Geneva Risk Insur Rev
https://doi.org/10.1057/s10713-018-0024-0

ORIGINAL ARTICLE

Welfare effects of insurance contract non-performance

Glenn W. Harrison[1,2,3] · **Jia Min Ng**[2]

Received: 1 March 2017

Abstract Non-performance lies at the heart of much of the regulation that insurance companies face. Consumers' concerns about non-performance of the insurance provider have also been cited as a possible explanation for low demand of microinsurance. We provide a behavioral evaluation of the welfare effects of non-performance risk. We test the hypothesis that the presence of non-performance risk negatively impacts not just take-up of insurance but more importantly the welfare of the insured. We also test if violations of the reduction of compound lotteries axiom could drive this decrease in take-up and welfare. The results show that the compound risk characteristic of non-performance risk does *not* significantly decrease the welfare of insurance choices made by individuals. This counter-intuitive result is sensitive to the structural modeling of risk preferences. If one assumes the reduction of compound lotteries axiom *does* characterize behavior towards risk, one finds evidence that non-performance risk reduces welfare for the insured. But if one correctly allows for violations in that axiom in the representation of risk preferences, which is appropriate if one is going to test for the effect of compound risk from non-performance, then the counter-intuitive result is obtained. Take-up is not a reliable proxy for welfare, and the behavioral drivers of take-up are again not the same drivers of welfare. These results provide structural behavioral insight to inform normative policy design with respect to insurance regulation.

✉ Glenn W. Harrison
gharrison@gsu.edu

Jia Min Ng
jiamin.ng@outlook.com

1 Department of Risk Management & Insurance, Robinson College of Business, Georgia State University, Atlanta, GA, USA

2 Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, GA, USA

3 School of Economics, University of Cape Town, Cape Town, South Africa

Keywords Insurance non-performance · Welfare · Behavioral economics · Risk preferences · Experiments

JEL Classification D61 · D81 · D90 · C90 · C91

An insurance contract is a promise by one party to pay the other party some money if unfortunate events lead to losses by the first party. Sometimes promises are not kept. In that case the premium is lost, and only some fraction of the claim is paid. How important is contract non-performance for the welfare effects of insurance purchase decisions? We consider the theoretical and empirical issues involved in answering that question, and provide some behavioral evidence from controlled laboratory experiments. Contract non-performance is obviously a concern of insurance regulators: the probability of non-performance is directly tied to reserving and solvency policies, as well as other issues such as misrepresentation and outright fraud.[1]

The primary theoretical model of the effects of insurance contract non-performance on demand is due to Schlesinger and Schulenburg (1987) and Doherty and Schlesinger (1990).[2] Assuming insurance purchases consistent with expected utility theory (EUT), this model establishes several core results for the simplest case in which the risk of non-performance is known.

The key theoretical result is that the demand for insurance might decrease with greater risk aversion, contrary to the standard results for traditional indemnity products. The intuition is simple enough. In the absence of the risk of non-performance, such products only reduce the final, post-claim variability of income (or income-equivalent) of the person purchasing the contract and facing the known loss contingency. Under EUT such products are always welfare-improving for the agent facing the loss contingency. But non-performance raises the possibility that the final, post-claim, *and post-performance* variability of income might be *larger* than in the *status quo*, when the product is not available or purchased. Hence, for sufficiently risk-averse EUT agents, it might be rational not to purchase the product.[3]

One empirical question, then, is whether the risk of non-performance is empirically relevant. How risk averse do EUT agents have to be in order for these rational non-purchase outcomes to be observed, and are these levels of risk aversion observed? Schlesinger and Schulenburg (1987, p. 314) provide a numerical example to suggest that this counter-intuitive result might not be practically relevant, requiring levels of risk aversion that are

[1] See Cummins et al. (1999) and (2002).

[2] Tapeiro et al. (1986) established the basic theoretical results for the actuarial determination of loading factors in the presence of non-performance risk.

[3] This result is also used in discussions of the demand for index insurance contracts by Clarke (2016). The idea of an index contract is that the insured gets coverage for an idiosyncratic risk of loss that they face that is positively correlated with some easily observed and verifiable index. Payment of a claim depends solely on outcomes with respect to the index, not with respect to outcomes that are specific to the insured. Thus the added risk of an index contract is akin to non-performance risk, but is two-sided.

implausibly high. Of course, such examples depend on reliable characterization of the risks of loss and non-performance, and these are generally poorly known.

Even if plausible levels of risk aversion for EUT agents do not reverse the usual qualitative result that insurance purchase is attractive, they might seriously mitigate the welfare gains to agents of insurance. Moreover, what happens when agents exhibit risk preferences that are not consistent with EUT, or just make "mistakes" relative to the model of risk preferences that best characterizes them? These are empirical questions, as much as theoretical questions. We show how to answer them, using controlled laboratory experiments.

Biener et al. (2017) conducted artefactual field experiments on insurance with non-performance risk. They presented subjects in the Philippines with insurance contracts that had a 10% default risk for an insurance product that had a loss probability of 30% and a premium of 50 Philippine Pesos, and asked if they wanted to purchase the product. Another product had a 0% default risk for the same loss probability and a premium of 60 Philippine Pesos. Their design does not allow a clean identification of the effect of default risk on take-up, since premium was varied as well, but the effect appears to be to *reduce* demand by at *least* 22.3%.[4] The effect of adding default risk is likely larger, since premiums were lowered as well, leading to an understatement of the pure effect of default risk. No evaluation of choices beyond take-up is provided.

No previous experiments evaluated the risk preferences of subjects in a way that would allow an evaluation of the welfare that their decisions about insurance with non-performance risk would imply.[5] We evaluate the expected welfare of one-sided contract non-performance in a simple setting in which we can control all potential confounds and yet still observe behavioral responses, a laboratory experiment. Given the importance of the issue for policies towards risk management in developing countries, and the unblinking enthusiasm of many policy-makers and non-governmental agencies for insurance in general, we make no apology for starting this evaluation in a laboratory. The confounds of field evaluations of the effects of insurance and the demand for the product make it impossible to make clean, simple evaluations of the welfare effects of the policy. Most evaluations, in fact, *only* talk about whether take-up is "too low" or "about right," with no coherent sense of what take-up is appropriate for the insured.[6] We view our laboratory experiment as a necessary precursor to informative and powerful field experiments.

[4] This is the estimate from the correct probit specification for a binary dependent variable (see their Table C1, p. 50), with numerous controls.

[5] There is a large behavioral literature on the concept of "probabilistic insurance," which shares some features with non-performance risk, but differs in important ways. We review this literature in Section 4.

[6] Many evaluations dodge the issue of the welfare effect of insurance by focusing on whether it is correlated with increased utilization of services or activities that are insured. That is not what insurance is traditionally designed to influence, and is at most a secondary benefit, or cost if negative, of insurance as a risk management instrument.

A decided advantage of undertaking a controlled experimental evaluation, whether in the laboratory or the field, is that we can investigate the structural reasons for welfare losses from decisions about insurance non-performance.[7] We say "decisions" rather than take-up, since it is possible that losses arise from not taking up the product when the individual should do so. Conversely, admitting that behavior is not always consistent, take up of the product is not even a reliable indicator of a welfare gain. In the case of insurance non-performance, the focus of theoretical attention has to be the compound risk that the contract generates. In theoretical terms this draws attention to violations of the reduction of compound lotteries (ROCL) axiom, which has been implicated in many experimental evaluations of EUT.[8] If non-performance risk has the same impact on welfare as basis risk, we would expect violations of the ROCL axiom to decrease expected welfare gain from purchase choices made on insurance (Harrison et al. 2016).

We lay out the basic theory of insurance non-performance in Section 1, identifying the role of ROCL, compound risk, risk preferences, and uncertainty in welfare evaluation. By "risk preferences" we mean both the level of risk aversion that an individual exhibits in choice behavior as well as the *type* of psychological processes underlying that level of risk aversion. To keep matters simple, we focus on EUT and rank-dependent utility (RDU) Theory. In Section 2 we set out the experimental design motivated by this theory, to allow us to identify welfare gains and losses at the individual level. A central subtlety of this design to undertake *normative* inferences, discussed by Harrison and Ng (2016), is that we must have a measure of risk preferences of the individual that is separate from the insurance choices, even if that might be viewed as *descriptively* restrictive. Results are presented in Section 3, Section 4 compares our results with related literature in "probabilistic insurance," and Section 5 draws conclusions and discusses extensions.

Our results show that the compound risk characteristic of non-performance risk significantly *decreases* the welfare of insurance choices made by individuals when we assume that individuals behave consistently with the ROCL axiom, even if they still might violate EUT. However, when we do *not* assume the ROCL axiom in our calculations of consumer surplus, the impact of these violations of the ROCL axiom by individuals on welfare when there is non-performance risk is *not* statistically significant. We can also identify demographic characteristics of individuals that appear to be correlated with less efficient choices, pointing to how normative policies might be efficiently designed. The drivers of welfare from non-performance risk are *not* the same drivers of take-up, so take-up is (yet again) not even a useful proxy for guiding policy to improve welfare in the face of this type of contract risk.

[7] References to "non-performance" henceforth refer to one-sided non-performance, in which there is some risk that a subjectively valid claim is not paid. Harrison et al. (2016) use similar methods to evaluate the welfare effects of the two-sided non-performance risk of index insurance.

[8] The ROCL axiom is that a decision-maker is indifferent between a compound lottery and the actuarially equivalent simple lottery in which the probabilities of the stages of the compound lottery have been multiplied out.

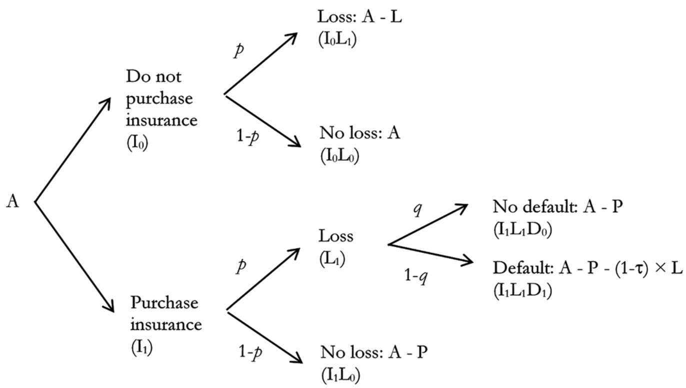

Fig. 1 Insurance purchase decision tree with non-performance risk

1 Theory

We follow the formal set-up in Doherty and Schlesinger (1990). An individual starts out with an initial endowment of A and is faced with a chance p of losing L. The individual could purchase insurance for π, which would fully compensate for the loss. There is a chance q that the insurance company will stay solvent, so there is a $(1 - q)$ chance that the insurance company would default. In the case of default, the insurance company would only repay a fixed proportion τ of the loss should the loss occur.[9] The possible monetary outcomes and their corresponding probabilities are summarized in Fig. 1.

There are five possible states, depending on the permutations of binary outcomes if the individual chooses to purchase insurance, if a loss occurs, and if the insurance company defaults. For instance, if the individual chooses not to purchase insurance, and a loss occurs, the individual would also experience a loss and be left with $A - L$. If she does not experience a loss she would keep A.

If the individual chooses to purchase insurance the outcomes are slightly more complex. If a loss does not occur she keeps her initial endowment less the premium paid, $A - \pi$. However if a loss occurs the insurance company may not be able to pay on a loss claim. If the insurance company remains solvent they can pay on the loss claim and the individual keeps her initial endowment less the premium paid, $A - \pi$. If the insurance company defaults they are only able to repay a fixed proportion τ of the loss instead of the entire loss, hence the individual only receives her initial endowment, less her insurance premium, and less the portion of the loss not covered by the insurance company when it defaults, $A - \pi - (1 - \tau) \times L$. This is the *non-performance* risk which creates a *compound* risk, hence ROCL must be considered when evaluating the welfare of insurance decisions in the presence of non-performance risk.

[9] Mahul and Wright (2004, 2007) consider the variant in which there is some probability that the indemnified claim is paid and where the probability depends on the size of the indemnified claim. This corresponds to a situation in which the insurance company suffers bankruptcy and receivers are able to repay claims on a "cents on the dollar basis" for certain, privileged claims. It also corresponds to a situation in which a government partially bails out a company, with rules for recovery rates that vary with claim value.

1.1 Evaluating welfare

Doherty and Schlesinger (1990) assume EUT, and hence assume ROCL, in their analysis of optimal insurance decisions by agents. Assuming EUT, the methodology to calculate consumer surplus (CS) is as follows. Let A denote initial wealth, L denote the loss amount, π denote the insurance premium, p denote the probability of a loss, q denote the non-performance risk of the insurer, and $\mathrm{U}(\cdot)$ denote the utility function of the individual. Utility is defined by $\mathrm{U}(A, \pi, L) = \mathrm{U}(A - \pi - L)$ if insurance is purchased at premium π and loss L occurs. The expected utility (EU) if the insurance is not purchased is $\mathrm{EU}^0 = p\mathrm{U}(A - L) + (1 - p)\mathrm{U}(A)$, and the EU if the insurance is purchased is $\mathrm{EU}^1 = [p \times (1 - q)]U[A - \pi - (1 - \tau) \times L] + (p \times q)\,\mathrm{U}(A - \pi) + (1 - p)\mathrm{U}(A - \pi)$. We can define the certainty equivalent (CE) of a lottery as the non-stochastic wealth level that is equivalent to that lottery in utility terms, so the CE of not purchasing insurance CE^0 is defined by $\mathrm{U}(\mathrm{CE}^0) = \mathrm{EU}^0$, and the CE of purchasing insurance CE^1 is defined by $\mathrm{U}(\mathrm{CE}^1) = \mathrm{EU}^1$. Expected welfare gain is measured by the CS from the option of purchasing insurance. This is the difference between the CE of purchasing insurance and the CE of not purchasing insurance: $\mathrm{CS} = \mathrm{CE}^1 - \mathrm{CE}^0$.

Doherty and Schlesinger (1990) allow for partial insurance to be purchased, so that the consumer is allowed to choose the level of coverage that would maximize EU. The choice in our experiment, however, is binary: the individual can only choose between full insurance coverage or no coverage at all. The calculated CE reveals the optimal choices. If the CE is positive, purchasing insurance would give a higher EU than not choosing insurance; if the CE is negative, having no coverage would be the optimal choice. In our experiment, we also vary premium loading, repayment percentage, loss probability, and solvency risk, each of which were considered by Doherty and Schlesinger (1990).

We relax the assumption of EUT and assume that individuals can have preferences that violate the Compound Independence Axiom (CIA).[10] We represent those preferences by the RDU model of Quiggin (1982). If we assume RDU as the decision-making model, the calculation of CS is similar once we calculate the corresponding CE values. The only complication is keeping track of how probabilities are transformed into decision weights.[11] The RDU of not purchasing

[10] The CIA states that a "constructed" *compound* lottery pair formed from two *simple* lotteries by adding a positive common lottery with the same probability to each of the simple lotteries will exhibit the same preference ordering as the simple lotteries. The CIA does not require than the decision-maker evaluate these compound lotteries by applying ROCL, so it does not imply ROCL. There is another version of the independence axiom, called the Mixture Independence Axiom (MIA), which does assume that ROCL is being used by the decision-maker when making decisions over the "constructed" compound lotteries. Many early experiments found evidence contrary to the MIA, which of course is consistent with evidence against CIA or ROCL or both. For our purposes, it is critical to design lotteries that allow us to identify pure ROCL violations, so we always refer to the CIA when talking about the independence axiom. Segal (1990) and Harrison et al. (2015, §1) have detailed formal statements of the axioms, and the latter discusses the way in which one tests them independently in experiments.

[11] In brief, the highest-ranked monetary outcome has a decision weight equal to the weighted probability, where the weighting function is yet to be defined. In our insurance choices there are only two monetary outcomes in each implied lottery. In this special case, the decision weight on the smallest-ranked monetary outcome is 1 minus the decision weight on the highest-ranked monetary outcome. The

insurance is then defined as RDU^0, and the RDU of purchasing insurance defined as RDU^1. The CE are then defined similarly, but using RDU instead of EU, so CE^0 is defined by $U(CE^0) = RDU^0$, and CE^1 is defined by $U(CE^1) = RDU^1$. The expected welfare gain is then calculated again as $CS = CE^1 - CE^0$. Since RDU^0 need not equal EU^0, and RDU^1 need not equal EU^1, and both will typically be quite different for a subject best characterized by RDU, the expected welfare gain of the option of purchasing insurance will depend on the characterization of risk preferences for the individual.

The same logic for evaluating the welfare gain extends to other variants on EUT, such as dual theory, disappointment aversion, and regret theory. We do not consider Prospect Theory, since all outcomes were in the gain domain in our experiments, but the logic extends immediately.[12]

1.2 Welfare and solvency risk

How does the CS from purchasing insurance vary as the solvency risk varies? To provide concrete illustrations, assume utility follows the constant relative risk aversion (CRRA) model so that $U(x) = x^{(1-r)}/(1 - r)$, where x is the monetary outcome and $r \neq 1$ is a parameter to be estimated. Thus r is the coefficient of CRRA under EUT: $r = 0$ corresponds to risk neutrality, $r < 0$ to risk loving, and $r > 0$ to risk aversion. Values between 0.3 and 0.7 are typical for our subjects and monetary stakes, as shown in the literature survey in Harrison and Rutström (2008, pp. 119–123).

Figure 2 shows how the CS varies as the probability of solvency varies for this insurance product, assuming the individual has EUT preferences.[13] Figures 2 and 3 both assume that the insurance decision is for an initial endowment of \$20 with a 10% chance of losing \$15, with a cost of \$1.80 to fully insure against the loss; these parameters are used in our experiment. If the insurance company defaults there is no payment in case of a claim. Regardless of the probability of solvency, CS from insurance is higher when the individual is more risk averse. This result follows because more risk averse individuals are willing to pay more for insurance. When the probability of solvency is 100%, the four most risk averse individuals shown in

---

Footnote 11 continued
probabilities of the top two monetary prizes are added prior to probability weighting, as are the probabilities of the bottom two monetary prizes. Thereafter, the RDU is evaluated as if it only had two outcomes.

12 To consider cumulative prospect theory (CPT) one would need some theory to identify losses and evidence that framed losses were evaluated as losses by our subjects even when they resulted in a net gain. One reference point might be the initial endowment A, so that the loss that is framed in our instructions and interface as a reduction of L from that endowment might then be subject to loss aversion. However, there are many other possible reference points that could be considered. We have extensive evidence, from Harrison and Swarthout (2016), that subjects in this population do not respond to framed losses of this kind in the manner assumed by CPT: they treat framed losses out of an endowment as a net gain. Therefore we did not consider CPT, although our approach extends immediately to do so.

13 We use the expression "solvency risk" when referring to the general risk of solvency or insolvency. When we need to be more precise we refer to the "probability of solvency" or "probability of insolvency," since the expression "solvency risk" is ambiguous.

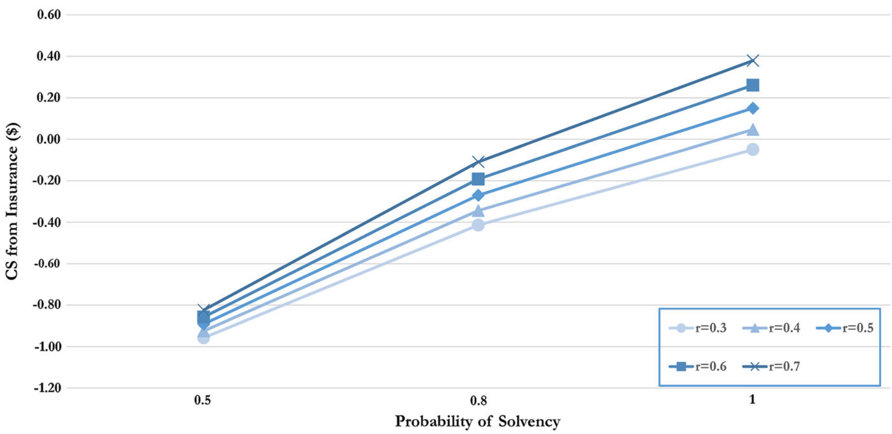

Fig. 2 Impact of solvency risk on consumer surplus assuming EUT

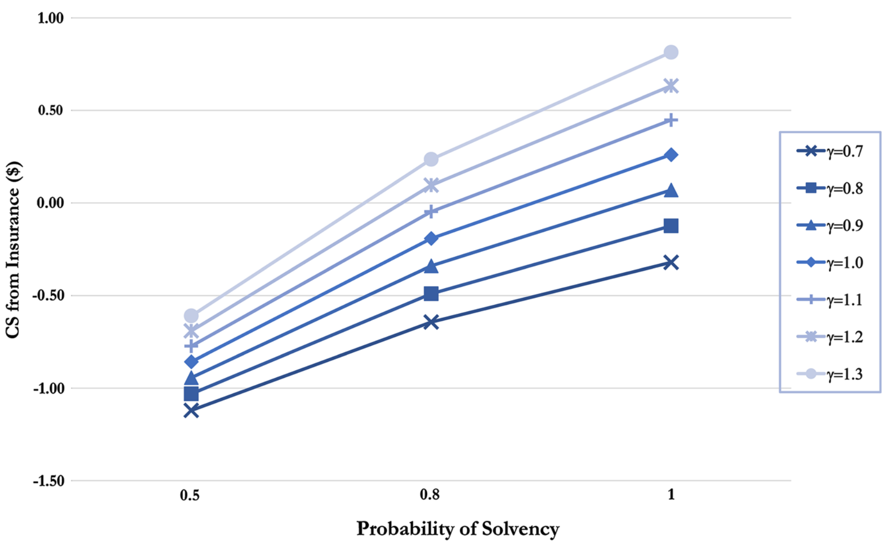

Fig. 3 Impact of solvency risk on consumer surplus assuming RDU with a power probability weighting function

Fig. 2 have a positive CS from insurance and the least risk averse individual has a negative CS from insurance. A decrease in the probability of solvency decreases the CS from insurance. With the presence of solvency risk, a decision to purchase insurance has a negative CS for sufficiently low probability of solvency (e.g., below 0.85 for the risk preferences shown in Fig. 2). Between probabilities of solvency of 0.85 and 1, the level of risk aversion determines if there is a positive CS from purchasing the insurance product or not. Clearly the risk preferences of the individual and solvency risk of the insurer can affect whether the individual's decision to purchase insurance would result in an expected welfare gain or loss.

Figure 3 shows how CS varies as the probability of solvency varies assuming an RDU decision-making model with a Power probability weighting function (pwf) given by $\omega(p) = p^{\gamma}$. In this case $\gamma \neq 1$ is consistent with a deviation from EUT. The probability weighting parameter γ spans our expected range of 0.7–1.3, and the CRRA coefficient r is held constant at 0.6. Convexity of the probability weighting function, with $\gamma > 1$, is said to reflect "pessimism" and generates, if one assumes for simplicity a linear utility function, a risk premium since $\omega(p) < p \; \forall p$ and hence the "RDU EV" weighted by $\omega(p)$ instead of p has to be less than the EV weighted by p. The converse is true for $\gamma < 1$, and is said to reflect "optimism." When there is no non-performance risk, and the probability of solvency is 100%, pessimism underweights the probability of no loss that generates a risk premium, which increases the expected welfare gain of purchasing insurance. Conversely, optimism overweights the probability of no loss and has the opposite effect of decreasing the expected welfare gain of purchasing insurance. This trend persists even in the presence of non-performance risk. Just as in the EUT model, a decrease in the probability of solvency decreases the expected welfare gain of purchasing insurance regardless of γ. Once again, not only do the probability weighting parameters impact whether the expected welfare gain is *positive or negative*, and hence whether or not the "correct" decision estimated for the individual is to purchase or not to purchase insurance, but they also affect *how much* the insurance product will or will not benefit the individual.

In our evaluation of risk preferences under RDU we also use the flexible Prelec (1998) probability weighting function $\omega(p) = \exp\{-\eta(-\ln p)^{\varphi}\}$, defined for $0 < p \leq 1$, $\eta > 0$ and $\varphi > 0$.[14] We also use the inverse-S probability weighting function $\omega(p) = p^{\gamma}/(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}$.

1.3 The normative metric for welfare evaluation

To make this theory operational, we need to make an assumption that we can indeed identify risk preferences independently of the insurance choice under evaluation. The reason is deceptively simple: in our setting there is almost always some assumption about risk preferences that can rationalize any insurance decision as generating a positive expected welfare gain.[15] It could be that the only models of

[14] When $\varphi = 1$ this function collapses to the Power function $\omega(p) = p^{\eta}$, and to EUT when $\eta = \varphi = 1$. Many apply the Prelec (1998, Proposition 1, part (B)) function with constraint $0 < \varphi < 1$, which requires that the probability weighting functions exhibit subproportionality (so-called "inverse-S" weighting). Contrary to received wisdom, many individuals exhibit estimated probability weighting functions that violate subproportionality, so we use the more general specification from Prelec (1998, Proposition 1, part (C)), only requiring $\varphi > 0$, and let the evidence determine if the estimated φ lies in the unit interval. This seemingly minor point often makes a major difference empirically. In addition, one often finds applications of the one-parameter Prelec (1998) function, on the grounds that it is "flexible" and only uses one parameter. The additional flexibility over the inverse-S probability weighting function is real, but minimal compared to the full two-parameter function.

[15] There are settings where this is not true, such as where observed insurance behavior appears to violate elementary requirements of all of the models of risk preferences we consider here, such as 1st-order

risk preferences that can rationalize certain decisions require some departure from EUT, as in Hansen et al. (2016) and Barseghyan et al. (2013) who stress the role of "probability distortions" akin to the RDU models we consider. But in order to identify an expected welfare loss, one must conceptually have some independent measure of risk preferences.

We make the simplest possible assumption here, that the risk task identifies these risk preferences for the individual, and then we use those estimated risk preferences to evaluate expected welfare gains or losses of that individual's insurance choices. An alternative assumption, of course, is that risk preferences for the same individual differ between our risk task and our insurance task, for whatever "framing" reason one might think of. This assumption might be correct, and indeed would be implied conceptually if we find, as we do, that risk preferences in the risk task do not explain every insurance choice. But note how that assumption is logically required if we are *ever* to declare some insurance purchase a mistake—we need to have some separate metric for declaring what is and is not a mistake than the choice itself.

Our statement of efficiency losses takes as given the type of risk preferences each individual employs, and uses that as the basis for evaluating welfare effects of insurance decisions: *periculum habitus non est disputandum*. One could go further and question if the RDU models themselves embody an efficiency loss for those subjects we classify as RDU. Many would argue that RDU violates some normatively attractive axioms, such as the independence axiom. Forget whether that axiom is descriptively accurate or not. If RDU is not normatively attractive then we should do a calculation of CS in which we *only* assume EUT parameters for subjects: we could estimate the EUT model and get the corresponding CRRA coefficient estimate (we would not just use the CRRA coefficient estimate from the RDU specification). Then we repeat the calculations. For subjects best modeled as EUT there is no change in the inferred CS, of course.

This suggested alternative raises many deeper issues with the way in which one should undertake behavioral welfare economics. For now, we take the agnostic view that the risk preferences we have modeled as best characterizing the individual are those that should be used, in the spirit of the "welfarism" axiom of welfare economics. Even though the alternatives to EUT were originally developed to relax one of the axioms of EUT that some consider attractive normatively, it does not

Footnote 15 continued
stochastic dominance. Or they imply a priori implausible levels of risk aversion, as in Sydnor (2010). We have concerns that many studies of naturally occurring insurance choices do not know the subjective loss probabilities that guided purchase decisions, and of course that is one good reason to start the welfare evaluation of insurance in a controlled experiment in which these can be induced. Some have attempted to measure risk preferences by observing naturally occurring insurance choices when there are controlled variations in contract features such as deductibles. But one must invariably make strong assumptions about the perception of losses: for instance, Cohen and Einav (2007, p. 746) note that "Two key assumptions—that claims are generated by a Poisson process at the individual level, and that individuals have perfect information about their Poisson claim rates—allows us to use data on (ex post) realized claims to estimate the distribution of (ex ante) claim rates." We view our controlled laboratory experiments as consistent with some of the conclusions of this behavioral insurance literature, but with far greater control of potential confounds.

follow that one is unable to write down axioms that make those alternatives attractive normatively.

We view this methodological issue as urgent, open, and important. There is a large, general literature on behavioral welfare economics, including Bernheim (2009), Bernheim and Rangel (2009), Manzini and Mariotti (2012, 2014), Rubinstein and Salant (2012), Salant and Rubinstein (2008), and Sugden (2004). Our general concern with this literature is that although it identifies the methodological problem well, none provide "clear guidance" so far to practical, rigorous welfare evaluation with respect to risk preferences as far as we can determine.

2 Experimental design

Our experimental design has two core tasks: one to elicit the risk preferences of the individual, and the other to elicit insurance choices. We varied the task eliciting insurance choices across three treatments on a between-subjects basis. In the control treatment subjects were asked for their preferences on purchasing a simple indemnity insurance product that has no chance of default. In the non-performance (NP) treatment, a different sample of subjects are offered the same insurance choices except that now there is a probability of non-performance of the insurance company. All instructions are provided in Appendix A of Harrison and Ng (2017).

2.1 Risky lottery choices

Each subject was asked to make choices for each of the 60 pairs of lotteries in the gain domain, designed to provide evidence of risk aversion as well as the tendency to make decisions consistently with EUT or RDU models. The battery is based on designs from Loomes and Sugden (1998) to test the CIA and designs from Harrison et al. (2015) to test the ROCL axiom. Each subject faced an individually randomized sequence of choices from this 60. The analysis of risk attitudes given these choices follows Harrison and Rutström (2008). The typical interface used is shown in Fig. 4, and all lottery pairs are documented in Appendix B of Harrison and Ng (2017).

The key insight of the Loomes and Sugden (1998) design is to vary the "gradient" of the EUT-consistent indifference curves within a Marschak-Machina (MM) triangle.[16] The reason for this design feature is to generate some choice patterns that are more powerful tests of EUT for any given risk attitude. Under EUT, the slope of the indifference curve within a MM triangle is a measure of risk aversion. So there always exists some risk attitude such that the subject is

[16] In the MM triangle, there are always one, two, or three prizes in each lottery that have positive probability of occurring. The vertical axis in each panel shows the probability attached to the high prize of that triple, and the horizontal axis shows the probability attached to the low prize of that triple. So when the probability of the highest and lowest prize is zero, 100% weight falls on the middle prize. Any lotteries strictly in the interior of the MM triangle have positive weight on all three prizes, and any lottery on the boundary of the MM triangle has zero weight on one or two prizes.

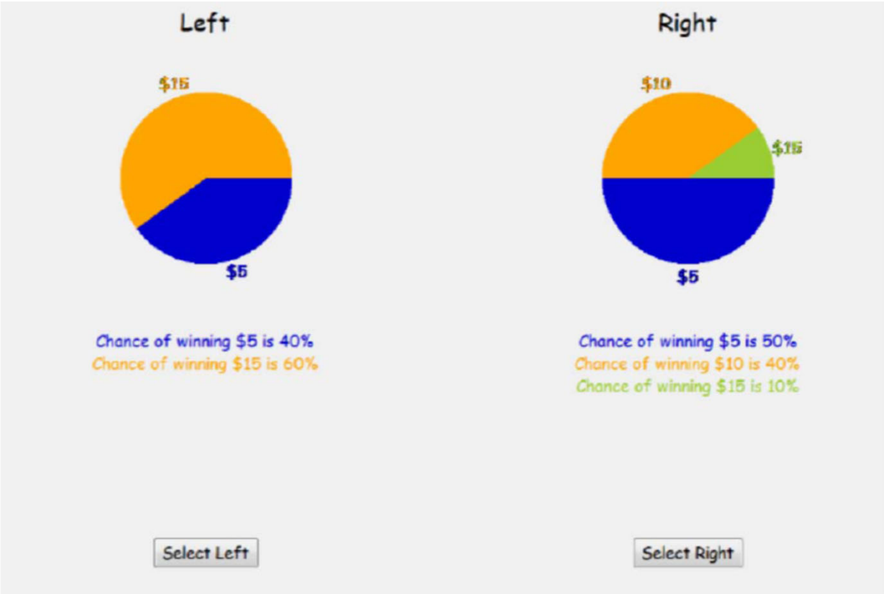

Fig. 4 Interface for risk aversion lottery choice

indifferent, as stressed by Harrison (1994), and evidence of common ratio (CR) violations in that case has virtually zero power.[17]

The beauty of this design is that even if the risk attitude of the subject makes the tests of a CR violation from some sets of lottery pairs have low power, then the tests based on other sets of lottery pairs *must* have higher power for this subject. By presenting subjects with several such sets, varying the slope of the EUT-consistent indifference curve, one can be sure of having some tests for CR violations that have decent power for each subject, without having to know a priori what their risk attitude is. Harrison et al. (2007) refer to this as a "complementary slack experimental design," since low-power tests of EUT in one set mean that there must be higher-power tests of EUT in another set.

A simple variant on these tests for a CR violation allows one to detect an empirically important pattern known as "boundary effects." These effects arise when one nudges the lottery pairs in CR and Common Consequence tests of EUT into the interior of the MM triangle, or moves them significantly into the interior. The striking claim is that EUT often performs better when one does this. Our battery replicates several of the sets of boundary CR tests originally proposed by Loomes and Sugden (1998), but also includes some lotteries moved into the interior of the MM triangle: we have 15 lottery pairs based on Loomes and Sugden (1998) and a

[17] EUT does not, then, predict 50:50 choices, as some casually claim. It does say that the expected utility differences will not explain behavior, and that then allows all sorts of psychological factors to explain behavior. In effect, EUT has *no* prediction in this instance, and that is not the same as predicting an even split.

corresponding 15 lottery pairs that are interior variants of those 15 that are "on the border."

Harrison et al. (2015) designed a battery to test ROCL by posing lottery pairs that include an explicit compound lottery and a simple (non-compound) lottery. These lottery pairs have a corresponding set of choice pairs that replace the explicit compound lottery with its actuarially equivalent simple lottery. Thus a ROCL-consistent subject would make the same choices in the first and second set. The compound lotteries are constructed by visually presenting two simple lotteries, but having some "double or nothing" option for one of them. We employ 30 lottery pairs from this battery.

2.2 Insurance choices

We are primarily interested in observing how subjects' choices vary as the non-performance risk varies across insurance choices, and how they compare to the traditional indemnity insurance product with no such risk studied by Harrison and Ng (2016). In the control treatment, subjects start with a \$20 endowment and a 10 or 20% chance of losing \$15. Each individual is offered 16 choices, displayed in Table 1, where the premium of indemnity insurance with full coverage is varied from \$0.50 to \$4.70 in 7 increments, and for each premium each individual must decide if they want to purchase insurance or not. With the loss probability 0.1 (0.2) the actuarially fair premium is \$1.50 (\$3.00). The typical interface used is shown in Fig. 5.

Table 1 Insurance contracts and parameters in the control treatment

Choice	Premium ($)	Loss probability	Initial endowment ($)	Loss ($)
1	0.50	0.1	20	15
2	1.20	0.1	20	15
3	1.80	0.1	20	15
4	2.30	0.1	20	15
5	2.90	0.1	20	15
6	3.50	0.1	20	15
7	4.10	0.1	20	15
8	4.70	0.1	20	15
9	0.50	0.2	20	15
10	1.20	0.2	20	15
11	1.80	0.2	20	15
12	2.30	0.2	20	15
13	2.90	0.2	20	15
14	3.50	0.2	20	15
15	4.10	0.2	20	15
16	4.70	0.2	20	15

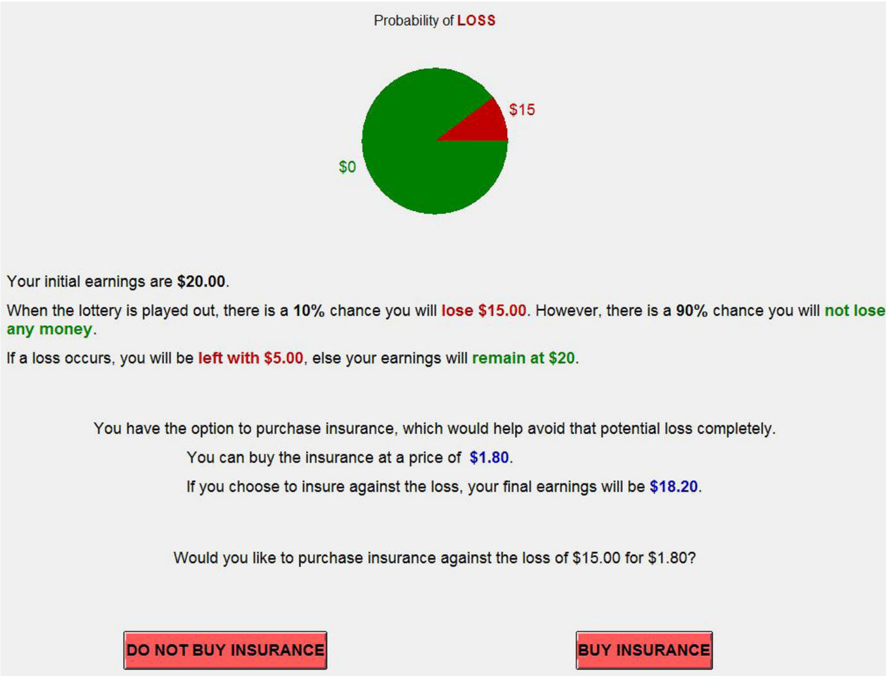

Fig. 5 Interface for insurance choice without non-performance risk

In the NP treatment, we use a $2 \times 2 \times 4 \times 2$ framework for a total of 32 choices, displayed in Table 2. The solvency probability q takes on values 0.8 or 0.5, so the non-performance probability is $0.2 = 1 - 0.8$ or $0.5 = 1 - 0.5$; the repayment proportion τ takes on the values 0 and 0.4; the premium varies over \$0.50, \$1.20, \$1.80, and \$3.50; and the loss probability is either 0.1 or 0.2. The far right column of Table 2 shows the actuarially fair premium, given the solvency probability, repayment proportion, and loss probability. The typical interface used is shown in Fig. 6. Our subjects are unable to control the *size* of the CS from an insurance purchase by selecting a level of indemnification or deductible: their only margin of choice is the binary decision to purchase or not purchase the product.

2.3 Procedures

The experiments were conducted over two sessions in January 2017 in the Experimental Economics Laboratory on the Georgia State University campus. The 40 subjects who attended the first session were given the control treatment, and the 37 subjects in the second session were given the NP treatment. All of the insurance choices came before the risk aversion task, and were presented to each individual in an individually randomized order rather than the order shown in Tables 1 and 2. Each subject was paid for one randomly selected choice from each task, and

Table 2 Insurance contracts and parameters in the non-performance treatment

Choice	Solvency probability	Repayment proportion	Premium ($)	Loss probability	Initial endowment ($)	Loss ($)	Actuarially fair premium ($)
1	0.8	0	0.50	0.1	20	15	1.20
2	0.5	0	0.50	0.1	20	15	0.75
3	0.8	0.4	0.50	0.1	20	15	1.32
4	0.5	0.4	0.50	0.1	20	15	1.05
5	0.8	0	0.50	0.2	20	15	2.40
6	0.5	0	0.50	0.2	20	15	1.50
7	0.8	0.4	0.50	0.2	20	15	2.64
8	0.5	0.4	0.50	0.2	20	15	2.10
9	0.8	0	1.20	0.1	20	15	1.20
10	0.5	0	1.20	0.1	20	15	0.75
11	0.8	0.4	1.20	0.1	20	15	1.32
12	0.5	0.4	1.20	0.1	20	15	1.05
13	0.8	0	1.20	0.2	20	15	2.40
14	0.5	0	1.20	0.2	20	15	1.50
15	0.8	0.4	1.20	0.2	20	15	2.64
16	0.5	0.4	1.20	0.2	20	15	2.10
17	0.8	0	1.80	0.1	20	15	1.20
18	0.5	0	1.80	0.1	20	15	0.75
19	0.8	0.4	1.80	0.1	20	15	1.32
20	0.5	0.4	1.80	0.1	20	15	1.05
21	0.8	0	1.80	0.2	20	15	2.40
22	0.5	0	1.80	0.2	20	15	1.50
23	0.8	0.4	1.80	0.2	20	15	2.64
24	0.5	0.4	1.80	0.2	20	15	2.10
25	0.8	0	3.50	0.1	20	15	1.20
26	0.5	0	3.50	0.1	20	15	0.75
27	0.8	0.4	3.50	0.1	20	15	1.32
28	0.5	0.4	3.50	0.1	20	15	1.05
29	0.8	0	3.50	0.2	20	15	2.40
30	0.5	0	3.50	0.2	20	15	1.50
31	0.8	0.4	3.50	0.2	20	15	2.64
32	0.5	0.4	3.50	0.2	20	15	2.10

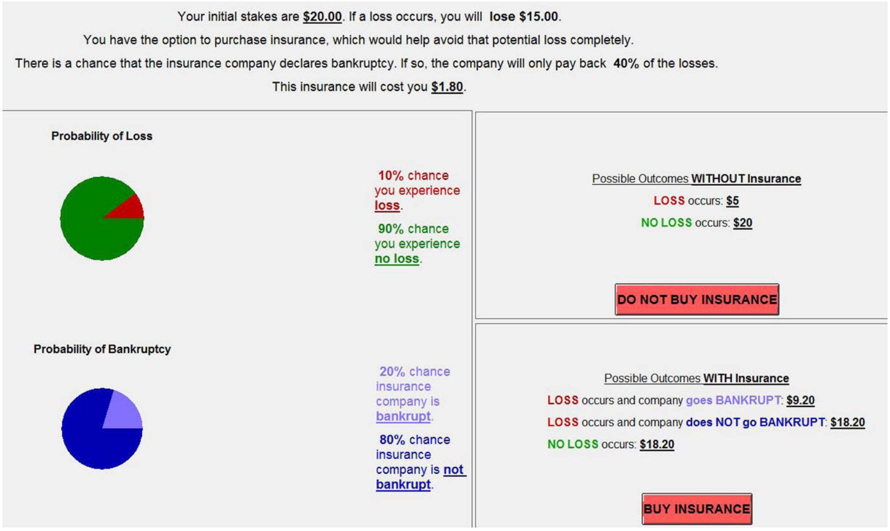

Fig. 6 Interface for insurance choice with non-performance risk

earnings for the insurance task were realized prior to the risk aversion task.[18] Average payoffs in the first (second) session were $18.01 ($17.38) for the insurance task and $27.88 ($31.62) for the risk task, for a total average payoff of $45.89 ($49.00) per subject. In addition, all subjects were given a $5 participation fee. The insurance task was programmed with the z-Tree software developed by Fischbacher (2007).

Each subject completed a survey of key demographic characteristics. The only apparent difference in sample composition is that the NP treatment had more females. This difference, as well as other smaller differences, is taken into account when later evaluating outcomes statistically.

3 Experimental evidence

3.1 Risk preferences

Overall, the proportion of model classifications as EUT or RDU is similar to previous experiments with this population. Figure 7 displays the classifications,

[18] We adopted this ordering and payment of tasks so that one could say that the estimated risk preferences were elicited independently of the insurance task. One could certainly study the empirical effect of order effects in the obvious manner, although our prior is that this is likely to be empirically unimportant. Paying subjects for the first task prior to the second task ensured that any "wealth effect" was from a known datum, rather than a subjective estimate based on prior choices. Using data from 63 subjects each making 60 binary choices in the gain domain from Harrison and Rutström (2009), one can directly test if a known initial wealth increment affects (pooled) risk preferences. In their design, each subject received a randomly generated wealth increment, between $1 and $10 in $1 increments. Using their statistical specifications, there is no significant effect of these wealth increments on estimated risk preferences using EUT or RDU models.

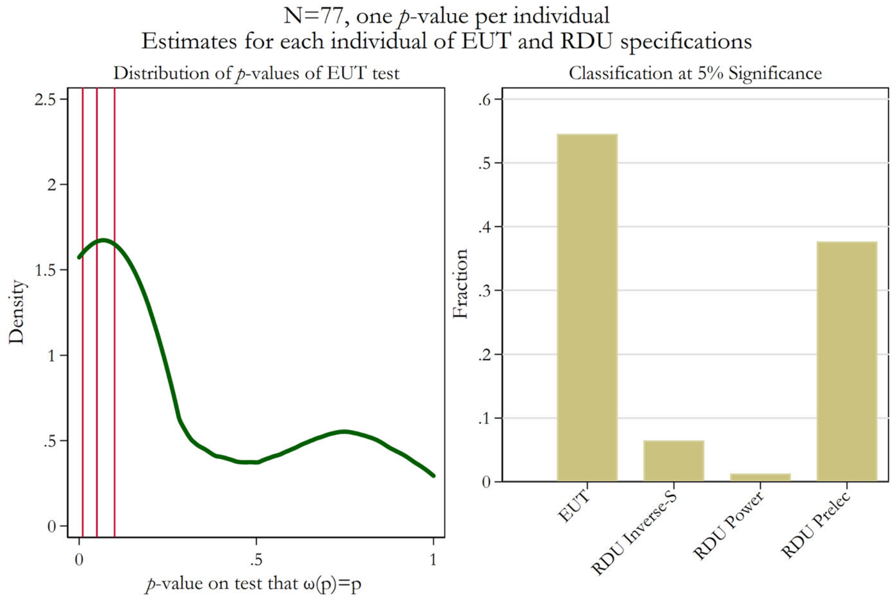

Fig. 7 Classifying subjects as EUT or RDU

based on tests of the null hypothesis that $\omega(p) = p$ and a 5% significance level. These estimates and hypothesis tests are undertaken *for each subject*. Slightly over half the subjects are classified as EUT, with the next most common model being the RDU specification with a Prelec probability weighting function.

It is important that we assign the appropriate *model* of risk preferences to each subject, since the model classification influences the expected welfare calculated for each insurance choice. To illustrate, consider subject #70. The risk parameters were estimated based on his choices on lotteries in the risk task, and are displayed in Fig. 8. If subject #70 was classified as EUT, he would be risk averse with a modestly concave utility function ($r = 0.60$). However, the preferred model is based on the log-likelihood *and* the hypothesis test that $\omega(p) = p$, and for subject #70 that preferred model is the RDU model with the Prelec probability weighting function. Classifying subject #70 as RDU (Prelec) means the utility function is more concave ($r = 0.90$), and the probability weighting function implies that the subject will overweight the better outcomes since it has a characteristically "optimistic" shape. Hence the subject would overestimate the probability of *not* experiencing a loss, and would be willing to pay a lower premium to purchase the insurance. This overweighting of the probability of no loss offsets the increase in risk aversion attributable to the more concave utility function under RDU, compared to when the risk premium is characterized entirely by curvature of the utility function under EUT.

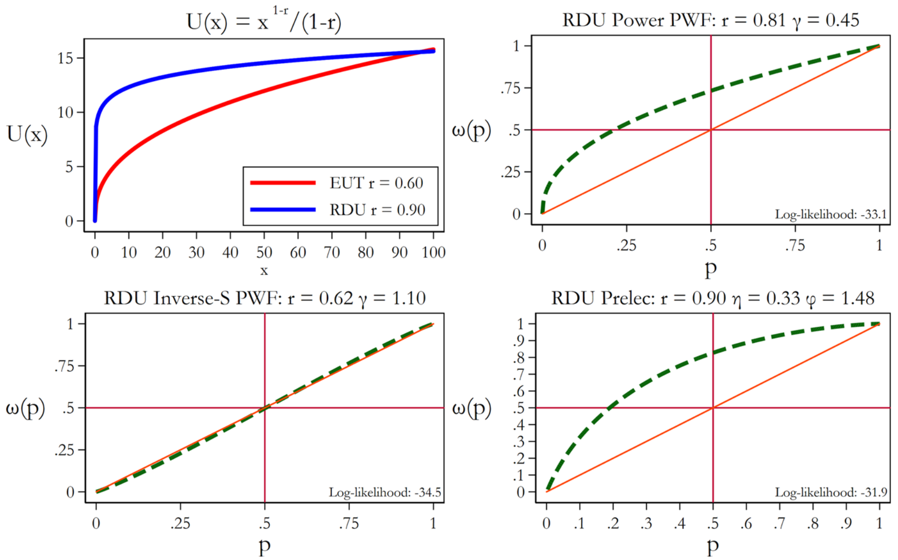

Fig. 8 Estimated risk parameters for subject #70. Subject #70 is classified RDU with EUT. p value = 0.037 (< 0.05)

Figures 9 and 10 illustrate the importance of this classification for the welfare calculations of subject #70. Each chart shows the CS calculated for each insurance choice made by subject #70. Light blue bars indicate that subject had chosen to purchase insurance and red bars indicate that subject had chosen not to purchase insurance. Figure 9 shows the CS distribution if we had assumed subject #70 had EUT risk preferences, and Fig. 10 shows the CS distribution assuming subject #70 had RDU risk preferences with Prelec probability weighting function, the preferred model. Different models of risk preference *type* can lead to *different* insurance decisions being recommended.[19]

Using a different model of risk preference *type* can also impact the *size* of the expected welfare gain from an insurance choice, and not just the sign. Choices 5 and 6 to take up insurance are *more* beneficial when subject #70 is classified as EUT compared to RDU (Prelec). Similarly choices 7 and 8 to not take up insurance are more detrimental when subject #70 is classified as EUT compared to RDU (Prelec). Again, subject #70 made one set of choices over the risky lotteries, so it is the classification of latent preferences given those choices that is driving these differences in implied CS.

[19] For choices 1 through 4 under EUT, subject #70's choices 1 and 3 to purchase insurance resulted in a positive CS, while choices 2 and 4 to not purchase insurance resulted in a negative CS. Under RDU, however, the expected welfare gains from these same choices are reversed: choices 1 and 3 resulted in a negative CS, while choices 2 and 4 resulted in a positive CS. Similarly for choices 9, 11–16 and choices 21–24: choices to purchase insurance resulted in positive CS under EUT but negative CS under RDU, while choices not to purchase insurance resulted in negative CS under EUT but positive CS under RDU.

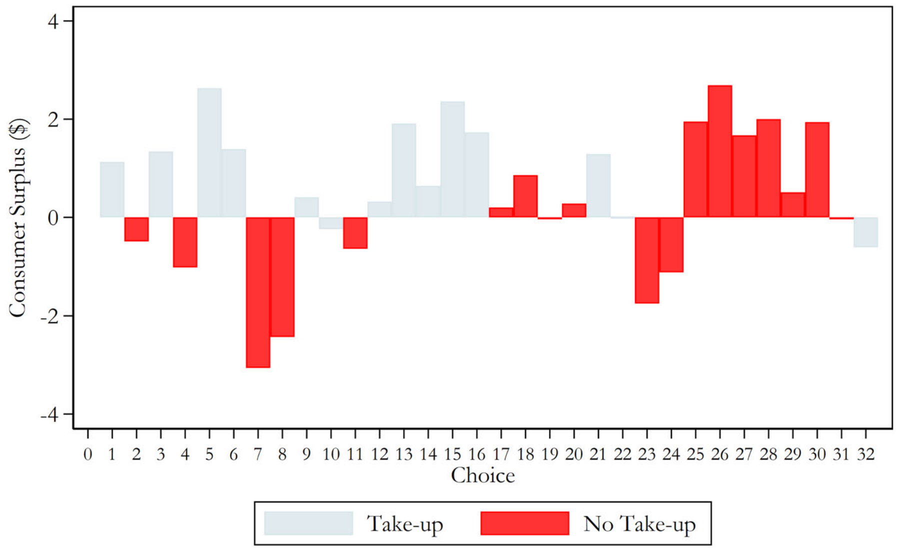

Fig. 9 Consumer surplus of choices of subject #70 expected utility theory risk preferences

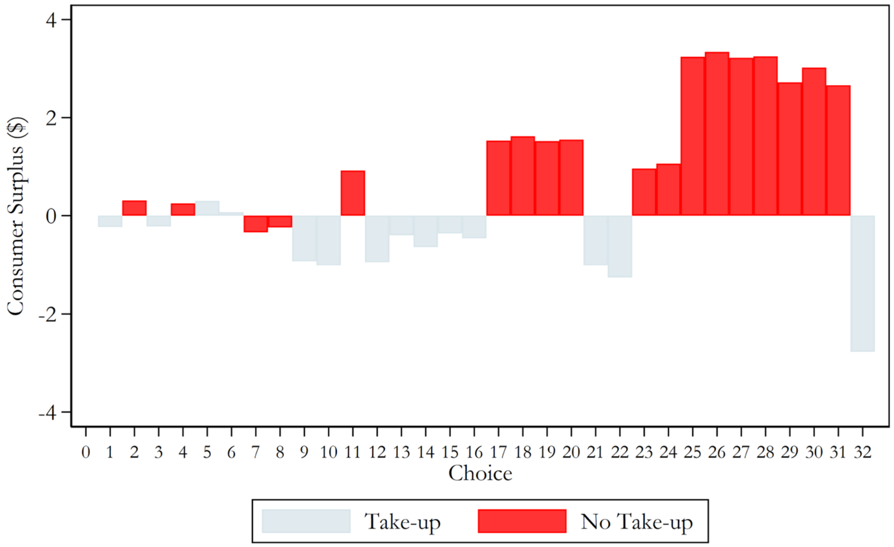

Fig. 10 Consumer surplus of choices of subject #70 rank-dependent utility (Prelec) risk preferences

The methodological lesson here is that understanding the structure of risk preferences can be essential to making the correct calculations about the sign *and* size of welfare.

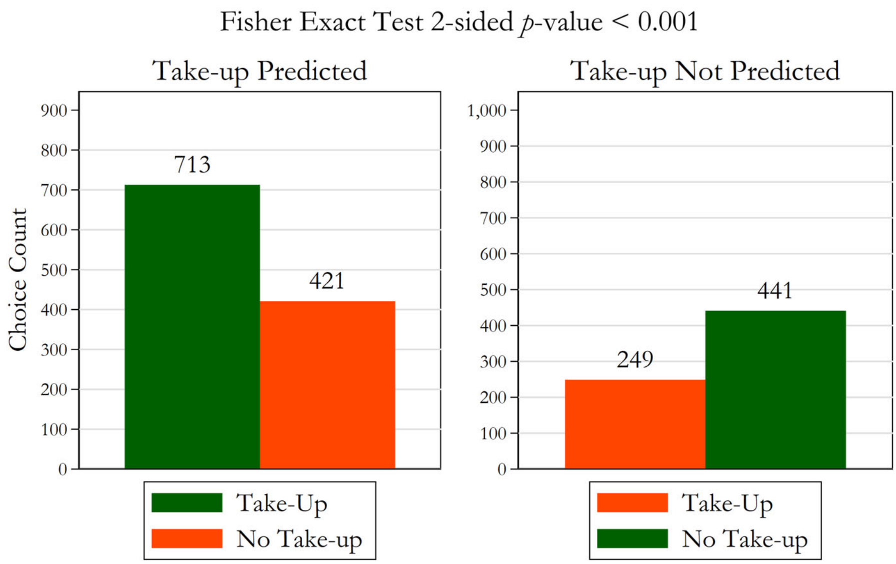

Fig. 11 Proportion of actual take-up to predicted choices for all subjects

3.2 Insurance take-up

The overall distribution of insurance choices is displayed in Fig. 11. We define a "correct" choice as one in which the subject makes the choice to purchase or not purchase the insurance product on offer that is predicted by *correctly applying the risk preferences we estimate for that subject*. In other words, if the certainty equivalent of the consumer surplus is positive when purchasing the insurance product, the "correct" decision is to purchase it; otherwise, the "correct" decision is not to purchase it.[20]

Subjects generally make the "correct" choice to purchase insurance when take-up is predicted, and to not purchase insurance when take-up is not predicted. Overall 63% of the choices are "correct" choices. There appears to be no significant pattern when the estimated risk preferences predict that the subject should not purchase insurance (the right panel). A Fisher Exact test indicates that these patterns of correct and incorrect decisions are significantly different across the two take-up predictions.

This pattern persists across our treatments, with two slight differences. The first difference is that, conditional on take-up being predicted, 72% of choices in the control treatment with no mention of non-performance risk made the "correct" choice to purchase insurance, but in the treatment with non-performance risk only 57% of the insurance choices to purchase insurance are "correct." Overall 70% of

[20] We use quotation marks for the word correct here, because our definition rests on theory and econometric inference about the risk preferences of individuals, and both of those might be wrong. But we firmly reject the view that one can determine what a correct insurance purchase decision is in the absence of some assumed theoretical and econometric structure.

choices in the control treatment are "correct," whereas only 60% of the choices in the treatment with non-performance risk are "correct." A Fisher Exact test shows that the percentage of choices that result in positive CS is significantly different across treatments. The second difference is that 60% of the control treatment choices involved the purchase of insurance, but only 49% of the non-performance treatment choices involved the purchase of insurance. Using a Fisher Exact test, this result shows that the presence of non-performance risk significantly decreases the take-up of insurance. The breakdown of distribution of insurance choices by treatment is in Appendix C of Harrison and Ng (2017).

These calculations of expected welfare are conditional on point estimates of risk preference, which in turn have estimated standard errors. We allow for these errors in the estimates and bootstrap the effects on calculated welfare. Harrison and Ng (2016, p. 110ff.) demonstrate how to allow for the sampling distribution of these estimates. Assuming a multivariate normal distribution on the risk parameters, 500 draws on the risk parameters for each individual were used to calculate the expected CS for each decision. Each decision was tested to determine if it was statistically significantly "incorrect." In other words, for decisions where insurance was actually purchased, was the expected CS significantly negative? And if insurance was not actually purchased, was the expected CS significantly positive?

Even after allowing for bootstrapping of the calculated welfare, the conclusion remains the same. Given the best-fitting decision-making model and risk preferences, a significant proportion of decisions made result in negative expected welfare gain. We actually find a stronger effect from the non-performance risk treatment. In the control we find that 73% of the choices where take-up was predicted actually occurred, whereas only 51% actually occurred in the case of non-performance risk.[21]

3.3 Consumer surplus and efficiency

The breakdown by treatment of actual choices compared to predicted choices provides an initial insight into potential welfare losses. But it does not weight these correct choices and incorrect choices: it is possible that all of the mistakes are *de minimus* in the sense that they entail minuscule losses in consumer surplus, and that the correct choices garner substantial consumer surplus, or vice versa. To address this issue we have to calculate and compare the size of the expected consumer surplus from all choices.

In Fig. 12 we compare the distribution of expected CS calculated from each insurance choice made in the control treatment to the expected CS calculated from each insurance choice made in the treatment with non-performance risk. The average CS in the control is indeed statistically significantly greater than the average CS in the treatment with non-performance risk, with a t test showing a p value < 0.01. It is important to stress that the mere existence of non-performance risk means that there is less consumer surplus possible from correct choices compared to the environment with no such risk.

[21] The comparison of actual take-up decisions to predicted decisions by treatment while allowing for bootstrapping can be found in Figures C3 and C4 in Appendix C of Harrison and Ng (2017).

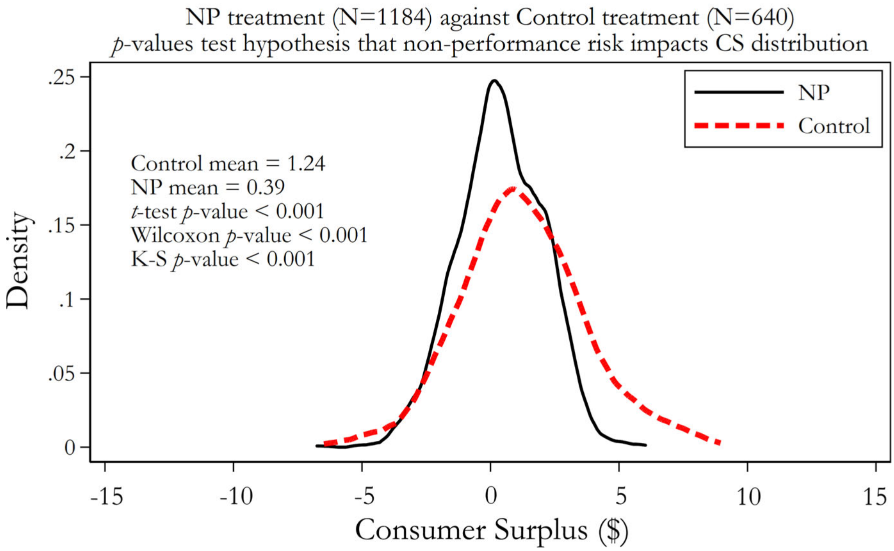

Fig. 12 Comparison of consumer surplus distribution for NP and control treatments

A more informative metric in this case is efficiency, defined as the sum of the actual CS each subject earns from all their insurance choices as a ratio of the total CS they could have earned if they had made every choice consistently with their risk preferences. The efficiency metric was developed by Plott and Smith (1978), and is defined at the level of the individual subject, whereas the expected CS is defined at the level of each choice by each subject. Efficiency provides a natural normalization of expected CS by comparing to the maximal expected CS for that choice and subject. Both metrics are of interest, and are complementary. Figure 13 displays the efficiency comparisons, with the same conclusion as with the CS comparisons: the control leads to significantly greater efficiency over the treatment with non-performance risk.

3.4 Factors affecting welfare

A regression analysis is useful in understanding what is driving the typical differences in the efficiency of insurance contracts in the presence of non-performance risk. We are interested in the impact of parameters that vary across insurance choices, which are the solvency probability, the recovery fraction, the loss probability, and the premium. We are also interested in how demographic characteristics of our subjects might influence their welfare choices.[22]

[22] Knowing which demographic groups seem to need help with these insurance decisions could help in the design of normative policies. However, this guidance need not take the form of targeting certain demographic groups, which could run afoul of social and legal anti-discrimination policies. It could be as simple as over-sampling subjects in surveys designed to assess the effects of regulatory policies, directing the dissemination of non-discriminatory information to different media, and so on.

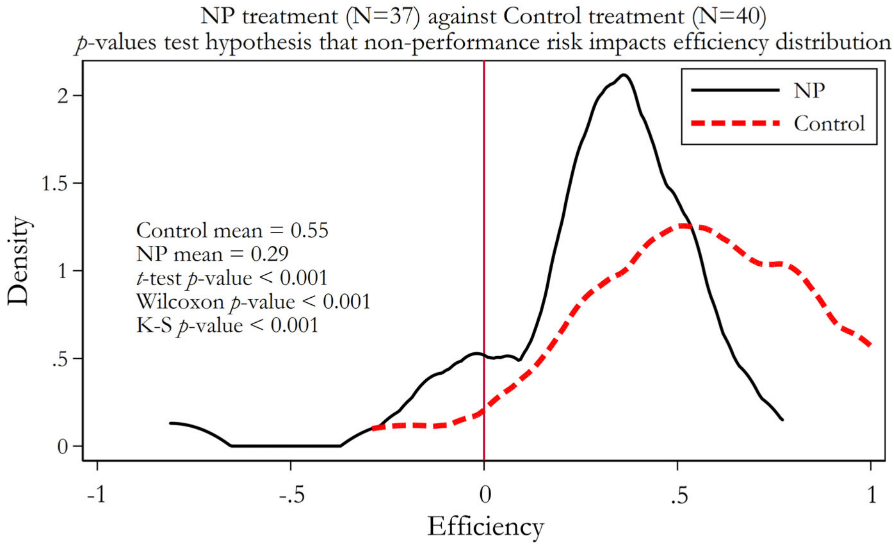

Fig. 13 Comparison of efficiency distribution for NP and control treatments

One natural characteristic to also look at is how a subject's behavior with respect to the ROCL axiom influences the welfare from choices over compound lotteries, which is what non-performance risk entails. We measure violations of the ROCL axiom non-parametrically by making use of the 15 ROCL lottery pairs in our risk battery. Each subject was given 15 lottery choices between a simple lottery and a compound lottery, as well as 15 corresponding lottery choices between the same simple lottery and a simple lottery that was actuarially equivalent to the compound lottery in the paired lottery choice. If the subject was making ROCL-consistent choices, the choices in each lottery pair would match: either choose the simple lottery in both choices or choose the compound and actuarially equivalent lottery. We count the number of pairs out of 15 that each subject does not make these ROCL-consistent choices as a measure of the degree to which each subject deviates from the ROCL axiom. This method of measuring compound risk preferences does not differentiate between compound-loving or compound risk averse preferences, and only measures if the lottery choice deviates from ROCL or not.

We use CS calculated for each insurance choice, as well as the efficiency of each subject, to estimate expected welfare gain from insurance. We also look at efficiency at the choice level (Choice), which is a binary variable indicating whether or not the "correct" choice was made to purchase insurance if it is expected to have positive welfare compared to the status quo, or not to purchase insurance if it is expected to have negative welfare compared to the status quo. Finally we also compare the results for the three welfare metrics to the results on take-up. Since Take-up and Choice are binary variables, a random effects probit model is used to measure the average marginal probability of insurance factors. Since CS is continuous, a random effects linear regression is used to measure the average marginal effect. A beta regression is applied to efficiency to measure the average marginal probability, since efficiency is a continuous variable between 0 and 1.[23]

Table 3 presents these regression results. We find that solvency probability, repayment percentage, premium, and loss probability all significantly affect take-up of the insurance product, in the a priori predicted directions.[24] However, as Fig. 11 makes apparent, take-up is most definitely not the same thing as welfare, which is what we are interested in.

The CS measure of welfare reflects a significant effect of repayment percentage, which is to be expected since repayment simply increases the potential expected payment in the event of a loss, *ceteris paribus* the premium. This is, again, why the efficiency measure is more informative for the evaluation of non-performance risk. The most striking finding is that efficiency is significantly and negatively impacted by the ROCL violation count, our proxy for each subject's *inconsistency* with the ROCL axiom. To reverse signs, for each *decrease* in the violation count, which is an *improvement* in the ROCL *consistency* of decision-making, a subject is on average 1.4% more likely to make a "correct" choice (p value $= 0.040$) that increases CS of that choice by \$0.08 (p value $= 0.001$) and increases the subject's efficiency by 1.9% (p value $= 0.002$).

Our results also show that there is an effect of age and race on the efficiency of insurance choices subject to non-performance risk. Younger subjects are more likely to make less efficient choices, as are black subjects. On the other hand, formal education, as measured by higher GPA, has no significant effect on the efficiency of decisions. Christians tend to make choices with *higher* expected welfare gain when faced with non-performance risk. This result, for *downside* non-performance risk, stands in stark contrast to the significantly *worse* decisions Christians make in both CS and efficiency terms when faced with *symmetric* non-performance risk of the type found in an index insurance contract (see Harrison et al. 2016; Table F5).[25] These same results hold when we consider the marginal effects at the mean of the covariates, instead of the average marginal effects.

3.5 Recursive methodology

Non-performance risk is a compound risk, as it is the risk of the insurance company defaulting in the event there is a payout in addition to the risk of a loss occurring. When the decision to purchase insurance involves a compound risk, we should consider how an individual's adherence to the ROCL axiom impacts the welfare of their insurance choices. To do so we should obviously not assume ROCL when calculating the expected welfare gain of insurance choices.

[23] Because all but one of these regression models are non-linear in the estimated parameters, it is possible for the margin, which is the derivative of the prediction function, to be greater than 1 due to numerical approximation.

[24] The effect of solvency probability has a p value of 0.086, and the other factors have much lower p values of 0.004 or less.

[25] In "carrot and stick" terms, it is as if the threat of eternal damnation by itself has a greater behavioral effect on motivation than the threat of eternal damnation with the risky promise of eternal salvation.

Table 3 Factors affecting welfare with non-performance risk

	Take-up	Choice	Consumer surplus	Efficiency
Risk aversion	− 0.0575	− 0.0122	0.0733	0.0287
	(0.622)	(0.813)	(0.667)	(0.586)
(Risk aversion)^2	− 0.0139	0.00411	0.0313	0.0128
	(0.820)	(0.869)	(0.731)	(0.627)
Solvency probability	0.196	0.141	0.374	
	(0.091)	(0.157)	(0.428)	
Repayment percentage	0.272**	0.104	0.784*	
	(0.004)	(0.228)	(0.027)	
Premium	− 0.0790***	0.0594*	0.134*	
	(< 0.001)	(0.030)	(0.046)	
Loss probability	0.855***	− 0.114	0.151	
	(< 0.001)	(0.751)	(0.915)	
ROCL violation count	0.000115	− 0.0138*	− 0.0773**	− 0.0194**
	(0.994)	(0.040)	(0.001)	(0.002)
Young	− 0.391**	− 0.0971	− 0.575*	− 0.160
	(0.003)	(0.323)	(0.020)	(0.064)
Female	0.0504	0.0368	0.241*	0.0613
	(0.649)	(0.395)	(0.022)	(0.192)
Black	0.137	− 0.0496	− 0.317**	− 0.112*
	(0.242)	(0.382)	(0.004)	(0.038)
Asian	− 0.0639	0.0707	0.148	0.0383
	(0.588)	(0.175)	(0.309)	(0.480)
Business major	− 0.108	0.0277	0.171	0.0477
	(0.259)	(0.625)	(0.116)	(0.404)
Freshman	0.0721	0.0520	0.191*	0.0575
	(0.402)	(0.213)	(0.018)	(0.196)
Senior	− 0.0659	0.0521	0.211	0.0725
	(0.474)	(0.176)	(0.060)	(0.079)
High GPA	0.153*	− 0.0179	− 0.127	− 0.0182
	(0.031)	(0.661)	(0.369)	(0.663)
Christian	− 0.225*	0.114	0.308**	0.113
	(0.017)	(0.084)	(0.003)	(0.061)
Insured	0.0726	− 0.00314	0.0206	− 0.00458
	(0.447)	(0.942)	(0.870)	(0.908)

Average marginal effects of appropriate regression models

Sample size is 1184 for take-up, choice, and consumer surplus, and 77 for efficiency

p values in parentheses

*$p < 0.05$

**$p < 0.01$

***$p < 0.001$

We relax the ROCL assumption by using two models: the source-dependent EUT model used in Harrison et al. (2015) and the recursive RDU model from Segal (1988, 1990). We use a source-dependent EUT model that allows for an individual to have one risk attitude for a simple lottery and a different risk attitude for a compound lottery, and a recursive RDU model that calculates the CE of the second-stage lottery before replacing the second-stage lottery with the CE to calculate the CE of the first stage lottery. A more detailed explanation of these two methods of calculating expected welfare gain can be found in Harrison et al. (2016).

3.5.1 Risk preferences

Figure 14 shows the classification of individuals using the two models. This figure should be compared with Fig. 7 where ROCL is assumed. A higher proportion of subjects are classified as source-dependent EUT (sdEUT). Only 13% of subjects are classified as recursive RDU (rRDU) with the Prelec probability weighting function. Since the sdEUT model cannot be nested in the rRDU model, non-nested hypothesis tests such as the Vuong test and Clarke test were used to determine if the sdEUT or rRDU model was a better fit.

For those subjects classified as sdEUT, we also tested the impact of using the recursive methodology instead of the standard methodology. Assuming all subjects were classified as EUT, we tested if the risk aversion parameter for compound risks was equal to the risk aversion parameter for simple risks. Only 9% of subjects had choices that used significantly different levels of risk aversion for simple risks and compound risks at a 5% significance level, reflecting "source dependence." Hence

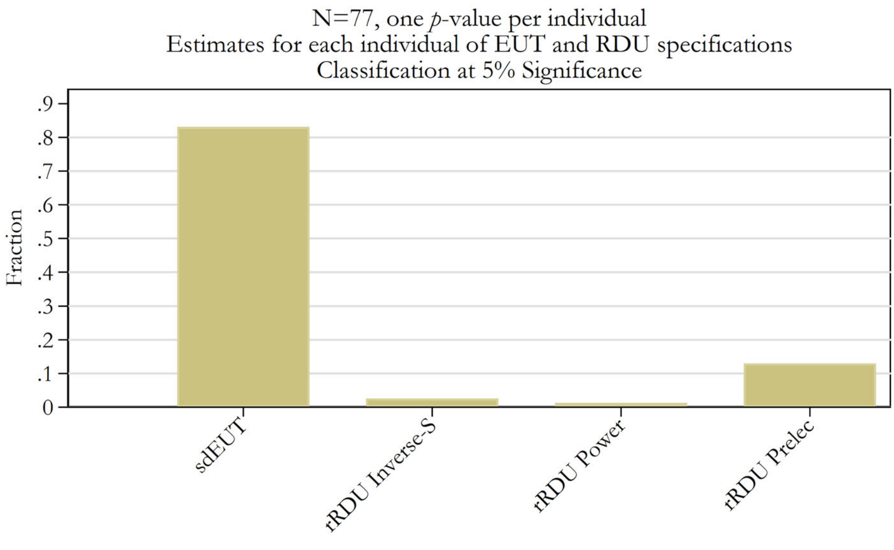

Fig. 14 Classifying subjects as source-dependent EUT or recursive RDU without assuming ROCL

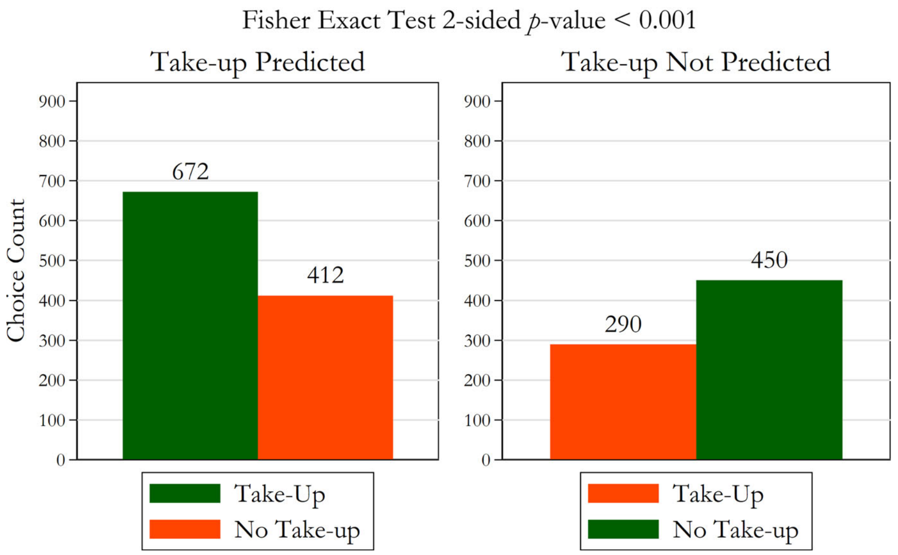

Fig. 15 Proportion of actual take-up to predicted choices for all subjects without assuming ROCL

the sdEUT model effectively collapses to the EUT model for the majority of subjects.

3.5.2 Comparison of insurance take-up

Relaxing ROCL in the calculation of welfare does not change our conclusions on the distribution of insurance choices (Fig. 15). The movement of insurance choice counts between bins is small, and the largest shift is from choices to take-up insurance: the number of insurance choices that matched the prediction to take-up insurance decreased by 41, from 713 to 672. Relaxing ROCL changes the "sign" of the expected welfare benefits. If the sign assuming ROCL is positive (negative) but changes to negative (positive) when relaxing ROCL, then the choice will switch from predicted to take-up (not take-up) to predicted to not take-up (take-up). The results show that when we relax the ROCL assumption there is a net shift in predicted choices from taking up insurance to not taking up insurance.

3.5.3 Comparison of consumer surplus and efficiency

When we relax the ROCL assumption, Figs. 16 and 17 show that the CS of insurance choices and efficiency of subjects' choices, respectively, are statistically significantly lower in the treatment with non-performance risk than in the control treatment. This result matches the conclusion when ROCL was assumed to calculate the expected welfare gain of purchasing insurance.

Once again we look to an individual's welfare benefits from choices on insurance to illustrate the impact of relaxing the ROCL assumption. Figures 18 and 19 show

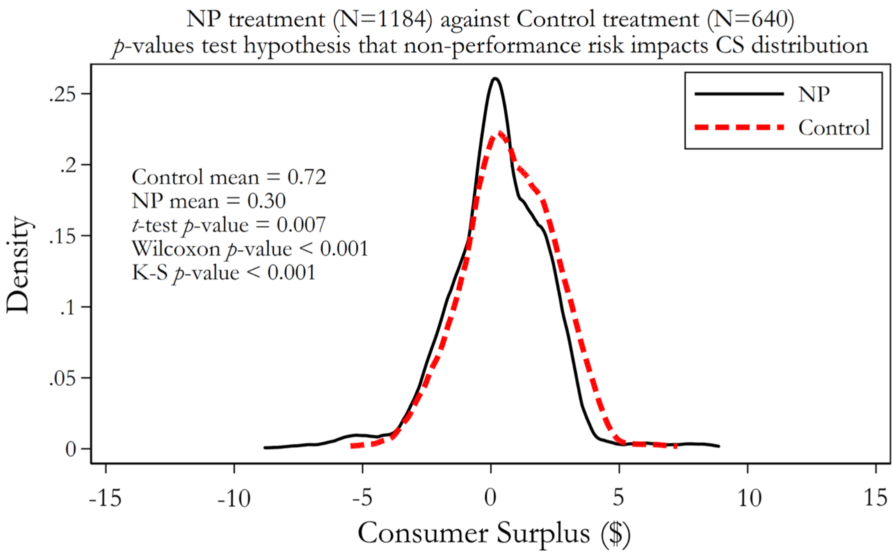

Fig. 16 Comparison of Consumer surplus distribution for NP and control treatments, without assuming ROCL

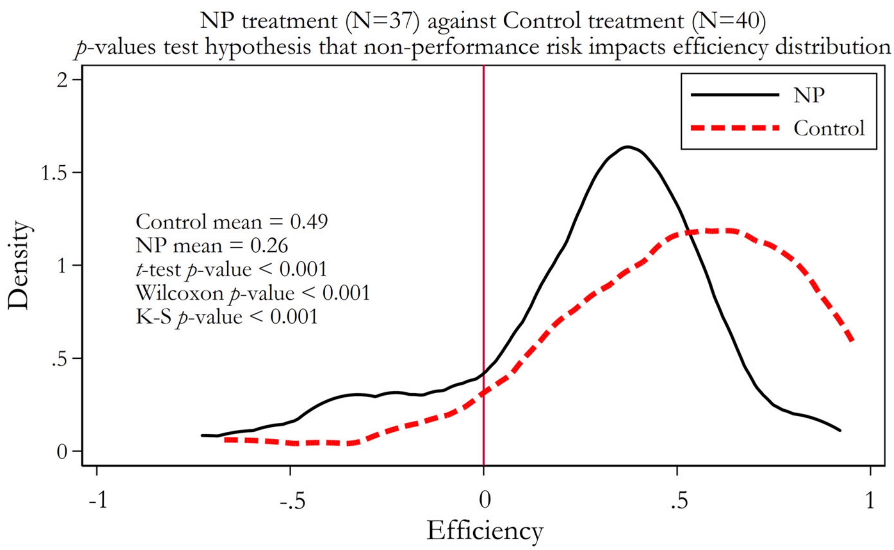

Fig. 17 Comparison of efficiency distribution for NP and control treatments, without assuming ROCL

the calculated CS for each insurance choice based on the risk model estimated for subject #58 with and without the ROCL assumption, respectively. When ROCL is assumed, subject #58 is classified as EUT with risk neutral risk preferences. If we

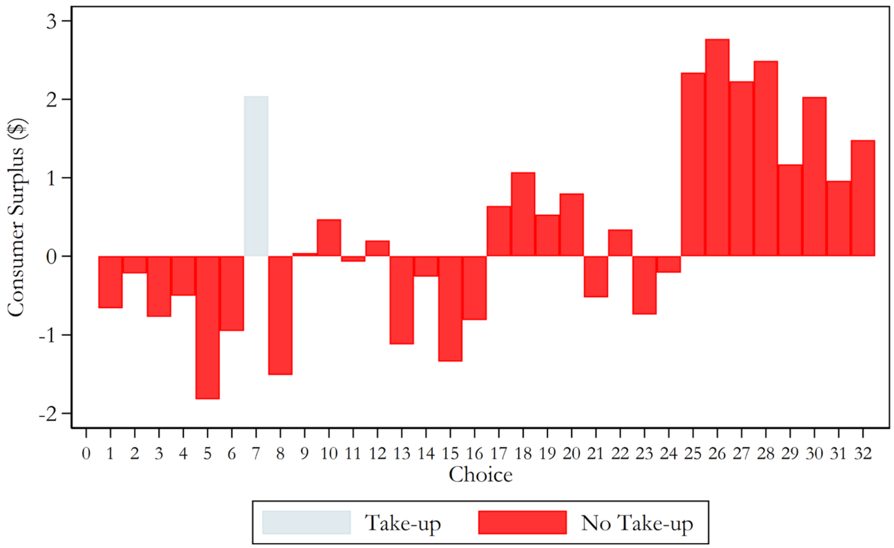

Fig. 18 Consumer surplus of choices of subject #58 expected utility theory risk preferences

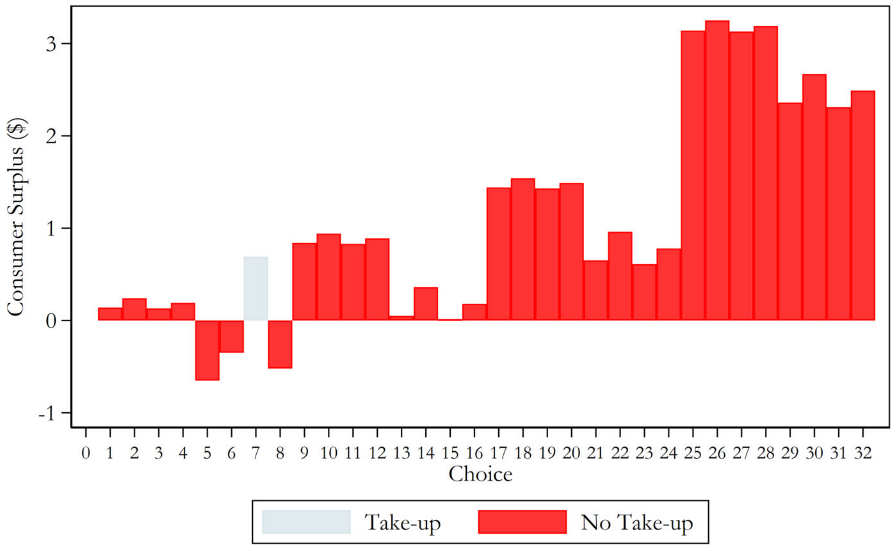

Fig. 19 Consumer surplus of choices of subject #58 recursive rank-dependent utility (Prelec) risk preferences

relax the ROCL assumption, however, subject #58 is classified as recursive RDU with a Prelec probability weighting function with a slightly concave utility function and a probability function that underweights extreme outcomes.

When we relax the ROCL assumption, the choice to not purchase insurance in choices 1–4, 11, 13–16, 21, 23, and 24 goes from being "incorrect" to "correct." For other choices, such as choices 17–20 and 29–32, we still infer that the decision to not purchase insurance resulted in positive expected welfare benefits when we relax the ROCL assumption. However, those benefits are greater for these decisions when the ROCL assumption is relaxed. This is also seen in the efficiency calculated for subject #58, which is only 0.31 if ROCL is assumed but 0.92 if it is not. Again, the persistent methodological theme here is that latent, structural theory is needed to get the correct welfare evaluations.

3.5.4 Factors affecting welfare

We can see the variation in the calculated expected welfare benefits with and without the ROCL assumption in the results of the regression analysis. Table 4 shows the regression results assuming subjects have source-dependent EUT preferences. Actuarial parameters such as repayment percentage, loss probability, and premium in the non-performance treatment still significantly impact take-up of insurance without significantly impacting the welfare of the insurance choices. When we remove the ROCL assumption to calculate expected welfare benefits, however, we see that subjects' consistency with ROCL *no longer significantly impacts the welfare measures* of "correct" choice, whether measured by the CS of choices or efficiency of subjects' choices. Age, race, and religion no longer significantly impact the efficiency of insurance choices.

Another natural characteristic of interest is a subject's attitude towards risk. We include a variable for the level of risk aversion for each subject, which is the risk parameter r, estimated assuming all subjects have CRRA utility functions and behave according to EUT. In this respect, we only use EUT descriptively, to provide a measure of the overall risk aversion of the subject, and not to claim that the subject is best characterized by EUT. These risk aversion characteristics are being considered heuristically here, since they are point estimates from a distribution and not data. For this reason, we present the results of considering them separately. The average marginal effects for all other variables were calculated excluding the point estimates for risk attitudes. When we relax the ROCL assumption, we see that greater risk aversion increases our welfare measures of the CS and efficiency of subjects' choices.

The difference in results when we use the standard methodology and when we use the recursive methodology that relaxes the ROCL axiom shows the importance of using the correct methodology to evaluate welfare. When we relax the ROCL axiom to evaluate welfare, we find that making decisions that are more consistent with the ROCL axiom no longer significantly improves welfare. Instead the level of risk aversion is now the factor that significantly increases our welfare measures. Using the inappropriate methodology could have policy implications. If the standard methodology was used to evaluate the welfare of choices made by this subject pool on insurance with non-performance risk, policies to improve welfare would have focused on educating consumers on how to make ROCL-consistent choices. Using the recursive methodology to evaluate the expected welfare gain of insurance choices with non-performance risk changes the focus to policies that promote such

Table 4 Factors affecting welfare with non-performance risk without assuming ROCL

	Take-up	Choice	Consumer surplus	Efficiency
Risk aversion	− 0.139	0.139	0.482***	0.203***
	(0.108)	(0.991)	(< 0.001)	(< 0.001)
(Risk aversion)^2	0.454	0.000857	− 0.00280	− 0.0660
	(0.058)	(0.993)	(0.992)	(0.312)
Solvency probability	0.183	0.136	0.529	
	(0.134)	(0.140)	(0.258)	
Repayment percentage	0.277**	0.0395	0.266	
	(0.005)	(0.546)	(0.449)	
Premium	− 0.0801***	0.0317	0.0920	
	(< 0.001)	(0.273)	(0.204)	
Loss probability	0.857***	0.129	1.321	
	(0.001)	(0.725)	(0.347)	
ROCL violation count	0.00434	0.0164	− 0.00313	− 0.00701
	(0.767)	(0.311)	(0.851)	(0.481)
Young	− 0.401**	− 0.0511	− 0.268	− 0.111
	(0.002)	(0.432)	(0.205)	(0.057)
Female	0.0179	− 0.0745	− 0.143	− 0.0455
	(0.878)	(0.170)	(0.093)	(0.461)
Black	0.146	− 0.0249	− 0.119	− 0.0422
	(0.223)	(0.622)	(0.189)	(0.459)
Asian	− 0.00902	0.0609	0.117	0.0511
	(0.935)	(0.205)	(0.247)	(0.311)
Business major	− 0.0802	0.0193	0.126	0.0573
	(0.415)	(0.695)	(0.092)	(0.258)
Freshman	0.0493	− 0.0179	0.103	0.0568
	(0.596)	(0.777)	(0.172)	(0.250)
Senior	− 0.109	− 0.00826	0.0973	0.0688
	(0.255)	(0.853)	(0.327)	(0.109)
High GPA	0.168*	0.0871	0.0827	0.0129
	(0.014)	(0.245)	(0.528)	(0.782)
Christian	− 0.229*	0.00575	0.0575	0.0394
	(0.011)	(0.906)	(0.567)	(0.497)
Insured	0.0743	− 0.0543	− 0.0897	− 0.0408
	(0.464)	(0.287)	(0.420)	(0.428)

Average marginal effects of appropriate regression models

Sample size is 1184 for take-up, choice, and consumer surplus, and 77 for efficiency

p values in parentheses

*$p < 0.05$

**$p < 0.01$

***$p < 0.001$

insurance to the more risk averse, and to promote other risk management strategies for the less risk averse.

4 Literature on "probabilistic insurance"

Kahneman and Tversky (1979, p. 269) introduced a concept of "probabilistic insurance," which incorporates some of the essential features of non-performance risk. Their hypothetical example also, however, allowed for the probabilistic reimbursement of the premium, which changes predictions from the pure non-performance risk case. Wakker et al. (1997) considered the case of pure non-performance risk, and showed that the same prediction under EUT applied: that a small risk of non-performance should not lead to a large change in willingness to pay for the product.

Kahneman and Tversky (1979), Wakker et al. (1997), and Zimmer et al. (2009), *inter alia*, report survey responses to suggest that subjects do not behave consistently with EUT, and appear to dislike probabilistic insurance in each of the forms proposed. Hypothetical survey responses are known to be generally unreliable, and the primary focus of these surveys was to question the empirical validity of EUT rather than evaluate the welfare effects of performance risk. Segal (1988) demonstrates that this "puzzle" from the perspective of EUT can be explained easily using RDU or Recursive RDU, where the latter is a model of Segal (1990) that allow one to relax ROCL while assuming the CIA. Wakker et al. (1997) also show that the "puzzle" can be resolved by RDU, even when probabilistic insurance is presented in the pure form.

Herrero et al. (2006) is the first experimental study to examine probabilistic insurance using real rewards applied in an incentive-compatible manner. They examined the original version of probabilistic insurance from Kahneman and Tversky (1979), and compared demand for what they refer to as no insurance (NI), full insurance (FI) and probabilistic insurance (PI). They used an elegant design to first elicit the loss probability that made an individual subject indifferent between NI and PI for the same final outcomes, but without ever framing choices in terms of NI, FI, or PI to subjects. For instance, a subject might have a choice between \$0 with probability λ and x_2 with probability $(1 - \lambda)$ or x_1 for sure. The x_1 for certain option is interpreted by the experimenters as FI, since it reflects the outcome of a full indemnity insurance contract with no deductible. Assume that the elicited probability for this subject and these outcomes is λ^*. The subject was then given a series of binary choices between NI, FI, and PI, defined simply as 3-prize lotteries:

- NI is where 0 is received with probability $\lambda^*/2$, 0 is received with $\lambda^*/2$, and x_1 is received with probability $1 - \lambda^*$.
- FI is where x_2 is received with probability $\lambda^*/2$, x_2 is received with probability $\lambda^*/2$, and x_2 is received with probability $1 - \lambda^*$.
- PI is where 0 is received with probability $\lambda^*/2$, x_2 is received with probability $\lambda^*/2$, and $(x_1 + x_2)/2$ is received with probability $1 - \lambda^*$.

Subjects were incentivized with a subtle method that is indeed incentive-compatible, although that might not seem apparent.[26] They find that subjects tend to prefer FI to PI, where the loss probability is the one that this subject revealed to make her indifferent between FI and NI, and that the subjects tend to prefer PI to NI. The preference for FI over PI is inconsistent with EUT, and consistent with RDU; but the preference for PI over NI is consistent with EUT and inconsistent with RDU. Both patterns are consistent with the Regret Theory of Loomes and Sugden (1982, 1987).

Zimmer et al. (2018) conduct the first *framed* lab experiment to examine probabilistic insurance using real rewards applied in an incentive-compatible manner. They framed the instructions in terms of insurance products, to make it easier for subjects to understand the task. As the exposition of the design of Herrero et al. (2006) illustrates, unframed experiments can seem very different than insurance decisions, perhaps disconnecting behavioral responses from field counterparts. Their experiments used a design which gave subjects a 1-in-200 chance of being paid: while a probability of 0.005 might satisfy a theorist as constituting a strictly positive probability of reward, it is surely a concern that subjects might have viewed this as "effectively hypothetical" given the low chance of being rewarded. However, if rewarded, the stakes were high: up to €800. Subjects were told that there was a 5% loss probability, and that the loss would be complete, resulting in earnings of €0. Four full indemnity contracts were offered, with non-performance risks of 0%, 1%, 2%, and 3%.[27]

Subjects were asked to state their maximum willingness to pay for the insurance contract. The Becker et al. (1964) elicitation method was used, with a well-known variation in which the random "buying price" is pre-selected and placed in an envelope. Although formally incentive-compatible, this elicitation method is widely avoided by experimental economists since subjects often fail to understand it without a great deal of hands-on training: see Plott and Zeiler (2005, p. 537). Moreover, even if subjects understand the incentives, the mechanism is known to generate *extremely* weak incentives for accurate reports: see Harrison (1992, 1994). Despite these caveats, the evidence suggests a sharp reduction in the valuation of insurance for small increases in non-performance risk, generally inconsistent with EUT.

[26] One choice was selected at random to be paid. If that choice was a direct binary choice of one lottery over another, the chosen lottery was paid out (the same method used in our experiments). If that choice was one in which the loss probability was elicited that made the subject indifferent, the subject would be paired with another subject. If the other subject reported a loss probability that was smaller (larger) than the subject being rewarded, the subject to be rewarded got to play the NI (FI) option. In effect, this is a variant on the Becker et al. (1964) elicitation method well-known to experimental economists, since the first subject does not know what probability the paired subject will state, and can only harm himself by stating a probability that is higher or lower than his true indifference probability. In practice, this logic is not obvious to subjects.

[27] The instructions did present the possibility of default with an unusually aggressive flourish, using the (translated) text: "*Default risk. 3%*, i.e., the insurer pays its valid claims in 97 out of 100 cases, and in the 3 out of 100 cases the insurer does not pay!" The use of exclamation marks is culturally specific, and Germans often use them for imperative sentences conveying simple advice. Given that the subjects were German, this exclamation mark should not be seen as biasing responses against the purchase of contracts with a default risk.

5 Conclusions

Non-performance lies at the heart of much of the regulation that insurance companies face. Our results provide a behavioral evaluation of the welfare effects of non-performance risk, keeping close to the canonical theoretical framework of Doherty and Schlesinger (1990). We hypothesize that there is a reduction in efficiency, our preferred measure of welfare in this instance, when there is non-performance risk. We stress that this measure does not just reflect the obvious fact that a less reliable product should be valued less by (risk averse) agents, *ceteris paribus*. Instead, efficiency normalizes the consumer surplus gains and losses from observed choices to naturally account for the fact that insurance that is subject to non-performance risk is a less reliable product than insurance that is free of that risk. Our behavioral comparison also matches the clean insights from theory, since in the field one would typically only see variations in non-performance risk if other factors were varied, such as reserving levels and premia.

Our behavioral evaluation goes beyond the EUT used in previous theory, by allowing for individuals to be characterized by risk preferences that relax the Compound Independence Axiom in the manner characterized by RDU. Indeed, this specification accounts for just under 45% of our sample, and makes a difference to the size and sign of consumer surplus impacts of observed choices.

We show here that the methodology used to estimate risk preferences and calculate the consumer surplus of insurance choices matters, and care must be taken to use suitable methodology. The initial result, that a violation of the ROCL axiom decreases the expected welfare gain of insurance choices with non-performance risk, is no longer significant if we relax the ROCL assumption in our estimation of risk preferences and calculation of welfare.

Four extensions to the theory and behavioral evaluation of non-performance risk would be valuable.

First, one can extend the analysis of non-performance risk to consider the implications of these risks being subjective rather than objective. Biener et al. (2017) and Liu and Myers (2016) suggest that *perceived* non-performance risk significantly decreases microinsurance take-up. In Appendix D of Harrison and Ng (2017) we extend the experimental design considered here, following Cummins and Mahul (2003), to allow the insurance company and buyers of insurance to have divergent subjective probabilities about the non-performance risk.[28] This extension further allows non-performance risk to be a probability distribution, *potentially* allowing agents to exhibit uncertainty aversion by taking into account the confidence with which they subjectively perceive that risk.[29] Cummins and Mahul (2003) note that if this weighted average of the distribution of the probability of non-performance is less (greater) than the objective non-performance probability,

[28] Contrary to Cummins and Mahul (2003, p. 121) this assumption does not require that the risks be uncertain. It could just be that the two groups have different priors or data, leading to different (posterior) subjective probabilities even if both apply Bayes rule.

[29] Following Harrison (2011, §4), we define uncertainty aversion as occurring when agents "boil down" probability distributions using some aspect of the distribution other than the weighted average: in effect, when agents do *not* apply ROCL with respect to that distribution.

we would expect the expected welfare gain from the insurance to decrease (increase). If agents employ ROCL then they evaluate such distributions as if they are a well-defined point-mass distribution exactly equal to that weighted average. Harrison and Ng (2017) assume that ROCL applies, that there is no uncertainty aversion, and that agents behave as if they use the weighted average of their subjective beliefs to make decisions about purchasing insurance. They find that allowing for subjective beliefs over non-performance risk does not significantly impact the welfare of insurance choices. One should extend this analysis of subjective beliefs to include uncertainty aversion. This would involve applying the ambiguity model developed in Klibanoff et al. (2005), for example, in the spirit of Peter and Ying (2016) and Biener et al. (2017) in the context of insurance contract non-performance.

Second, the core theorem of the probabilistic insurance thought experiment of Kahneman and Tversky (1979) does not survive generalization that relaxes the perfect asset integration assumption. That assumption is that agents treat wealth, income, and loss amounts as perfect substitutes. In this case the intuition behind the claim that EUT agents always prefer probabilistic insurance stems from the fact that risk has a second-order effect under EUT and agents prefer a sure gain (from a cheaper policy) since it has a first-order effect: see the general case for the familiar optimality-of-deductibles theorem in Gollier and Schlesinger (1996). In the theoretical framework of Section 1, this assumption involved defining $\mathrm{U}(A, \pi, L) = A - \pi - L$ if insurance is purchased at premium π and loss L occurs. However, a generalization proposed by Cox and Sadiraj (2006) allows these arguments of the utility function to be less than perfect substitutes.[30] It is easy to show that in this general case that the EUT agent does *not* always prefer probabilistic insurance to a traditional non-probabilistic full indemnity contract.[31] Hence it would be valuable to consider the welfare evaluation of insurance contracts with non-performance risk when these generalizations are allowed.

Third, our insurance contract did not allow individuals to choose the level of indemnification, other than the binary decision to purchase the product or not. Particularly when decision-making is undertaken with subjects exhibiting uncertainty aversion, such margins of choice might serve to mitigate the efficiency cost of poor decisions. Of course, the reverse is true: such margins might lead subjects to make the right choice in terms of when to purchase the product, but not to optimally indemnify.

[30] Some claim that EUT requires perfect asset integration, but this is not true. On the other hand, whether or not EUT does require this assumption is irrelevant for present purposes.

[31] For simplicity we consider the original probabilistic insurance contract proposed by Kahneman and Tversky (1979), in which the premium is returned with some probability. Consider the functional form used by Andersen et al. (2016), in which $v(A, \pi, L)$ is a constant elasticity of substitution function, and $\mathrm{U}(v)$ is the usual CRRA function $\mathrm{U}(v) = v^{(1-r)}/(1 - r)$ over the composite good. This specification allows perfect asset integration, null asset integration, and partial asset integration as special cases. Andersen et al. (2016) show that the evidence for adult Danes supports the partial asset integration case, and Harrison et al. (2017) show that the evidence for the population sampled in our experiments supports the null asset integration case. The only case in which the probabilistic insurance contract dominates is when A and L are perfect substitutes.

Finally, non-performance often reflects a breakdown of trust between insured and insurance company, or between insured and insurance regulator. Behavioral responses to trust in other agents might be processed differently than behavioral responses to objective or even subjective probabilities of non-performance,[32] and affect the efficiency of insurance decisions in the presence of non-performance risk.

Acknowledgements We are grateful for comments from J. Todd Swarthout, Martínez-Correa, participants at the *CEAR/MRIC Behavioral Insurance Workshop*, Munich 2016, and two anonymous referees.

References

Andersen, Steffen, James C. Cox, Glenn W. Harrison, Morten I. Lau, E. Elisabet Rutström, and Vjollca Sadiraj. 2016. Asset Integration and Attitudes to Risk: Theory and Evidence. *CEAR Working Paper 2011–17*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2016; forthcoming, *Review of Economics & Statistics*.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. 2013. The Nature of Risk Preferences: Evidence from Insurance Choices. *American Economic Review* 103 (6): 2499–2529.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak. 1964. Measuring Utility by a Single-Response Sequential Method. *Behavioral Science* 9: 226–232.

Bernheim, B. Douglas. 2009. Behavioral Welfare Economics. *Journal of the European Economic Association* 7: 267–319.

Bernheim, B. Douglas, and Antonio Rangel. 2009. Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *Quarterly Journal of Economics* 124: 51–104.

Biener, Christian, Andreas Landmann, and Maria I. Santana. 2017. Contract Nonperformance Risk and Ambiguity in Insurance Markets. *Working Papers on Finance No. 2017/1*, University of St. Gallen. Available at SSRN: https://doi.org/10.2139/ssrn.2908072, January 2017.

Bohnet, Iris, and Richard Zeckhauser. 2004. Trust, Risk and Betrayal. *Journal of Economic Behavior & Organization* 55: 467–484.

Clarke, Daniel J. 2016. A Theory of Rational Demand for Index Insurance. *American Economic Journal: Microeconomics* 8 (1): 283–306.

Cohen, Alma, and Liran Einav. 2007. Estimating Risk Preferences from Deductible Choice. *American Economic Review* 97 (3): 745–788.

Cox, James C., and Vjollca Sadiraj. 2006. Small- and Large-Stakes Risk Aversion: Implications of Concavity Calibration for Decision Theory. *Games and Economic Behavior* 56: 45–60.

Cummins, J. David, Neil Doherty, and Anita Lo. 2002. Can Insurers Pay for the "Big One"? Measuring the Capacity of the Insurance Market to Respond to Catastrophic Losses. *Journal of Banking & Finance* 26: 557–583.

Cummins, J. David, Martin F. Grace, and Richard D. Phillips. 1999. Regulatory Solvency Prediction in Property-Liability Insurance: Risk-based Capital, Audit Ratios, And Cash Flow Simulation. *Journal of Risk and Insurance* 66: 417–458.

Cummins, J. David, and Olivier Mahul. 2003. Optimal Insurance with Divergent Beliefs about Insurer Total Default Risk. *Journal of Risk and Uncertainty* 27 (2): 121–138.

Doherty, Neil A., and Harris Schlesinger. 1990. Rational Insurance Purchasing: Consideration of Contract Nonperformance. *Quarterly Journal of Economics* 105 (1): 243–253.

Fischbacher, Urs. 2007. z-Tree: Zurich Tool Box for Ready-made Economic Experiments. *Experimental Economics* 10 (2): 171–178.

Gollier, Christian, and Harris Schlesinger. 1996. Arrow's Theorem on the Optimality of Deductibles: A Stochastic Dominance Approach. *Economic Theory* 7: 359–363.

Hansen, Jan V., Rasmus H. Jacobsen, and Morten I. Lau. 2016. Willingness to Pay for Insurance in Denmark. *Journal of Risk and Insurance* 83 (1): 49–76.

[32] Bohnet and Zeckhauser (2004) call this "betrayal aversion," as distinct from risk aversion.

Harrison, Glenn W. 1992. Theory and Misbehavior of First-Price Auctions: Reply. *American Economic Review* 82: 1426–1443.
Harrison, Glenn W. 1994. Expected Utility Theory and The Experimentalists. *Empirical Economics* 19 (2): 223–253.
Harrison, Glenn W. 2011. Experimental Methods and the Welfare Evaluation of Policy Lotteries. *European Review of Agricultural Economics* 38 (3): 335–360.
Harrison, Glenn W., Eric Johnson, Melayne M. McInnes, and E. Elisabet Rutström. 2007. Measurement with Experimental Controls. In *Measurement in Economics: A Handbook*, ed. M. Boumans. San Diego, CA: Elsevier.
Harrison, Glenn W., Morten I. Lau, Don Ross, and J.Todd Swarthout. 2017. Small Stakes Risk Aversion in Experiments: A Reconsideration. *Economics Letters* 160: 24–28.
Harrison, Glenn W., Jimmy Martínez-Correa, Jia Min Ng, and J. Todd Swarthout. 2016. Evaluating the Welfare of Index Insurance. *CEAR Working Paper 2016-07*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2016.
Harrison, Glenn W., Jimmy Martínez-Correa, and J.Todd Swarthout. 2015. Reduction of Compound Lotteries with Objective Probabilities: Theory and Evidence. *Journal of Economic Behavior & Organization* 119: 32–55.
Harrison, Glenn W., and Jia Min Ng. 2016. Evaluating the Expected Welfare Gain from Insurance. *Journal of Risk and Insurance* 83 (1): 91–120.
Harrison, Glenn W., and Jia Min Ng. 2017. Welfare Effects of Insurance Non-Performance. *CEAR Working Paper 2017-03*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2017.
Harrison, Glenn W., and E. Elisabet Rutström. 2008. Risk Aversion in the Laboratory. In *Risk Aversion in Experiments (Research in Experimental Economics)*, vol. 12, ed. J.C. Cox, and G.W. Harrison. Bingley, UK: Emerald.
Harrison, Glenn W., and E. Elisabet Rutström. 2009. Expected Utility Theory and Prospect Theory: One Wedding and a Decent Funeral. *Experimental Economics* 12 (2): 133–158.
Harrison, Glenn W., and J. Todd Swarthout. 2016. Cumulative Prospect Theory in the Laboratory: A Reconsideration. *CEAR Working Paper 2016-05*, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, 2016.
Herrero, Carmen, Josefa Tomás, and Antonio Villar. 2006. Decision Theories and Probabilistic Insurance: An Experimental Test. *Spanish Economic Review* 8 (1): 35–52.
Kahneman, Daniel, and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47: 263–291.
Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. A Smooth Model of Decision Making Under Ambiguity. *Econometrica* 73 (6): 1849–1892.
Liu, Yanyan, and Robert J. Myers. 2016. The Dynamics of Microinsurance Demand in Developing Countries Under Liquidity Constraints and Insurer Default Risk. *Journal of Risk and Insurance* 83 (1): 121–138.
Loomes, Graham, and Robert Sugden. 1982. Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. *Economic Journal* 92: 805–824.
Loomes, Graham, and Robert Sugden. 1987. Some Implications of a More General Form of Regret Theory. *Journal of Economic Theory* 41: 270–287.
Loomes, Graham, and Robert Sugden. 1998. Testing Different Stochastic Specifications of Risky Choice. *Economica* 65: 581–598.
Mahul, Olivier, and Brian D. Wright. 2004. Implications of Incomplete Performance for Optimal Insurance. *Economica* 71 (2004): 661–670.
Mahul, Olivier, and Brian D. Wright. 2007. Optimal Coverage for Incompletely Reliable Insurance. *Economics Letters* 95 (3): 456–461.
Manzini, Paola, and Marco Mariotti. 2012. Categorize Then Choose: Boundedly Rational Choice and Welfare. *Journal of the European Economic Association* 10: 1141–1165.
Manzini, Paola, and Marco Mariotti. 2014. Welfare Economics and Bounded Rationality: the Case for Model-Based Approaches. *Journal of Economic Methodology* 21 (4): 343–360.
Peter, Richard, and Jie Ying. 2016. Optimal Insurance Demand when Contract Nonperformance Risk is Perceived as Ambiguous. *Unpublished Manuscript*, Department of Finance, Tippie College of Business, University of Iowa, October 2016.
Plott, Charles R., and Vernon L. Smith. 1978. An Experimental Examination of Two Exchange Institution. *Review of Economic Studies* 45 (1): 133–153.

Plott, Charles R., and Kathryn Zeiler. 2005. The Willingness to Pay-Willingness to Accept Gap, the 'Endowment Effect', Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review* 95 (3): 530–545.

Prelec, Drazen. 1998. The Probability Weighting Function. *Econometrica* 66: 497–527.

Quiggin, John. 1982. A Theory of Anticipated Utility. *Journal of Economic Behavior & Organization* 3 (4): 323–343.

Rubinstein, Ariel, and Yuval Salant. 2012. Eliciting Welfare Preferences from Behavioral Datasets. *Review of Economic Studies* 79: 375–387.

Salant, Yuval, and Ariel Rubinstein. 2008. (A, f): Choice with Frames. *Review of Economic Studies* 75: 1287–1296.

Schlesinger, Harris, and J.M.Graf vd Schulenburg. 1987. Risk Aversion and the Purchase of Risky Insurance. *Journal of Economics (Zeitschrift für Nationalönomie)* 47 (3): 309–314.

Segal, Uzi. 1988. Probabilistic Insurance and Anticipated Utility. *Journal of Risk and Insurance* 55 (2): 287–297.

Segal, Uzi. 1990. Two-Stage Lotteries without the Reduction Axiom. *Econometrica* 58 (2): 349–377.

Sugden, Robert. 2004. The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences. *American Economic Review* 94: 1014–1033.

Sydnor, Justin. 2010. (Over)insuring Modest Risks. *American Economic Journal: Applied Economics* 2 (4): 177–199.

Tapeiro, Charles S., Yehuda Kahane, and Laurent Jacque. 1986. Insurance Premiums and Default Risk in Mutual Insurance. *Scandinavian Actuarial Journal* 1986 (2): 82–97.

Wakker, Peter P., Richard H. Thaler, and Amos Tversky. 1997. Probabilistic Insurance. *Journal of Risk and Uncertainty* 15 (1): 7–28.

Zimmer, Anja, Helmut Gründl, Christian D. Schade, and Franca Glenzer. 2018. An Incentive-Compatible Experiment on Probabilistic Insurance and Implications for an Insurer's Solvency Level. *Journal of Risk and Insurance* 85 (1): 245–273.

Zimmer, Anja, Christian D. Schade, and Helmut Gründl. 2009. Is Default Risk Acceptable when Purchasing Insurance? Experimental Evidence for Different Probability Representations, Reasons for Default, and Framings. *Journal of Economic Psychology* 30 (1): 11–23.