53	0.40	0.33	0.27
Paycheck to paycheck	0.33	0.20	0.20	0.07
Payday loan	0.20	0.07	0.07	0.00
Retirement savings	0.33	0.27	0.20	0.13
Smoking	0.07	0.00	0.00	0.00

Notes. Table summarizes the 285 regressions, 19 outcomes regressed on 15 time-preference measures, reported in Table C.1. Regressions control for cognitive ability, financial literacy, 17 income categories, 6 age categories, indicator variable for female, 16 education categories, and 5 race categories. Risk preferences are controlled by taking the average of all the economic (i.e., excluding the self-report) risk-preference measures in the overlap sample.

Table 4. Predictability of Outcomes Using Risk-Preference Measures

	Proportion of risk-preference coefficients that are significant			
	$p \leq 0.05$	$p \leq 0.01$	$q \leq 0.05$	$q \leq 0.01$
Alcohol	0.00	0.00	0.00	0.00
Any stock	0.50	0.39	0.39	0.14
Bankruptcy	0.00	0.00	0.00	0.00
Cereal fiber	0.06	0.00	0.00	0.00
Checking savings amount	0.08	0.06	0.03	0.00
Fruit vege	0.14	0.08	0.06	0.03
Go to doctor routinely	0.25	0.11	0.11	0.00
Has retirement savings	0.14	0.06	0.06	0.06
Health	0.42	0.31	0.28	0.19
Live longer 85	0.47	0.31	0.28	0.17
Low dose of aspirin	0.06	0.00	0.00	0.00
Medical insurance	0.08	0.03	0.03	0.00
Moderate exercise time	0.11	0.06	0.03	0.00
Pay bill full	0.03	0.00	0.00	0.00
Pay off debt	0.06	0.00	0.00	0.00
Paycheck to paycheck	0.03	0.00	0.00	0.00
Payday loan	0.06	0.00	0.00	0.00
Retirement savings	0.36	0.19	0.11	0.06
Smoking	0.00	0.00	0.00	0.00

Notes. Table summarizes the 684 regressions, 19 outcomes regressed on 36 time-preference measures, reported in Table C.2. Regressions control for cognitive ability, financial literacy, 17 income categories, 6 age categories, indicator variable for female, 16 education categories, and 5 race categories. Time preferences are controlled by taking the average of all the economic (i.e., excluding the self-report) time-preference measures in the overlap sample.

best predicted. Outcomes that are very poorly predicted include bankruptcy, paying off debt in full, paying bills in full, living paycheck to paycheck, alcohol consumption, cereal fiber, low dose of aspirin, smoking, and payday loans.

We now turn to comparing the different measures to each other. Table 5 summarizes Table C.1, displaying the frequency of $p < 0.05$ significant time-preference coefficients by method. Table 5 should be viewed as a rough first-cut comparison of the measures. Different measures have very different sample sizes so the results should be taken with a grain of salt. Significance may be an indication of a good predictor, but it could also be an indication of a large overlap sample. The results suggest that titration, the self-report "for future," and the MPL predict best, in that order, with significant coefficients 40%, 37%, and 32%, respectively, of the time.

Table 6 summarizes Table C.2, displaying the frequency of $p < 0.05$ significant risk-preference coefficients by method. Again, different measures have very different sample sizes so the results should be viewed as preliminary evidence. Here, the self-report measures have by far the strongest performance, with 20% of coefficients significant. In contrast, the next best-performing category, titration, has only 7.9% of its coefficients significant. Tables 6 and 5 show stark variation in the predictiveness of the measures, but part of this variation will be because of variation in sample size.

We next attempt to *horse race* the various measures against each other using approaches that control for the differences in sample size. To serve as our outcome for horse racing, we construct a behavior index for time preference that includes any outcome in Table C.1 that has a $p < 0.01$ for at least one time-preference measure. We then stepwise drop outcomes and time-preference measures that enlarge the sample size the most until it is above 200. Our final index includes any stock, bankruptcy, checking and saving amount, has retirement savings, health, low dose of aspirin, pay off debt, and retirement savings. The index standardizes each behavior, takes the average, and then standardizes the average. We regress the behavior index on the various time-preference measures while controlling for financial literacy, risk aversion, cognitive ability, log income, gender, age, age squared, 16 education categories, and 5 race categories. Coefficients can be interpreted as the effect measured in standard deviations of the index. Table 7 shows the results. Each row uses a different measure of time preferences.

Table 5. Proportion of Significant Coefficients by Time-Preference Elicitation Method and Outcome

	MPL	Titration	CTB	Lifetime	Real effort	For future	Total
Alcohol	0.33	0.4	0	0	0	1	0.27
Any stock	0.33	0.8	0	0	0	1	0.4
Bankruptcy	0	0.8	0.5	0	0	0	0.33
Cereal fiber	0.33	0.2	0	0	0	0	0.13
Checking savings amount	0.67	0.6	0	0	1	1	0.47
Fruit vege	0	0.2	0	0	0	0	0.067
Go to doctor routinely	0	0.4	0	0	0	0	0.13
Has retirement savings	0	0.2	0	0	0	0	0.067
Health	0.33	0.2	0	0	0	1	0.2
Low dose of aspirin	0.33	0.4	0	0	0	0	0.2
Live longer 85	0	0	0	0	1	0	0.067
Medical insurance	0.33	0	0	0	0	0	0.067
Moderate exercise time	0.67	0.2	0	0.33	0	0	0.27
Pay bill full	1	0.6	0	0	0	0	0.4
Pay off debt	1	0.8	0	0	0	1	0.53
Paycheck to paycheck	0.33	0.6	0	0	0	1	0.33
Payday loan	0.33	0.4	0	0	0	0	0.2
Retirement savings	0	0.8	0	0	0	1	0.33
Smoking	0	0	0.5	0	0	0	0.067
Total	0.32	0.4	0.053	0.018	0.11	0.37	0.24

Notes. Table summarizes the 285 regressions, 19 outcomes regressed on 15 time-preference measures, reported in Table C.1. Regressions control for cognitive ability, financial literacy, 17 income categories, 6 age categories, indicator variable for female, 16 education categories, and 5 race categories. Risk preferences are controlled by taking the average of all the economic (i.e., excluding the self-report) risk-preference measures in the overlap sample.

Table 6. Proportion of Significant Coefficients by Risk-Preference Elicitation Method and Outcome

	Titration	Eckel and Grossman	MPL	Self-report	Total
Alcohol	0	0	0	0	0
Any stock	0.25	0	0	0.81	0.64
Bankruptcy	0.13	0	0	0	0.024
Cereal fiber	0	0	0	0.16	0.12
Checking savings amount	0	0	0	0.097	0.071
Fruit vege	0.13	0	0	0.13	0.12
Go to doctor routinely	0.13	0	0	0.35	0.29
Has retirement savings	0.13	0	0	0.13	0.12
Health	0.13	0	0	0.55	0.43
Live longer 85	0.13	0	0	0.74	0.57
Low dose of aspirin	0.13	0	1	0	0.048
Medical insurance	0	0	0	0.097	0.071
Moderate exercise time	0	0	0	0.13	0.095
Pay bill full	0	0	0	0.032	0.024
Pay off debt	0	0	0	0.065	0.048
Paycheck to paycheck	0	0.5	0	0.032	0.048
Payday loan	0.13	0	0	0.065	0.071
Retirement savings	0.25	0	0	0.45	0.38
Smoking	0	0	0	0	0
Total	0.079	0.026	0.053	0.20	0.17

Notes. Table summarizes the 684 regressions, 19 outcomes regressed on 36 time-preference measures, reported in Table C.2. Regressions control for cognitive ability, financial literacy, 17 income categories, 6 age categories, indicator variable for female, 16 education categories, and 5 race categories. Time preferences are controlled by taking the average of all the economic (i.e. excluding the self-report) time-preference measures in the overlap sample.

Table 7. Horse Race: Behavior Index Regressed on $\beta\delta$ Measures

Survey	Coefficient	Standard error	Naive p value	FDR q-value	R^2	N
$\beta\delta$_11 (Titration Pay)	0.127**	0.05	0.018	0.133	0.466	227
$\beta\delta$_14 (MPL Pay)	0.037	0.04	0.410	0.958	0.454	227
$\beta\delta$_248 (MPL Pay)	0.005	0.06	0.931	1.000	0.453	227
$\beta\delta$_10 (Titration Receive)	0.127**	0.06	0.026	0.175	0.465	227
$\beta\delta$_189 (Titration Receive)	0.223***	0.05	0.000	0.001	0.492	227
$\beta\delta$_399 (Titration Receive)	0.123**	0.06	0.029	0.191	0.466	227
ForFuture (SPQ)	0.094*	0.05	0.057	0.289	0.461	227

Notes. OLS with robust standard errors. Demographic controls include log value of income, indicator variables for female, age, age square term, 16 education categories, 5 race categories, cognitive ability, financial literacy, and risk aversion index. The $\beta\delta$ is standardized. The behavior index for $\beta\delta$ only includes the behaviors in Table C.1 for which at least one $\beta\delta$ was significant at $p < 0.01$. The behavior index for time preference includes health, any stock, bankruptcy, pay off debt, checking and saving amount, has retirement savings, retirement savings, and low dose of aspirin. The index standardizes each behavior, takes the average, and then standardizes the average.

*, **, and ***Significant at the 10% level, 5% level, and 1% level, respectively.

Measures from surveys 10, 189, and 399 perform well, with p values of 0.03, 0.000, and 0.002, respectively. These three surveys all use titration methods. Later in Section 5, we show that these three measures have the highest factor loadings. In contrast, survey 11 (pay-format titration), survey 14 (pay-MPL), and survey 248 (pay-titration) are not significant predictors of behavior in this sample. Interestingly the self-report, “I am willing to give up something that is beneficial today in order to receive a greater benefit in the future,” from Falk et al. (2016), performs better than the MPL-pay methods.

Table 7 keeps the sample fixed across all regressions. Unfortunately, to horse-race all the different measures with a fixed sample would require using only individuals that took all the surveys, resulting in a very small sample. Hence, several measures are left out of this table. If Table 7 is the fair horse race in which a limited number of horses run on the same track, it is still possible to conduct a more comprehensive horse race in which more horses are included, but each runs on its own track. In Table 8, we include eight more measures, and the sample is no longer fixed. To control for variation in sample size, we use bootstrapped standard errors with N set by the measure with the lowest sample ($N = 224$). Although variation in sample may be akin to variation in racing tracks, the fixed N used to compute the bootstrapped

Table 8. Horse Race: Behavior Index Regressed on $\beta\delta$ Measures (Bootstrap Version)

Survey	Coefficient	Standard error	Naive p value	FDR q-value	N
$\beta\delta$_11 (Titration Pay)	0.110*	0.06	0.054	0.280	377
$\beta\delta$_10 (Titration Receive)	0.117**	0.05	0.026	0.175	385
$\beta\delta$_276 (Titration Receive)	0.169***	0.04	0.000	0.003	443
$\beta\delta$_189 (Titration Receive)	0.129***	0.04	0.003	0.040	923
$\beta\delta$_399 (Titration Receive)	0.142***	0.05	0.009	0.084	1,287
$\beta\delta$_14 (MPL Pay)	0.045	0.05	0.386	0.921	424
$\beta\delta$_248 (MPL Pay)	0.057	0.05	0.296	0.793	1,138
δ_169 (MPL Receive)	0.125***	0.05	0.006	0.064	285
$\beta\delta$_*ctb*_212 (CTB)	0.018	0.06	0.754	1.000	918
$\beta\delta$_*ctb*_263 (CTB)	0.036	0.03	0.258	0.733	224
Lifetime_14_1	0.015	0.05	0.767	1.000	366
Lifetime_14_2	0.014	0.05	0.791	1.000	377
Lifetime_33	0.046	0.05	0.387	0.922	313
$\beta\delta$_263 (Real Effort)	−0.027	0.04	0.511	1.000	247
ForFuture (SPQ)	0.080	0.05	0.110	0.449	1,288

Notes. OLS with robust standard errors. Demographic controls include log value of income, indicator variables for female, age, age square term, 16 education categories, 5 race categories, cognitive ability, financial literacy, and risk aversion index. The $\beta\delta$ is standardized. The behavior index for $\beta\delta$ only includes the behaviors in Table C.1 for which at least one $\beta\delta$ was significant at $p < 0.01$. The behavior index for time preference includes health, any stock, bankruptcy, pay off debt, checking and saving amount, has retirement savings, retirement savings, and low dose of aspirin. The index standardizes each behavior, takes the average, and then standardizes the average.

*, **, and ***Significant at the 10% level, 5% level, and 1% level, respectively.

Table 9. Horse Race: Behavior Index Regressed on $\beta\delta$ Methods

	(1)	(2)
$\beta\delta$_Titration	0.252***	
	(0.06)	
$\beta\delta$_NonTitration		0.027
		(0.05)
R^2	0.501	0.453
N	227	227
Coefficient equality test *p* value	0.001	

Note. The coefficient of $\beta\delta$_Titration is significantly different from that of $\beta\delta$_NonTitration at $p < 0.01$.
*, **, and ***Significant at the 10% level, 5% level, and 1% level, respectively.

standard errors is like forcing the different tracks to at least be the same distance. The results are consistent with Table 7, with the titration-receive format doing well but pay formats doing poorly. Additionally, the MPL-receive performs quite well. Notably CTB, lifetime, and the real-effort methods do not seem to predict the behavior index.

Some measures are more predictive than others but that does not imply that the coefficients are significantly different from each other. In Table 9 we pool the measures into two categories: titration and nontitration. We define a new $\beta\delta$ for titration. The $\beta\delta$_Titration standardizes each titration measure, takes the average, and then standardizes the average again. We define a new $\beta\delta$ for nontitration similarly. We then regress the behavior index for time preference against $\beta\delta$ for titration and nontitration. We find that the coefficient on $\beta\delta$_Titration is significantly larger than the coefficient on $\beta\delta$_NonTitration ($p = 0.001$). This is a more definitive test, showing that the average of the titration measures is a better predictor than the average of the nontitration measures.

We then conduct a similar analysis for risk preferences with the same controls in Table 10. We construct a behavior index for risk preference that includes any outcome in Table C.2 that has $p < 0.01$ for at least one risk-preference measure. We then stepwise drop outcomes and risk-preference measures that enlarge the sample size the most until it is above 200. The outcomes include any stock, checking savings amount, fruit/vegetable consumption, go to doctor routinely, has retirement savings, health, live longer 85, moderate exercise time, and retirement savings. The index standardizes each behavior, takes the average, and then standardizes the average. We find that none of the economic measures predict the index well. Two self-reports actually perform better than all the economic measures: self-reported general risk aversion and self-reported risk aversion about life changes. Both questions originate from Dohmen et al. (2011).

In Table 11, we allow the sample to vary but use bootstrapped standard errors based on the measure with the smallest sample ($N = 226$). This is the horse race with different but equal length tracks. We add 21 new measures. The table exhibits a startling pattern: the economic measures perform poorly but the qualitative self-reports perform much better. No economic measure has a *p* value less than 0.1. However, the general and financial self-reported risk-aversion

Table 10. Horse Race: Behavior Index Regressed on Risk Aversion Measures

Survey	Coefficient	Standard error	Naive *p* value	FDR q-value	R^2	*N*
Risk Avers 197 (Titration Hypothetical Job)	−0.09*	0.05	0.071	0.327	0.42	231
Risk Avers 197 (Titration Alternative Job)	−0.05	0.06	0.388	0.924	0.41	231
Risk Avers 197 (Titration Insurance)	−0.08	0.06	0.195	0.653	0.41	231
Risk Avers 50 (Titration Inheritance)	−0.09	0.07	0.193	0.653	0.42	231
Risk Avers 189 (Titration Inheritance)	−0.02	0.06	0.771	1.000	0.41	231
Risk Avers 243 (Titration Ball)	0.01	0.05	0.901	1.000	0.41	231
Risk Avers 118 (Titration Retire Savings)	−0.10*	0.06	0.092	0.394	0.42	231
Risk Avers 210 (MPL)	−0.01	0.06	0.821	1.000	0.41	231
Risk Avers SelfReport 197 (General)	−0.12**	0.05	0.018	0.130	0.42	231
Risk Avers SelfReport 197 (Driving)	−0.02	0.06	0.716	1.000	0.41	231
Risk Avers SelfReport 197 (Financial)	−0.10*	0.05	0.059	0.295	0.42	231
Risk Avers SelfReport 197 (Occupation)	−0.07	0.05	0.203	0.663	0.41	231
Risk Avers SelfReport 197 (Health)	0.04	0.05	0.521	1.000	0.41	231
Risk Avers SelfReport 197 (Social Relation)	−0.01	0.05	0.828	1.000	0.41	231
Risk Avers SelfReport 197 (Life Change)	−0.14**	0.06	0.012	0.104	0.43	231

Notes. Demographic controls include log value of income, indicator variables for female, age, age square term, 16 education categories, 5 race categories, cognitive ability, financial literacy, and $\beta\delta$ index. The risk aversion is standardized. The behavior index for risk preference only includes the behaviors in Table C.2 for which at least one risk aversion was significant at $p < 0.01$. The behaviors include health, any stock, live longer 85, checking savings amount, has retirement savings, retirement savings, moderate exercise time, fruit vege, and go to doctor routinely. The index standardizes each behavior, takes the average, and then standardizes the average.
*, **, and ***Significant at the 10% level, 5% level, and 1% level, respectively.

Table 11. Horse Race: Behavior Index Regressed on Risk Aversion Measures (Bootstrap Version)

Survey	Coefficient	Standard error	Naive p value	FDR q-value	N
Risk Avers 167 (Titration Hypothetical Job)	−0.02	0.04	0.544	1.000	288
Risk Avers 197 (Titration Hypothetical Job)	−0.02	0.05	0.600	1.000	692
Risk Avers 197 (Titration Alternative Job)	−0.06	0.04	0.178	0.613	692
Risk Avers 197 (Titration Insurance)	0.00	0.04	0.974	1.000	692
Risk Avers 50 (Titration Inheritance)	−0.03	0.05	0.535	1.000	621
Risk Avers 189 (Titration Inheritance)	−0.04	0.04	0.259	0.733	923
Risk Avers 399 (Eckel & Grossman)	−0.03	0.06	0.655	1.000	2,272
Risk Avers 263 (Eckel & Grossman)	−0.05	0.05	0.385	0.921	340
Risk Avers 243 (Titration Ball)	−0.00	0.04	0.963	1.000	1,044
Risk Avers 118 (Titration Retire Savings)	−0.04	0.04	0.261	0.737	808
Risk Avers 210 (MPL)	−0.02	0.04	0.697	1.000	1,030
Risk Avers SelfReport 133 (General)	−0.11**	0.04	0.017	0.127	824
Risk Avers SelfReport 197 (General)	−0.11***	0.03	0.000	0.006	693
Risk Avers SelfReport 12 (General)	−0.07*	0.04	0.083	0.366	226
Risk Avers SelfReport 48 (General)	−0.03	0.03	0.276	0.763	612
Risk Avers SelfReport 169 (General)	−0.07**	0.03	0.011	0.096	295
Risk Avers SelfReport 248 (General)	−0.16***	0.05	0.003	0.034	749
Risk Avers SelfReport 189 (General)	−0.09**	0.04	0.029	0.191	929
Risk Avers SelfReport 349 (Financial)	−0.05	0.03	0.111	0.449	1,157
Risk Avers SelfReport 342 (Financial)	−0.15***	0.05	0.001	0.021	527
Risk Avers SelfReport 197 (Financial)	−0.13***	0.03	0.000	0.001	693
Risk Avers SelfReport 193 (Financial)	−0.09**	0.04	0.029	0.190	436
Risk Avers SelfReport 342 (Financial)	−0.11***	0.04	0.001	0.021	526
Risk Avers SelfReport 130 (Financial)	−0.13***	0.04	0.001	0.014	298
Risk Avers SelfReport 260 (Financial)	−0.11**	0.04	0.014	0.110	584
Risk Avers SelfReport 167 (Financial)	−0.10***	0.03	0.001	0.019	290
Risk Avers SelfReport 179 (Financial)	−0.03	0.04	0.549	1.000	894
Risk Avers SelfReport 130 (Daily Act)	−0.08*	0.05	0.088	0.382	298
Risk Avers SelfReport 260 (Daily Act)	−0.07	0.05	0.107	0.445	584
Risk Avers SelfReport 167 (Daily Act)	−0.06	0.04	0.149	0.548	290
Risk Avers SelfReport 197 (Driving)	−0.00	0.05	0.975	1.000	693
Risk Avers SelfReport 197 (Occupation)	−0.07*	0.04	0.079	0.352	693
Risk Avers SelfReport 197 (Health)	0.03	0.04	0.374	0.905	692
Risk Avers SelfReport 197 (Social Relation)	−0.03	0.04	0.525	1.000	693
Risk Avers SelfReport 197 (Life Change)	−0.10**	0.04	0.026	0.175	693
Risk Avers SelfReport 179 (Safe Invest Importance)	−0.07	0.05	0.156	0.564	893

Notes. OLS with bootstrap estimates of standard errors. The size of bootstrap is 226. Demographic controls include log value of income, indicator variables for female, age, age square term, 16 education categories, 5 race categories, cognitive ability, financial literacy, and $\beta\delta$ index. The risk aversion is standardized. Survey 2, 186, 338, and 339 are not used because of insufficient sample size. The behavior index for risk preference only includes the behaviors in Table C.2 for which at least one risk aversion was significant at $p < 0.01$. The behaviors include health, any stock, live longer 85, checking savings amount, has retirement savings, retirement savings, moderate exercise time, fruit vege, and go to doctor routinely. The index standardizes each behavior, takes the average and then standardizes the average.

*, **, and ***Significant at the 10% level, 5% level, and 1% level, respectively.

measures perform well with five of seven predicting the index with $p < 0.05$, and seven of nine measures predicting the index with $p < 0.05$.

To provide a more definitive comparison between the economic risk-aversion measures and the qualitative self-report measures, we pool all measures into the two categories. We define a new risk-aversion measure for the economic elicitations. The index standardizes each economic measure, takes the average, and then standardizes the average again. Likewise, we define a new risk-aversion measure for the self-report elicitations. We then regress the outcome index for risk preference on the two new risk aversion measures, RiskAversion_Econ and RiskAversion_SelfReport, in Table 12. The coefficients are not significantly different from each other. Although the relative performance between these categories is stark in Table 11 in which the sample varies, the self-reports do not perform significantly better in the overlap sample.

To summarize, we find considerable heterogeneity in the quality of measurements for predicting economic behavior. Although some measures are insignificant, others are highly significant at the $q < 0.01$ level after controlling for multiple hypothesis testing. For time-preference elicitations, we find that the titration-receive method produces measures that consistently predict well. For risk preferences, no economic measures robustly predict the index. However, the self-reports on general, financial, and life-change risk aversion robustly predict well.

Table 12. Horse Race: Behavior Index Regressed on Risk Aversion Methods

	(1)	(2)
RiskAversion_Econ	−0.128**	
	(0.06)	
RiskAversion_SelfReport		−0.094
		(0.06)
R^2	0.422	0.417
N	231	231
Coefficient Equality Test p-value	0.636	

Note. The coefficient of RiskAversion_Econ is not significantly different from that of RiskAversion_SelfReport.

*, **, and ***Significant at the 10% level, 5% level, and 1% level, respectively.

5. Distinctness

In this section, we attempt to answer the following questions. (1) Are the preference measures proxies for other individual attributes such as cognitive ability, education, or income? (2) Do time-preference measures and risk-preference measures reflect the same underlying latent trait? (3) Are the measures for a given preference parameter sufficiently correlated that there is a single dimension of variation?

To answer these questions, we conduct exploratory factor analysis. Table 13 displays exploratory factor analysis among all measures and additionally includes FICO,[11] income, education, and two measures of cognitive ability. Our data for exploratory factor analysis is incomplete because individuals will have missing values for some of these measures. We use the expectation-maximization method suggested by Truxillo (2005) for incomplete data.[12] Orthogonal varimax rotation is performed for factor rotation. Keeping all factors with positive eigenvalues produces 17 factors. Figure C.6 in the online appendix displays the scree plot and parallel analysis. The figure indicates that only the first five factors should be retained. The eigenfactors of the five rotated factors are 8.58, 3.33, 2.45, 2.34, and 1.65. The percent of variance explained by the first five factors are, respectively, 28.35%, 11%, 8.10%, 7.72%, and 5.45%. Altogether, they explain 81.65% of total variance. Table 13 reports the rotated factor loadings. Factor loadings less than 0.3 are omitted and loadings greater than 0.5 are in bold.

Factor 1 loads on risk-preference self-reports across various contexts. Factor 2 loads on the five cognitive ability questions, the three IQ scores, income, education, and FICO. Factor 3 loads on time-preference measures and FICO. The factor loading on FICO is small relative to the loadings on the time-preference measures. Factor 4 loads on questions related to risk in driving, occupation, health, and social relations. Factor 5 loads primarily on the titration-job measure.

There are five lessons from this table. (1) Time-preference measures primarily load on one factor (factor 3). This suggests that time preferences in our data may be adequately represented by a single latent parameter.[13] (2) Factor 1 loads highly on self-report risk preferences and does not load on much else. This suggests that risk preference is indeed distinct from time preference or other characteristics like income, cognitive ability, or education. (3) Some of the risk self-report measures and titration-job measures load on secondary factors, factors 4 and 5, respectively, suggesting that these measures may capture different latent variables that are distinct from factor 1. Risk preferences may be multidimensional; on the other hand, factor 1 explains more than double the variation of factors 4 and 5 combined, implying that unidimensional risk preferences may be sufficiently close to the truth. (4) Time-preference and risk-preference measures load on different factors indicating that they are measuring distinct latent characteristics. Confirmatory factor analysis in Table C.3 of the online appendix supports this claim. (5) Cognitive ability does not load on the same factors as the preference measures indicating that the preference measures are distinct from cognitive ability.

Table B.5 in the online appendix displays how closely the time-preference and risk-aversion measures correlate with each other. Because of limitations of overlapping size, we limit surveys to five time-preference surveys and eight risk-aversion surveys. The sample includes only subjects who respond to all 13 surveys ($n = 449$). The time preference measures are not correlated with the risk-preference measures. The correlation is −0.03, on average, with a median of −0.01.

In summary, the results are consistent with the claims that time preferences are unidimensional and well captured by the titration measures, and risk preferences are unidimensional and well captured by the self-report measures. Time preferences and risk preferences are unrelated characteristics, and they are also distinct from cognitive ability, FICO, income, and education.

If time preferences and risk preferences are indeed separate latent characteristics that are distinct from other observables, we might expect that these measures would have independent effects on outcomes. Inspired by the analysis in Stango et al. (2017b) on the independence of behavioral biases, we regress economic outcomes on one of our top-performing time-preference measures and top-performing risk-preference measures and then remove various control sets.[14] The exercise is meant to test the degree of omitted-variable bias when regressing real-life outcomes on these preference measures. We vary the control sets and then observe the extent to which the

Table 13. Factor Analysis on Standardized Parameters

	F1	F2	F3	F4	F5
Risk Aversion Titration Job Survey2_1					
Risk Aversion Titration Job Survey2_2					
Risk Aversion Titration Job Survey167					0.43
Risk Aversion Titration Job Survey197_Hypothetical Job Context					**0.61**
Risk Aversion Titration Job Survey197_Alternative Job Context					**0.53**
Risk Aversion Titration Insurance Survey197_Insurance Context					
Risk Aversion Titration Inheritance Survey50					
Risk Aversion Titration Inheritance Survey189					
Risk Aversion Titration Inheritance Survey186					
Risk Aversion Eckel & Grossman Survey399					
Risk Aversion Eckel & Grossman Survey263					
Risk Aversion Titration Ball Survey243					
Risk Aversion Titration Ball Survey338					
Risk Aversion Titration Ball Survey339					
Risk Aversion Titration Retire Savings Survey118					
Risk Aversion MPL Survey210					
BetaDelta Titration Pay Survey11					
BetaDelta MPL Pay Survey14					
BetaDelta MPL Pay Survey248					
BetaDelta Titration Receive Survey10			0.49		
BetaDelta Titration Receive Survey276			**0.54**		
BetaDelta Titration Receive Survey125					
BetaDelta Titration Receive Survey126					
BetaDelta Titration Receive Survey189			**0.62**		
BetaDelta Titration Receive Survey399			**0.58**		
BetaDelta Real Effort Survey263					
BetaDelta CTB Survey212					
BetaDelta CTB Survey263					
Delta MPL Receive Survey169			0.37		
Lifetime Survey14_1					
Lifetime Survey14_2					
Lifetime Survey33					
SPQ DoWhenPlan					
SPQ ForFuture					
SPQ Postpone					
Risk Aversion SelfReport Generally Survey133	**0.55**				
Risk Aversion SelfReport Generally Survey197	**0.74**				
Risk Aversion SelfReport Generally Survey12	**0.69**				
Risk Aversion SelfReport Generally Survey48	**0.65**				
Risk Aversion SelfReport Generally Survey169	0.47				
Risk Aversion SelfReport Generally Survey284	**0.70**				
Risk Aversion SelfReport Generally Survey189	**0.67**				
Risk Aversion SelfReport Financial Matters Survey349	**0.61**				
Risk Aversion SelfReport Financial Matters Survey342	**0.72**				
Risk Aversion SelfReport Financial Matters Survey197	**0.74**				
Risk Aversion SelfReport Financial Matters Survey193	**0.60**				
Risk Aversion SelfReport Financial Matters Survey186	0.39				
Risk Aversion SelfReport Financial Matters Survey342	**0.57**	−0.34			
Risk Aversion SelfReport Financial Matters Survey130	**0.72**				
Risk Aversion SelfReport Financial Matters Survey260	**0.75**				
Risk Aversion SelfReport Financial Matters Survey167	**0.72**				
Risk Aversion SelfReport Financial Matters Survey179	0.37				
Risk Aversion SelfReport Daily Activities Survey130	**0.54**				
Risk Aversion SelfReport Daily Activities Survey260	**0.58**				
Risk Aversion SelfReport Daily Activities Survey167	**0.51**				
Risk Aversion SelfReport while Driving Survey197				**0.62**	
Risk Aversion SelfReport in your Occupation Survey197	**0.50**			0.35	
Risk Aversion SelfReport with your Health Survey197				**0.62**	
Risk Aversion SelfReport in your Social Relation Survey197	0.32			0.41	
Risk Aversion SelfReport in making Life Changes Survey197	0.45				
Risk Aversion SelfReport Safe Investment Importance Survey179	0.44				
FICO		0.39	0.32		

Table 13. (Continued)

	F1	F2	F3	F4	F5
Cognitive Agesister Survey399		**0.50**			
Cognitive Onefifith Survey399		**0.52**			
Cognitive Nextletter Survey399		0.40			
Cognitive Bestfigure Survey399		0.35			
Cognitive Rotatecube Survey399					
IQ Number Survey286		**0.66**			
IQ Picture Survey286		**0.53**			
IQ Analogy Survey286		**0.67**			
Income		0.42			
Education		0.43			

Notes. Factors obtained with maximum likelihood and orthogonal varimax rotation. Incomplete data are estimated by expectation-maximization method. Factor loadings less than 0.3 are omitted and loadings greater than 0.5 are in bold.

coefficients change. The supposition is that if omitted-variable bias appears limited when omitting observables from the regression, then it is more likely that omitted-variable bias is limited when omitting unobservables from the regression.

Table 14 shows the results for time preferences. The rows each represent an outcome, and the columns represent a specification. As we look across the columns, we see remarkable consistency in the estimated coefficients. Of the 10 estimated coefficients that are significant at the $p < 0.05$ level in the full specification, all are significant at $p < 0.05$ when the full control set is removed, and of the 17 significant coefficients at the $p < 0.05$ in the uncontrolled regression, 10 remain significant at the $p < 0.05$ when the full control set is added. Many of the coefficients vary by less than a third. The effect of the time-preference measurements on economic outcomes appears to be mostly unmediated by the control variables. As we eliminate control sets the coefficients remain remarkably unaltered. This gives some reassurance that omitted-variable bias may not be a major concern more generally.

Table 15 shows the results for risk preferences. Observe that there are far less specifications that are significant. For the one coefficient with $p < 0.05$ in the full specification, removing the full control set does not alter this fact. Of all the six significant coefficients at the $p < 0.05$ in the uncontrolled regression only one remains significant at the $p < 0.05$ when the full control set is added. It is harder to draw any conclusions regarding risk preferences because the predictiveness is much weaker than for time preferences.

6. Discussion

6.1. Comparison with Earlier Studies

There is a sizable amount of literature that has correlated time- and risk-preference measures with real-world outcomes. We report papers that we found on this topic in Tables 16 and 17. We do not include papers that do not include a real-world outcome (e.g., papers that are conducted entirely in the laboratory), that use the ALP sample, or papers that use related but conceptually different measures (e.g., psychological measures of self-control instead of time preference). There may be relevant papers that we missed in our search, but this is not meant to be an entirely comprehensive survey, which is a large undertaking and deserving of a paper in its own right. Nonetheless, we believe that our coverage is wide and sufficient to give perspective to our results. Statistically significant coefficients at $p < 0.05$ are marked with a plus sign in the far right column. Some papers include multiple specifications in which the coefficient is sometimes significant and sometimes not. If the coefficients are not consistently significant across specifications, we do not count this as significant in the table. If there are multiple coefficients for different parameters (e.g., β versus δ) with at least the coefficient of one parameter being consistently significant, we count this as significant in the table. Sample size may vary across specifications. We report the largest observed sample size.

We offer several observations on Table 16 in connection to our work. (1) The methods used differ from those in our ALP sample. Notably, MPL and the method offered by Kirby and Maraković (1996) are quite common outside our sample, and titration seems to be almost entirely absent. (2) There are no clear patterns of significance based on method. Evident from the table, there is wide variation in terms of methods, sample size (varying by nearly three orders of magnitude), and outcomes. Not reported in the table, there is additionally wide variation in the samples, how the measures are constructed (e.g., discount rate versus δ discount factor versus $\beta\delta$ discount factor), and which control variables are used. Both observations show that the results in the current paper fill a hole in the literature. First, we compare methods, some of which have undergone relatively little study previously. Second, we analyze the

Table 14. Standardized $\beta\delta$ Coefficients When Removing Covariates

	Full specification	Remove cognitive ability	Remove risk aversion	Remove age, gender, education	Remove financial literacy	Remove all covariates	*N*
Health	0.00	0.00	0.00	0.01	0.00	0.10***	920
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Any stock	0.06***	0.07***	0.07***	0.08***	0.07***	0.13***	862
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Live longer 85	0.72	0.81	0.81	1.78*	0.72	2.23**	918
	(0.97)	(0.95)	(0.97)	(0.98)	(0.97)	(0.90)	
Bankruptcy	−0.03***	−0.03***	−0.03***	−0.04***	−0.03***	−0.04***	894
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Pay off debt	0.11***	0.12***	0.11***	0.13***	0.11***	0.15***	791
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Checking savings amount	2493***	2734***	2476***	2951***	2505***	4907***	806
	(541)	(543)	(540)	(534)	(539)	(589)	
Has retirement savings	0.02	0.01	0.02	0.02	0.02	0.08***	766
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Pay bill full	0.15***	0.16***	0.14***	0.16***	0.15***	0.21***	477
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Paycheck to paycheck	−0.30***	−0.30***	−0.30***	−0.36***	−0.30***	−0.39***	284
	(0.09)	(0.08)	(0.09)	(0.09)	(0.09)	(0.08)	
Retirement savings	17759***	18797***	18283***	22872***	17790***	50684***	862
	(5373)	(5386)	(5354)	(5322)	(5366)	(6106)	
Alcohol	0.01	0.01	0.00	0.01	0.01	0.04	785
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	
Moderate exercise time	0.15*	0.14*	0.14	0.16*	0.15*	0.22***	785
	(0.09)	(0.08)	(0.09)	(0.08)	(0.09)	(0.08)	
Fruit vege	0.10	0.12*	0.11*	0.13**	0.11*	0.19***	784
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	
Cereal fiber	0.12*	0.12*	0.12*	0.10	0.12*	0.11**	785
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	
Go to doctor routinely	0.02	0.01	0.01	0.03	0.02	0.00	785
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Low dose of aspirin	0.04**	0.04**	0.04**	0.07***	0.04**	0.07***	785
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Smoking	−0.08**	−0.08**	−0.08**	−0.11***	−0.08**	−0.17***	921
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	
Payday loan	−0.09**	−0.10**	−0.09**	−0.09**	−0.09**	−0.11***	502
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	
Medical insurance	0.00	0.00	0.00	0.01	0.00	0.04**	655
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	

Notes. All regressions are OLS with robust standard errors. Demographic controls include 17 income categories, 6 age categories, indicator variables for female, 16 education categories, and 5 race categories. Both $\beta\delta$ and risk aversion are from survey 189.

*, **, and ***Significant at the 10% level, 5% level, and 1% level, respectively.

predictability of methods in a way that allows for meaningful comparisons. (3) The table also reveals that the present paper is the first to find an association between time-preference measures and having stocks and bankruptcy.

The observations on Table 16 mostly apply to Table 17 as well. Risk-preference measures used outside of the ALP are similar to those in our sample, unlike for time-preference measures. However, observation (2) largely applies to risk-preference measures, with literally more than three orders of magnitude difference between the smallest sample and largest sample. Our finding that self-report measures of risk predict well appears consistent with previous results. Dohmen et al. (2011) find that the general risk question predicts behavior better than any other measure in their sample. Beauchamp et al. (2017) also find that the general and financial risk questions are good predictors of behavior. This is also complementary with the findings of Mata et al. (2018), who conducted a

Table 15. Standardized Risk-Aversion Coefficients When Removing Covariates

	Full specification	Remove cognitive ability	Remove risk aversion	Remove age, gender, education	Remove financial literacy	Remove all covariates	N
Health	−0.04	−0.04	−0.04	−0.04	−0.04	−0.10***	920
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Any stock	−0.04***	−0.04***	−0.04***	−0.02*	−0.04***	−0.06***	862
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Live longer 85	−1.33	−1.42*	−1.37	−0.02	−1.34	−1.13	918
	(0.84)	(0.83)	(0.84)	(0.83)	(0.84)	(0.82)	
Bankruptcy	0.01	0.01	0.01	0.01	0.01	0.01	894
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Pay off debt	−0.00	−0.01	−0.01	0.01	−0.00	−0.01	791
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Checking savings amount	241	−5	87	548	238	-1,228**	806
	(503)	(505)	(511)	(478)	(502)	(540)	
Has retirement savings	−0.03*	−0.03	−0.03*	−0.04**	−0.03*	−0.07***	766
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Pay bill full	0.02	0.01	0.01	0.03	0.01	−0.03	477
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Paycheck to paycheck	0.03	0.02	0.04	0.01	0.03	0.01	284
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	
Retirement savings	-8,332*	-9,292**	-9,315*	−3980	-8,379*	-21,203***	862
	(4,740)	(4,715)	(4,761)	(4,654)	(4,735)	(6,032)	
Alcohol	0.04	0.04	0.04	0.02	0.04	−0.01	785
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Moderate exercise time	0.07	0.07	0.05	0.02	0.06	−0.03	785
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	
Fruit vege	−0.07	−0.09	−0.08	0.00	−0.07	−0.04	784
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	
Cereal fiber	−0.02	−0.02	−0.03	−0.03	−0.02	−0.06	785
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	
Go to doctor routinely	0.04*	0.05**	0.04*	0.05**	0.04*	0.06***	785
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Low dose of aspirin	0.01	0.02	0.01	0.03*	0.02	0.02	785
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Smoking	−0.01	−0.01	−0.01	−0.03	−0.01	0.03	921
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Payday loan	0.03	0.03	0.03	0.01	0.03	0.01	502
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Medical insurance	0.01	0.01	0.01	0.01	0.01	−0.01	655
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	

Notes. All regressions are OLS with robust standard errors. Demographic controls include 17 income categories, 6 age categories, indicator variables for female, 16 education categories, and 5 race categories. Both $\beta\delta$ and risk aversion are from survey 189.
*, **, and ***Significant at the 10% level, 5% level, and 1% level, respectively.

meta-analysis of studies that conduct test-retests on risk preference elicitations. They found that economic measures tend to have fairly low test-retest correlations compared with self-reports/self-reports.

Our factor analysis results are consistent with the results of Dohmen et al. (2011) who conduct factor analysis on five different self-report risk-preference measures and find that a single factor explains 60% of the variation. Frey et al. (2017) found similar results. We find no association between the time-preference measures and the risk-preference measures. This is consistent with the claim that time preferences and risk preferences are distinct latent variables. This is in contrast to Dean and Ortoleva (2015), who find that risk aversion (method unstated) is significantly associated with time preference as measured by MPL, with a correlation of 0.32.

We find that cognitive ability is a distinct factor from either time-preference or risk-preference. Burks et al. (2009) find that cognitive ability is correlated with their time-preference measure (MPL Receive), and their risk-preference measure (MPL Receive), with correlations ranging in magnitude from 0.126 to 0.202. Dohmen et al. (2010) find a correlation

Table 16. Papers That Use a Time-Preference Measure to Predict a Real-World Outcome

Paper	Sample size	Method	Notes	Outcome	Significance
Ashraf, Karlan and Yin (2006)	715	MPL		Commitment savings product	
Borghans and Golsteyn (2006)	2,059	Other	Six binary questions	BMI	
Eisenhauer and Ventura (2006)	3,237	Other	BDM method	Wealth class	+
				Prepayments on property	
				Own home	+
				Contributions to pension	+
Chabris et al. (2008)	126	Other	Kirby and Maraković (1996) method	BMI	+
				Smoking	+
				Exercise	+
				Diet	
Daly, Harmon and Delaney (2009)	149	Other	Kirby and Maraković (1996) method	Body fat %	
				BMI	
				R-R interbeat interval	
				SD of R-R	
				Systolic blood pressure	+
				Diastolic blood pressure	
				Glucose	
Menon and Perali (2009)	6,496	MPL		School performance	
				Reservation wage	
Meier and Sprenger (2010)	541	MPL		Credit card borrowing	+
Tanaka, Camerer and Nguyen (2010)	5,340	MPL		Education	
			instrumental variable for income	Income	
Burks et al. (2012)	1,069	MPL		Smoking	+
				Credit score	+
				BMI	
				Leave	
				Washout	+
				awol	+
		Other	Big red button method	Smoking	
				Credit score	
				BMI	
				Leave	
				Washout	
				awol	
		SPQs		Smoking	
				Credit score	
				BMI	
				Leave	
				Washout	+
				awol	
Courtemanche et al. (2015)	5,982	Other	BDM method	BMI	+
Augenblick et al. (2015)	102	CTB		Demand for commitment	
		CTBRealEffort		Demand for commitment	+
Kaur et al. (2015)	111	Other	Six binary questions	Demand for commitment	
				Goal setting	
				Earnings	+
Bradford et al. (2017)	872	MPL		Health	+
				Mental health	+
				Activity limited	+
				Health insurance	+
				Bought own health insurance	+
				Obese	
				BMI	+
				Exercise	+
				Snacks per day	+
				Cigarettes per day	+
				Binge drinking last month	+
				Use sunscreen	
				Use seatbelts	+
				Fuel efficiency	
				Installed CFL	+

Table 16. (Continued)

Paper	Sample size	Method	Notes	Outcome	Significance
				Added insulation	
				Weather stripping	+
				Programmable thermostat	
				Summer temperature in home	
				Less energy than average	
				Energy audit	
				Has credit card	+
				Credit card balance	+
				Nonretirement savings	+
				Any retirement savings	+
Blumenstock et al. (2018)	702	CTB	Not a full CTB, a few descrete choices	Default effect for savings	+
Falk et al. (2018)	70,000	Titration+ForFuture	71% titration+29% ForFuture, cross-global sample	Saved last year	+
				Education	+

Notes. Papers included contain a real-world outcome, use a sample other than the ALP, and have a measure of time preference. Largest observed sample size reported. Sample size may vary across specifications. Statistically significant coefficients at $p < 0.05$ are marked with a plus sign in the right column.

between cognitive ability and a risk-preference measure (MPL Receive) of magnitude 0.233 and with a time-preference measure (MPL Receive) of 0.124. Both research teams find that the associations between the preference measures and cognitive ability are highly significant. Although this may seem to conflict with our results, they do not. We find similar rank correlations ranging from 0.00 (not significant) to 0.25 ($p < 0.0001$) for time-preference and cognitive ability and from 0.00 (not significant) to 0.20 ($p < 0.0001$) for risk-preference and cognitive ability, depending on the pair of measures. In this sense, our results are consistent with past findings. These correlations, although sometimes significant, are relatively low compared with the within-category correlations of cognitive ability, time preference, and risk preference, thereby explaining why we find that these are distinct factors.

6.2. Interpretation

For time preferences, the titration-receive method yields measures that predict a large number of economic behaviors in the direction indicated by theory. The measures also show high factor loadings in exploratory factor analysis on the primary time-preference factor. Collectively, the results suggest that the titration-receive method is measuring a latent trait that is distinct from other observables and is related to economic behavior. Other methods that have low predictiveness and low factor loadings may exhibit a high degree of measurement error.

For risk preferences, the self-reports about general or financial risk aversion predict behavior the best and in the direction indicated by theory. The first factor to emerge in the exploratory factor analysis loads heavily on these questions as well. The evidence suggests that simple self-reports for risk aversion are effective predictors and represent a latent trait that is distinct from other observables. In contrast, the economic measures predict poorly and do not cohere in the factor analysis.

6.3. Limitations

It is important to keep in mind the limitations of the data and analysis. First, all elicitations are conducted via Internet survey, primarily for hypothetical stakes. The quality of the measures may change under different circumstances and with different incentives. Second, the exact implementation may differ across surveys, even within a given method, reducing control. For example, two titration-receive time-preference elicitations could use different time horizons, different stakes, and a different number of total titrations. Even if each survey was a self-contained controlled study, no control was exercised across surveys. The data, for our purposes, is thus observational as opposed to controlled. Third, the elicitations may not be paragons of their methodology. For example, survey 263, which used the CTB time preference elicitation, contained 12 time budgets, whereas the original version in Andreoni and Sprenger (2012) contained 45 time budgets.

Table 17. Papers That Use a Risk-Preference Measure to Predict a Real-World Outcome

Paper	Sample size	Method	Notes	Outcome	Significance
Barsky et al. (1997)	11,707	Titration–job		Smoke	+
				Drinks	+
				Education	
				Self-employed	
				Immigrant	
				No health insurance	+
				No life insurance	+
				Owns home	+
Khwaja, Sloan and Salm (2006)	20,975	Titration–job		Smoker	+
Bonin et al. (2007)	4,094	Self report–general		Earnings	+
Anderson and Mellor (2008)	978	MPL	Holt and Laury (2002) method	Smoking	
				Heavy drinking	
				Overweight/obese	+
				Does not use seat belt	
				Drives fast	
Menon and Perali (2009)	6,496	MPL		School performance	
				Reservation wage	
Tanaka, Camerer and Nguyen (2010)	181	MPL	Prospect theory parameters, IV for income	Education	+
				Income	
Dohmen et al. (2011)	13,571	Self report–general		Stocks	+
				Active sports	+
				Self-employed	+
				Smoking	+
		Self report–driving		Stocks	+
				Active sports	+
				Self-employed	+
				Smoking	+
		Self report–financial		Stocks	+
				Active sports	+
				Self-employed	+
				Smoking	
		Self report–sports/leisure		Stocks	+
				Active sports	+
				Self-employed	
				Smoking	
		Self report–career		Stocks	+
				Active sports	+
				Self-employed	+
				Smoking	+
		Self-report–health		Stocks	+
				Active sports	+
				Self-employed	+
				Smoking	+
Ahn (2010)	27,650	Titration–job		Self-employed	+
Beauchamp et al. (2017)	11,418	Self report–general	Corrects for measurement error	Portfolio risk	+
				Equity share	+
				Own business	+
				Alcohol	+
				Smoking	+
		Self report–financial		Portfolio risk	+
				Equity share	+
				Own business	+
				Alcohol	+
				Smoking	
		Titration–job		Portfolio risk	+
				Equity share	+
				Own business	+
				Alcohol	
				Smoking	
Falk et al. (2018)	~ 70,000			Own business	+

Table 17. (Continued)

Paper	Sample size	Method	Notes	Outcome	Significance
		Titration and self-report–general	47% titration + 53% self-report–general	plan start business	+
				smoking intensity	+
Galizzi et al. (2016)	608	MPL	Holt and Laury (2002) method	Smoking	
				BMI	+
				Junk food	
				Fruit veges	+
				Heavy drinking	
				Savings	
				Pension	
		Eckel and Grossman		Smoking	
				BMI	+
				Junk food	
				Fruit veges	
				Heavy drinking	
				Savings	
				Pension	+
		General		Smoking	
				BMI	
				Junk food	
				Fruit veges	
				Heavy drinking	
				Savings	+
				Pension	
		Finance		Smoking	
				BMI	
				Junk food	
				Fruit veges	
				Heavy drinking	+
				Savings	
				Pension	
		Health		Smoking	
				BMI	
				Junk food	
				Fruit veges	
				Heavy drinking	
				Savings	
				Pension	
Andreoni and Kuhn (2019)	66	MPL	MPL is Holt and Laury (2002) method. All become sig, except for probability equivalent on gambles, after controlling for measurement error.	Gambles	
				Buys warranties	
		Certainty equivalent		Gambles	
				Buys warranties	
		Probability equivalent		Gambles	
				Buys warranties	
		mCRB		Gambles	
				Buys warranties	

Notes. Papers included contain a real-world outcome, use a sample other than the ALP, and have a measure of risk preference. Largest observed sample size reported. Sample size may vary across specifications. Statistically significant coefficients at $p < 0.05$ are marked with a plus sign in the right column.

7. Conclusion

The results in this paper may be informative for researchers who plan to run time- or risk-preference elicitations in surveys. We find that titration time-preference measures and self-report risk-preference measures predict behaviors that theory suggests should depend on time and risk preference. Although we do not claim that this paper provides a definitive horse race comparing methods of elicitation, the heterogeneity in measure performance implies that researchers should be cautious when selecting their elicitation method. Research is needed to test the relative performance of elicitation methods under more controlled circumstances. Nonetheless, even imperfect comparisons can be informative and provide guidance to practitioners.

Knowing how to best measure time and risk preferences has managerial implications. The financial wellness of employees has become a concern at some firms. As we demonstrated in the prediction section of the paper, effective measures predict many important financial behaviors including owning stock, paying off one's debt, the amount in one's checking account, and retirement savings. To better cater to employees' needs, firms may wish to estimate their employees' financial preferences (Blumenstock et al. 2018), and having good predictive measures versus poor measures makes all the difference.

Acknowledgments
The authors thank Monica Capra, David Huffman, Thomas Kniesner, Charlie Sprenger, Nora Szech, Alastair Wilson, and participants at seminars at the Economic Science Association North American Meeting–Los Angeles, Southwest Experimental and Behavioral Economics Workshop, University of Pittsburgh, and the University of San Francisco for valuable comments.

Endnotes

[1] For example, compare Andreoni et al. (2015) with Dave et al. (2010).

[2] In Online Appendix B, we also explore the measures' stability across methods and over time.

[3] An important and related paper that deserves mention is Falk et al. (2016). They present the product of an extensive comparison of elicitation methods. For time preferences, risk preferences, and many other preferences, the researchers compared roughly 30 different measures in their comparative ability to predict behavior in incentivized-choice experiments. The paper presents the winning measures and their properties, but it does not present the comparative analysis.

[4] The CTB method is a notable exception as it is intended to simultaneously measure utility curvature along with the empirical discount rate.

[5] Daly et al. (2009) have a related study where they conduct factor analysis on many constructs related to the economic concept of time preference, but they do not include measures of a discount factor or a discount rate.

[6] This holds if consumption does not change over time and the monetary quantities are relatively small.

[7] In principle, we could use a risk-preference measure for that individual from another survey (if it exists), but then we would have to make an arbitrary choice regarding which risk-preference measure to use if multiple ones were available.

[8] The functional form for the CTB is provided (Andreoni and Sprenger 2012), fortunately leaving zero degrees of freedom to us.

[9] In several other places in the paper, we report FDR q-values. The FDR q-values are computed using the full set of statistical tests for which we report a q-value.

[10] The main purpose of controlling for risk preferences in the time-preference regressions and vice versa is to address omitted-variable bias. As discussed in Section 2.2, there is reason to expect that time preferences and risk preferences are correlated. Without controlling one for the other, the estimated effects could be biased.

[11] When we have multiple measures of FICO for the same individual, we take the average.

[12] We use maximum likelihood to estimate the covariance matrix and then use the matrix for factor extraction. Graham (2009) argues that the expectation-maximization method is the preferred approach for factor analysis with missing data. Kamakura and Wedel (2000) show with both synthetic data and real marketing data, maximum likelihood methods perform well, even when up to 60% of the data are missing.

[13] Whether this factor is distinct from other characteristics may be called into question given that the factor loads on FICO. However, FICO, a person's credit score, is an endogenous outcome that is predicted by theory to be related to time preference. Theory predicts that more patient people will be more likely to pay off loans, because the cost is in the present, but the benefits are in the future, thereby achieving better credit scores.

[14] We use survey 189 because it has several strengths: (1) it is one of the few surveys that has both time and risk preferences, (2) the sample size is relatively large, and (3) both measures predict well.

References

Ahn T (2010) Attitudes toward risk and self-employment of young workers. *Labour Econom.* 17(2):434–442.

Andersen S, Harrison G, Lau M, Rutström EE (2006) Elicitation using multiple price list formats. *Experiment. Econom.* 9:383–405.

Andersen S, Glenn W, Harrison MIL, Rutström EE (2008) Eliciting risk and time preferences. *Econometrica* 76(3):583–618.

Anderson LR, Mellor JM (2008) Predicting health behaviors with an experimental measure of risk preference. *J. Health Econom.* 27(5):1260–1274.

Andreoni J, Kuhn MA (2019) Is it safe to measure risk preferences? Assessing the completeness, predictive validity, and measurement error of various techniques. Working paper, University of California, San Diego, San Diego.

Andreoni J, Sprenger C (2012) Estimating time preferences from convex budget sets. *Amer. Econom. Rev.* 102(7):3333–3356.journal

Andreoni J, Kuhn MA, Sprenger C (2015) Measuring time preferences: A comparison of experimental methods. *J. Econom. Behav. Organ.* 116:451–464.

Ashraf N, Karlan D, Yin W (2006) Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quart. J. Econom.* 121(2):635–672.

Augenblick N, Niederle M, Sprenger C (2015) Working over time: Dynamic inconsistency in real effort tasks. *Quart. J. Econom.* 130(3):1067–1115.

Barsky RB, Juster FT, Kimball MS, Shapiro MD (1997) Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *Quart. J. Econom.* 112(2):537–579.

Beauchamp J, Cesarini D, Johannesson M (2017) The psychometric and empirical properties of measures of risk preferences. *J. Risk Uncertainty* 54:203–237.

Berg J, Dickhaut J, McCabe K (2005) Risk preference instability across institutions: A dilemma. *Proc. National Acad. Sci. USA* 102(11): 4209–4214.

Blais A-R, Weber EU (2006) A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgmental Decision Making* 1(1): 33–47.

Blumenstock J, Callen M, Ghani T (2018) Why do defaults affect behavior? Experimental evidence from Afghanistan. *Amer. Econom. Rev.* 108(10):2868–2901.

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

Borghans L, GolsteynBHH (2006) Time discounting and the body mass index: Evidence from the Netherlands. *Econom. Human Biology* 4(1):39–61.

Bradford D, Courtemanche C, Heutel G, McAlvanah P, Ruhm C (2017) Time preferences and consumer behavior. *J. Risk Uncertainty* 55:119–145.

Burks S, Carpenter J, Götte L, Rustichini A (2012) Which measures of time preference best predict outcomes: Evidence from a large-scale field experiment. *J. Econom. Behav. Organ.* 84(1):308–320.

Burks SV, Carpenter JP, Goette L, Rustichini A (2009) Cognitive skills affect economic preferences, strategic behavior, and job attachment, *Proc. National Acad. Sci. USA* 106(19):7745–7750.

Chabris CF, Laibson D, Morris CL, Schuldt JP, Taubinsky D (2008) Individual laboratory-measured discount rates predict field behavior. *J. Risk Uncertainty* 37(2-3):237–269.

Chapman J, Dean M, Ortoleva P, Snowberg E, Camerer CF (2018) Econographics. NBER Working Paper 24931, National Bureau of Economic Research, Cambridge, MA.

Charness G, Gneezy U, Imas A (2013) Experimental methods: Eliciting risk preferences. *J. Econom. Behav. Organ.* 87:43–51.

Chuang Y, Schechter L (2015) Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *J. Development Econom.* 117:151–170.

Condon DM, Revelle W (2014) The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence* 43:52–64.

Courtemanche C, Heutel G, McAlvanah P (2015) Impatience, incentives and obesity. *Econom. J. (London)* 125(582):1–31.

Crosetto P, Filippin A (2016) A theoretical and experimental appraisal of four risk elicitation methods. *Experiment. Econom.* 19(3): 613–641.

Daly M, Harmon CP, Delaney L (2009) Psychological and biological foundations of time preference. *J. Eur. Econom. Assoc.* 7(2–3): 659–669.

Dave C, Eckel CC, Johnson CA, Rojas C (2010) Eliciting risk preferences: When is simple better? *J. Risk Uncertainty* 41(3):219–243.

Dean M, Ortoleva P (2015) Is it all connected? A testing ground for unified theories of behavioral economics phenomena. Preprint, submitted August 14, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2643355.

Dohmen T, Falk A, Huffman D, Sunde U (2010) Are risk aversion and impatience related to cognitive ability? *Amer. Econom. Rev.* 100(3):1238–1260.

Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG (2011) Individual risk attitudes: Measurement, determinants, and behavioral consequences. *J. Eur. Econom. Assoc.* 9(3):522–550.

Eckel CC, Grossman PJ (2008) Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *J. Econom. Behav. Organ.* 68(1):1–17.

Eisenhauer JG, Ventura L (2006) The prevalence of hyperbolic discounting: Some European evidence. *Appl. Econom.* 38:1223–1234.

Falk A, Becker A, Dohmen T, Enke B, Huffman D, Sunde U (2018) Global evidence on economic preferences. *Quart. J. Econom.* 133(4):1645–1692.

Falk A, Becker A, Dohmen T, Huffman D, Sunde U (2016) The preference survey module: A validated instrument for measuring risk, time, and social preferences. IZA Discussion Paper No. 9674, Bonn, Germany.

Frey R, Pedroni A, Rui M, Rieskamp J, Hertwig R (2017) Risk preference shares the psychometric structure of major psychological traits. *Sci. Adv.* 3(10):e1701381.

Galizzi MM, Machado SR, Miniaci R (2016) Temporal stability, cross-validity, and external validity of risk preferences measures: Experimental evidence from a UK representative sample. Working paper.

Gaudecker H-MV (2015) How does household portfolio diversificaiton vary with financial literacy and financial advice. *J. Finance* 70(2):489–507.

Goda GS, Levy MR, Manchester CF, Sojourner A, Tasoff J (2019) Predicting retirement savings using survey measures of exponential-growth bias and present bias. *Econom. Inquiry* 57(3):1636–1658.

Graham JW (2009) Missing data analysis: Making it work in the real world. *Annu. Rev. Psych.* 60:549–576.

Heckman JJ, Stixrud J, Urzua S (2006) The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *J. Labor Econom.* 24(3):411–482.

Hertwig R, Wulff DU, Mata R (2019) Three gaps and what they may mean for risk preference. *Philosophical Trans. Royal Soc. London B Biological Sci.* 374(20180140).

Holt CA, Laury SK (2002) Risk aversion and incentive effects. *Amer. Econom. Rev*. 92(5):1644–1655.

Hung AA, Parker AM, Yoong JK (2009) Defining and measuring financial literacy. RAND Labor and Population Working Paper Series.

Kamakura WA, Wedel M (2000) Factor analysis and missing data. *J. Marketing Res.* 37:490–498.

Kaur S, Kremer M, Mullainathan S (2015) Self-control at work. *J. Political Econom.* 123(6):1227–1277.

Khwaja A, Sloan F, Salm M (2006) Evidence on preferences and subjective beliefs of risk takers: The case of smokers. *Internat. J. Industry Organ.* 24(4):667–682.

Kirby KN, Maraković NN (1996) Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bull. Rev.* 3:100–104.

Laibson D (1997) Golden eggs and hyperbolic discounting. *Quart. J. Econom.* 62(2):443–477.

Lusardi A, Mitchell O (2007) Financial literacy and retirement planning: New evidence from the Rand American Life Panel. University of Michigan Retirement Research Center Working Paper 2007-157, Ann Arbor.

Lusardi A, Mitchell OS (2014) The economic importance of financial literacy: Theory and evidence. *J. Econom. Literature* 52(1):5–44.

Mata R, Frey R, Richter D, Schupp J, Hertwig R (2018) Risk preference: A view from psychology. *J. Econom. Perspective* 32(2):155–172.

Meier S, Sprenger C (2010) Present-biased preferences and credit card borrowing. *Amer. Econom. J. Appl. Econom.* 2(1):193–210.

Menon M, Perali F (2009) Eliciting risk and time preferences in field experiments: Are they related to cognitive and non-cognitive outcomes? Are circumstances important? *Riv. Internat. Sci. Soc.* 117(3/4):593–629.

Pedroni A, Frey R, Bruhin A, Dutilh G, Ralph HJR (2017) The risk elicitation puzzle. *Natl. Human Behav.* 1:803–809.

Stango V, Yoong J, Zinman J (2017a) The quest for parsimony in behavioral economics: new methods and evidence on three. NBER Working Paper No. 23057, National Bureau of Economic Research, Cambridge, MA.

Stango V, Yoong J, Zinman J (2017b) Quicksand or bedrock for behavioral economics? Assessing foundational empirical questions. NBER Working Paper 23625, National Bureau of Economic Research, Cambridge, MA.

Tanaka T, Camerer CF, Nguyen Q (2010) Risk and time preferences: Linking experimental and household survey data from Vietnam. *Amer. Econom. Rev.* 100(1):557–571.

Truxillo C (2005) Maximum likelihood parameter estimation with incomplete data. *Proc. 30th Annual SAS (r) Users Group Internat. Conf.*, 111–130.

van Rooij M, Lusardi A, Alessie R (2011) Financial literacy and stock market participation. *J. Financial Econom.* 101(2):449–472.

Weber EU, Blais A-R, Betz NE (2002) A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *J. Behav. Decision Making* 15(4):263–290.