THREE ESSAYS ON DECISION MAKING IN INSURANCE CHOICES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Lin Xu

August 2018

THREE ESSAYS ON DECISION MAKING IN INSURANCE CHOICES

Lin Xu, Ph.D.

Cornell University 2018

This dissertation consists of three essays on decision making in real and hypothetical insurance contexts and the factors that impact those decisions. Risky and complicated dynamic decisions are virtually unavoidable in health and property insurance contexts, and these studies help us to gain a greater understanding of how individuals respond and their underlying preferences.

In Chapter 1, I examine prescription-drug purchasing behavior of enrollees in Medicare Part D, and in particular I study the extent to which their behavior changes in anticipation of uncertain but predictable changes in insurance-coverage generosity. I develop a simple heuristic model that illustrates how enrollees ought to behave. The model predicts that any changes in spending should be smooth and occur far in advance of changes in insurance-coverage generosity (except at the end of the year). I then find empirical evidence that enrollees do anticipate and respond more optimally to the Medicare Part D pricing schedule than suggested by prior literature.

In Chapter 2, with coauthors, I conduct a novel experiment to understand how risk preferences in insurance contexts differ between the lab and the field. We employ an experimental setting to identify objective versus subjective beliefs and to evaluate how ambiguity might contribute to those beliefs. We find that, similar to prior work using field data, subjects are virtually risk-neutral and probability distortions are an important factor to explain behavior; however, there is more heterogeneity in probability distortions in the lab. We further find that subjects

respond to ambiguity by making less risky choices, and their preferences may be consistent with maxmin utility.

Finally, in Chapter 3, with coauthors, I examine the stability of risk preferences across contexts involving different stakes. Using data on households' deductible choices in three property-insurance coverages and their limit choices in two liability-insurance coverages, we assess the stability across the five contexts in the ordinal ranking of the households' willingness to bear risk. We find evidence of stability across contexts involving stakes of the same magnitude, but not across contexts involving stakes of very different magnitudes. Our results appear to be robust to heterogeneity in wealth and access to credit.

BIOGRAPHICAL SKETCH

Lin Xu was born in China and immigrated with her family as a child to the United States. She grew up in Plano, Texas, where she graduated from Plano West Senior High School. She then went on to graduate from Wellesley College in Massachusetts with a dual major in Economics and Mathematics. After college, she had a career as a software developer at Goldman Sachs in New York prior to returning to school to pursue a Ph.D. in Economics at Cornell University. She is joining the Joint Committee on Taxation of the United States Congress in Washington, D.C.

For my parents and teachers

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Ted O'Donoghue for the continuous support of my research and for his expertise, guidance, and indispensable feedback through copious iterations of my work at all stages of development. I am also grateful to Levon Barseghyan for his majestic advice and the countless hours spent under his tutelage to learn the art of estimation. Their consistent involvement and technical and editorial advice was essential to the completion of this dissertation and has taught me how to be a better researcher and economist.

I want to also thank my committee members Colleen Carey and Francesca Molinari for their investment in my intellectual growth. Colleen possesses an incredible knowledge of health economics, which she has generously shared. Francesca's encouragement and enthusiasm for econometrics has inspired me in both research and teaching.

It has been an honor to work with my co-author Josh Teitelbaum whose discerning judgment and rigor taught me lessons that I will take with me to future projects. I also want to thank my colleagues Miriam Larson-Koester, Maura Coughlin, Yang Zhang, Behavioral Economics Reading Group participants, and others who have helped me work through many modeling and data challenges throughout this process.

I am deeply indebted to my parents Shan Xu and Yanhua Han for all they have sacrificed so that I could succeed. I would not have had the opportunity to pursue my dreams without their love and unwavering support.

Last, but not least, I am incredibly fortunate to share this experience with Daniel B. Reeves who is both my partner and colleague. Since we met in graduate school, he has been my constant companion through adventures and struggles and is the first one I turn to for advice. His endless encouragement, patience, and levity

have enriched both my academic and personal life.

This work is supported in part by the National Science Foundation under grant SES-1031136.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

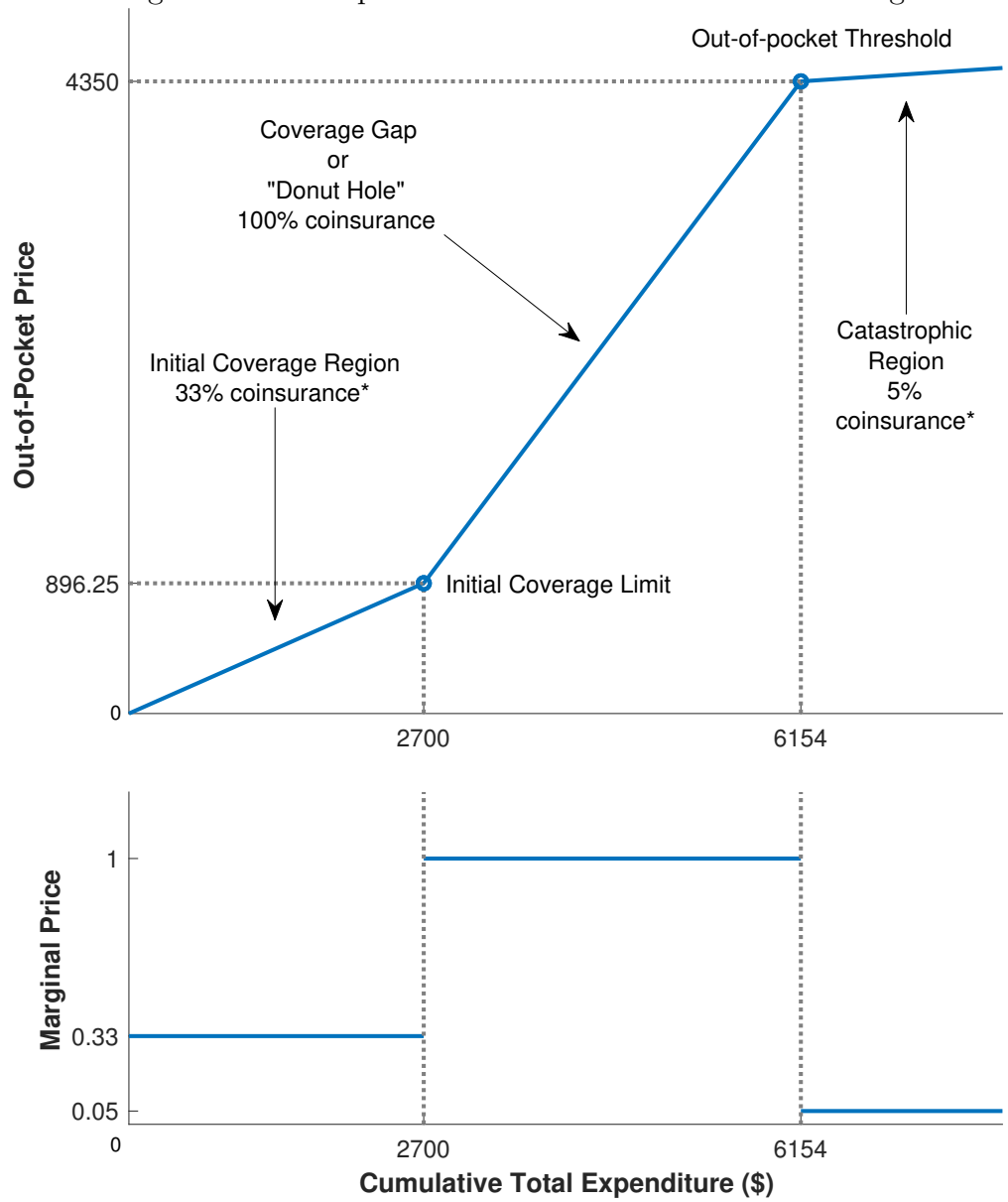
DYNAMIC PURCHASING BEHAVIOR IN HEALTHCARE CONSUMPTION

1.1 Introduction

With the rapid growth of the US healthcare sector and the accompanying financial burden it poses, the government and insurers have responded by experimenting with insurance plan characteristics and cost-sharing aimed at reducing their costs. Policy makers and academics have long had an interest in understanding how people's healthcare spending and health outcomes respond under these plans. However, it is only recently that economic research has begun to emphasize the extent of the dynamics and complexity of beneficiaries' decisions when facing multiple levels of cost-sharing within health insurance plans. This paper studies beneficiary behavior in the context of the Medicare Part D prescription drug insurance market in 2009-2012, where the government has had significant regulatory oversight on imposing plans with these features.

I focus on beneficiary responses to plans where the marginal out-of-pocket (OOP) price that the beneficiary is required to pay out of the total cost of a prescription is not constant (or non-linear) between coverage regions. Figure 1.1 illustrates an example Medicare Part D 2009 plan contract. The generosity of the plan explicitly depends on the cumulative total expenditure, i.e. the cumulative amounts that patients, the insurance company, and Medicare have spent on prescriptions within the plan-year. In this example, there is an initial coverage region (ICR) where the beneficiary is responsible for 33.2% of the total prescription costs (33.2% coinsurance), followed by a coverage gap or "donut hole" where beneficia-

Figure 1.1: Example Medicare Part D 2009 contract design



Note: The figure depicts the nonlinear structure of an example Medicare Part D benefit contract with no deductible from 2009. The plan depicted here is actuarially equivalent (with no deductible) to the government-defined contract depicted in Appendix Figure A.1. The premium or the amount the patient pays for the benefit package is not displayed. The bottom panel displays the marginal cost in each of the coverage regions, or the proportion that the beneficiary is responsible to pay of the total expenditure cost.

ries are fully responsible for costs (100% coinsurance), after which beneficiaries reach the catastrophic region and have a minimal 5% coinsurance.[1]

[1]The actual 2009 coverage benefit requires beneficiaries to pay the maximum of either 5%

See Section 1.3 for more information on the types of plans.[2] After the plan-year elapses, plans reset or beneficiaries switch to new plans, and the cumulative total spending for the new plan-year is reset to zero. Within a plan-year, beneficiaries face a truly dynamic problem—consuming healthcare today can impact the marginal price of future healthcare consumption.

It is important then to study the actual beneficiary response to these types of nonlinear pricing structures even before the actual nonlinearities. The coverage gap was initially included as a cost-saving measure for the government similar to a deductible but positioned in the middle of the patient benefit schedule (Baker, 2006). In addition to directly saving money in the donut hole, the presence of the donut hole mechanism was meant as an incentive to beneficiaries to restrain their spending even prior to the donut hole and avoid the catastrophic spending region altogether.

Much of the literature has focused on beneficiary behavior at or around the controversial coverage gap, which has been shown to have a negative effect on drug adherence. Joyce et al. (2013) and Zhang et al. (2009) find that having a coverage gap disrupts the use of prescription drugs, with a higher decline on more expensive medications as compared to cheaper ones. And while Joyce et al. (2013) fail to detect a corresponding substitution from drugs to medical treatment in concurrent Medicare claims, if one believes that adherence to drug treatments is good for patient outcomes, discontinuing the use of these drugs would have a negative welfare effect on patients.

In order to fully understand the ways in which changes to nonlinear plans can

the cost of the drug or $2.40 and $6.00 for a one-month supply of generic and branded drugs respectively. This means that patients may pay either the copay dollar amount or a percentage share of the drug price where the remainder is covered by insurance or the government.

[2]In general plans are structured similar to this example with either three or four regions.

impact beneficiary behavior and welfare, researchers must also determine whether beneficiaries respond sub-optimally to them. The current consensus in the literature is somewhat mixed. Some of the literature cites the initial lack of knowledge of the donut hole, where in 2007 only 40% of individuals knew about the coverage gap (Polinski et al., 2010). There are papers where beneficiaries appear to be somewhat optimal and fully forward-looking (Aron-Dine et al., 2015; Einav et al., 2015), and there are papers that support the conclusion that beneficiaries overly respond to the "spot" or current price (Dalton et al., 2018; Abaluck et al., 2018). As evidence of suboptimal behavior, Dalton et al. (2018) also present the stylized fact that even among individuals in their sample who were very likely to end the year in the coverage gap or beyond, there is a sharp drop in average spending on prescription purchases at the coverage gap.

Ignoring liquidity constraints, fully forward-looking optimal behavior suggests that beneficiaries should use their expected end-of-year marginal price for each purchase decision throughout the year. For example, a Medicare Part D beneficiary who fully expects to end the year in the catastrophic phase of their insurance coverage should not respond to temporary changes in their plan coverage and spot prices as they spend through earlier benefit phases. Under uncertainty about the end-of-year region and price, however, the beneficiaries may adjust their expected marginal prices as risks are realized. Assuming standard geometric discounting ($\delta >> 0$), these transitions should be smooth, especially early in the year.

In Section 1.2, because of the beneficiary's complicated optimization problem under uncertainty, this paper first approximates optimal behavior by developing a heuristic for constructing a beneficiary's perceived marginal out-of-pocket coinsurance rate. This rate is generated from the entire population's probabilities of

ending the year in each region and is meant to help visualize how the average beneficiary's expected coinsurance rate may evolve across the different weeks of the year and cumulative spending levels. It differs from the optimal beneficiary's expected prices because it uses ex-post population outcomes, which may not be representative of rational agents. Further, because beneficiaries may themselves have difficulties anticipating their cumulative total end-of-year spending and prices, the heuristic approach may provide intuitions on behavior as well. The heuristic approach predicts that, because probability distributions are quite smooth outside of the last weeks of the year, the expected marginal prices and thus spending should also be smooth with any pricing updating occurring prior to the kink.

Further, intuitions from this paper's heuristic approach indicate that errors in time discounting such as present bias may not be an appropriate model to represent beneficiary behavior that resemble a sharp spending drop at discrete changes in spot prices. A key takeaway from the heuristic model is that, outside of the last weeks of the year, only sharp discontinuities in a beneficiary's perceived marginal price or coinsurance rate should result in sharp discontinuities in beneficiary spending. As long as beneficiaries continuously update their expectations, only zero discounting would predict sharp changes at the coverage gap.

Section 1.3 provides background information on the structure of the Medicare health insurance program and details on the prescription insurance program, Medicare Part D. The section also discusses the data and sample. The data studied comes from the Center for Medicare and Medicaid Services' (CMS's) 5% sample of all beneficiaries in Medicare, with administrative information on enrollee selected plans and claims. The sample selected include enrollees, who are continuously in Medicare Part D plans without deductibles over the 2009-2012 period and who

meet criteria to make sure they do not have unique pricing or non-optional choices. Then, the heuristic from Section 1.2 is applied to the empirical data in Section 1.4, supporting the intuitions developed.

In Section 1.5, this paper uses a flexible nonparametric regression approach to provide a graphic representation of the empirically observed shape of spending patterns through the entirety of a beneficiary's plan year and across a wide range of cumulative spending levels. In part, the goal of this empirical exercise is to generate a low-assumption view of the extent of a beneficiary's anticipatory response to the different pricing regions as a function of both the time of the year and the beneficiary's cumulative total spending. The regression approach uses an individual fixed effect to control for individual heterogeneity in spending frequencies in a dynamic panel that includes four years of claims data.[3]

My estimates show that across four quarters of the year and across cumulative total spending amounts, beneficiaries have a statistically significant anticipatory response far in advance of the coverage gap. As expected, the reduction in spending in advance of the coverage gap occurs at higher cumulative total expenditure values (closer to the ICL) in later parts of the year.

This paper contributes to the existing literature studying how individuals respond to nonlinear contracts in healthcare and supports findings that people do respond to these plan structures. Previous papers had found that individuals are at least partially forward-looking in the non-linear contract context. Aron-Dine et al. (2015) find that employees that join health insurance plans in the later parts of the year (and thus have less time in their respective plans) respond to future expectations of prices of healthcare. Einav et al. (2015) show that people exhibit

[3]See Section 1.5.3 for a discussion regarding dynamic panel bias and why it is less of a concern in this setting.

bunching behavior at the donut hole and decrease their probability of spending at the entrance to the donut hole early in the year.

Our findings differ from the literature in that we do not observe the same extent of "myopic" behavior and do not detect a large discontinuity in beneficiary spending frequencies directly at the coverage gap except in the last quarter of the year. The estimation strategy in this paper is most comparable to a specification in Dalton et al. (2018), who use a fixed-effects-regression as their empirical evidence of "myopic" behavior. They analyze Medicare Part D claims of a 2008 subset of employer-sponsored Medicare Part D individuals that were likely to end the year in the coverage gap. Unlike me, they found few changes in spending in anticipation of entering the donut hole. Their analysis differs from mine in that they only had indicators to measure the level of spending response (and other dependent variables including prescription occurrence) in four cumulative total spending zones ($310 before, between $310-$110 before, $110 before, and after the coverage gap), while my paper measures a continuous response to cumulative total and cumulative out-of-pocket spending. They found no economic or statistically significant evidence of spending or claims frequency decreases in the $310-$110 leading up to the coverage gap, but they found a sharp decrease in spending and claims frequency in the $110 right before the coverage gap.

My paper's reduced-form estimates provide a simpler alternative to studying beneficiary behavior under nonlinear contracts as compared to the estimates of Einav et al. (2015) and Dalton et al. (2018). The reduced-form estimates provide a graphical explanation of why these papers found both evidence of forward-looking behavior and over-response to spot prices respectively. Einav et al. (2015) estimate a model with standard geometric discounting that allows for five types of individu-

als with different risk levels and sensitivities to the coverage gap, and they recover a weekly discount factor δ equal to 0.96 that roughly translates to a yearly discount factor of only 11%, far lower than what is accepted as standard in the literature.[4] In order to explain observed drops in spending just prior to the coverage gap, Dalton et al. (2018) estimate two models that allow for a sharp drop in spending at the coverage gap, comparing a model with salience to the gap with a model with beta-delta time-inconsistency or present bias.[5] In both models their estimates support the empirical finding of beneficiaries lacking an anticipatory spending response to the coverage gap, and in the beta-delta model, the discounting rates they estimate are indistinguishable between $\beta = \delta = 0$ indicating that beneficiaries only consider the spot price.

Another unique feature of my paper is that unlike the majority of the healthcare literature, I mainly focus on the frequency of beneficiary claims. When a beneficiary has a prescription to be filled, they have two broad decision options: wait to fill the prescription, or switch the prescription either from branded to generic or with more effort acquire an alternative prescription. The decision to stop taking a course of chronic treatment likely has more of a negative effect on beneficiary health than a decision to switch medication. Thus, this paper measures changes in claims frequency, which should capture the beneficiary's first decision to postpone or stop (postpone indefinitely) taking a course of treatment.

[4]Einav et al. (2015) refer to this estimate as a "'behavioral' parameter that also reflects individual's understanding of the insurance coverage contract, in particular the salience of the (future) nonlinearities of the contract", as part of the reason why δ is so low.

[5]Present bias, also known as hyperbolic discounting, is a form of time-inconsistent preferences where subjects show a tendency to overweigh the "present" and have self-control problems.

1.2 Intuition

In the Medicare Part D setting, patients make a combination of periodic and unexpected purchases of prescription drugs within a year. Within the decision to fill each prescription purchase, assume a beneficiary compares her costs with the perceived benefits of purchasing and then consuming drugs.

Suppose, in week w of the year, a beneficiary with observables X faces a decision whether to fill a prescription that has a total cost of s, where the total cost is to be paid between the insurer and beneficiary. The beneficiary chooses to purchase if the perceived benefit of the drug exceeds her perceived cost of the drug under the insurance plan—that is, if

$$B(s|X) > sPMC(Z_w, w|X)$$

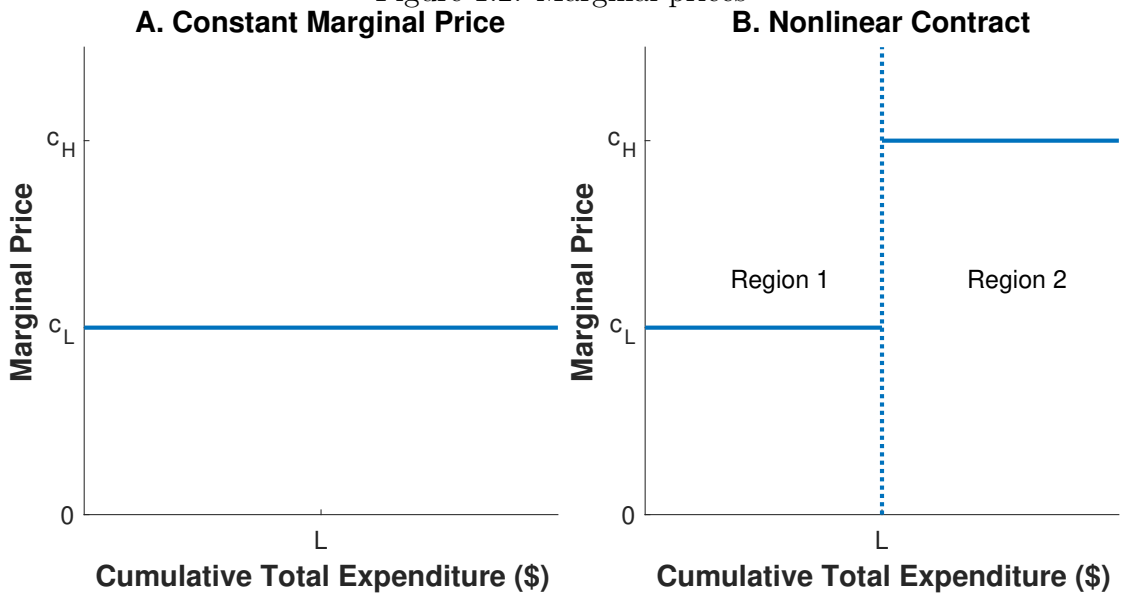
where $B(s|X)$ is the benefit of the drug, $Z_w = \sum_{u=1}^{w-1} s_u$ is the cumulative total amount spent up until week w, and $PMC(Z_w, w|X)$ is the perceived marginal cost of the drug to the beneficiary who has spent Z_w by week w conditional on their observables X.

Because of the structure of her nonlinear insurance plan, the cost a beneficiary considers is not necessarily just her out-of-pocket (OOP) cost, so her marginal cost for an additional dollar of total prescription spending is not just her coinsurance amount. Certainly she should consider her OOP cost, which in the plan is defined by how much she has cumulatively spent in the year Z_w up until that week w, but if she is forward looking, the impact that spending today has on her future costs should also enter into her decision. This paper focuses its discussion on the beneficiary's beliefs on his or her monetary costs with insurance and not on the beneficiary's beliefs on the medical benefits of consuming certain drugs. I assume

that the benefits of drug purchase and consumption do not depend directly on the arbitrary contract design.

To build intuition, consider how the beneficiary's perceived marginal cost may differ under two simple pricing contracts in Panel A and B of Figure 1.2. If the beneficiary were in a plan with contract A, she should always expect that her marginal cost of purchasing a drug would be equal to c_L, as it is the coinsurance rate applied to all purchases. It does not matter if she has to make a purchasing decision in the beginning of the year or the end, her $PMC(Z_w, w|X) = c_L$. Then her out-of-pocket (OOP) payment or the amount that she is responsible for is $c_L s$.

Figure 1.2: Marginal prices



Note: The marginal price (or cost) is the beneficiary's price for an additional dollar of prescription spending.

In contrast, Plan B is a nonlinear cost structure, and thus a beneficiary's perceived marginal price for purchasing a drug is less clear. Under this plan if $Z_w < L$, her spot price and actual out-of-pocket payment in a given week would be $MC(Z_w) = c_L$ and $c_L s$, respectively, and if $Z_w \geq L$, they would be $MC(Z_w) = c_H$ and $c_H s$, respectively. At the two extremes of behavior, a beneficiary may only

respond to the spot marginal price determined by the current insurance region, or a beneficiary may be a "fully forward-looking", perfectly rational economic agent. If she only responds to the spot price, then her $PMC(Z_w, w|X) = MC(Z_w)$. Her PMC is just her coinsurance rate in the region she is currently in, and she does not take into consideration how her spending can impact her future marginal costs. If a beneficiary is "fully forward-looking", in each purchasing period decision, she optimizes her decision by discounting the future stream of expected benefits and costs that result from her current decision, including any changes to her expected marginal prices due to the non-linear pricing. Ultimately if a beneficiary is fully forward-looking, she should anticipate her full stream of expected payments, and an additional prescription should be considered with her expected year-end marginal price. Thus, her expectation of ending in Region 1 or 2 matters. If she is entirely confident in week w that her end-of-year (week W) cumulative total expenditure will be below the limit L, $Pr(Z_W < L|Z_w) = 1$, then her perceived marginal price should be $PMP_w = c_L$. Similarly if she is confident that her end-of-year cumulative total expenditure is greater than L, $P(Z_W \geq L|Z_w) = 1$, then her end-of-year marginal cost should be the cost in Region 2, or $PMP_w = c_H$. Depending on the beneficiary's perceived uncertainty about spending past L, and conditional on not passing L when evaluating her problem in week w, her perceived marginal price in w may be somewhere between c_L and c_H.

Within the "in-between" response, Dalton et al. (2018) have tested a model of inconsistent time-discounting and salience of the "donut hole" to explain beneficiaries' behaviors in a subset of 2008 Medicare Part D claims. Their structural model that allowed for present bias was a better fit than a standard discounting model, but their estimation result of discount factors that were indistinguishable from $\beta = \delta = 0$ is indicative that present bias may not be an appropriate model.

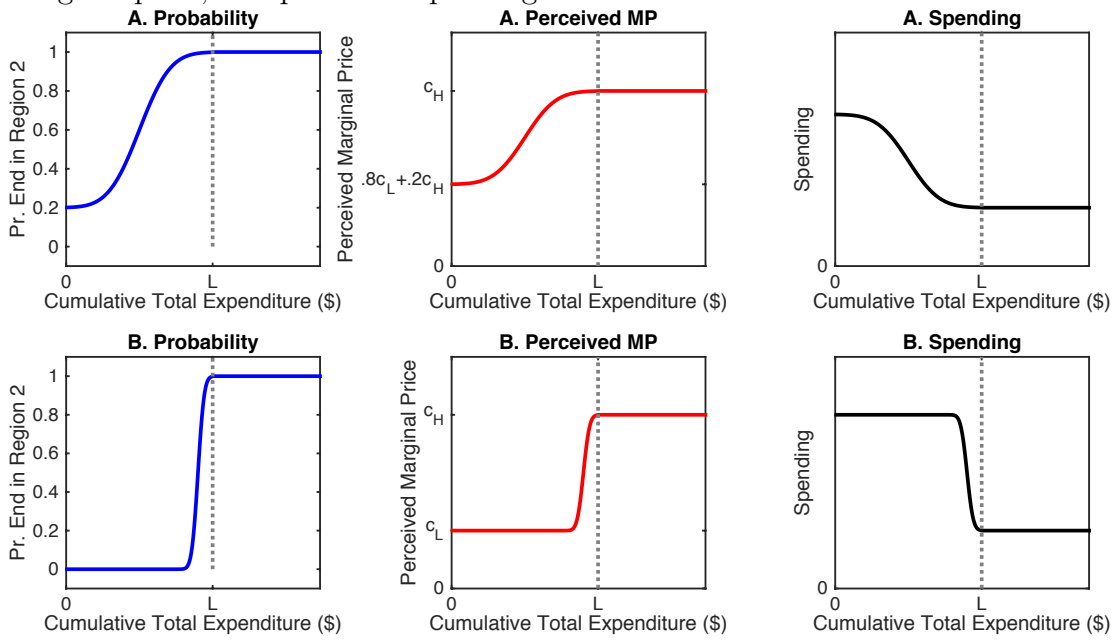
Abaluck et al. (2018) also allow for an "in-between" response to inter-year changes in the coinsurance rates by estimating the weights beneficiaries in Medicare Part D place on changes in coinsurance rates in the initial coverage region or coverage gap. While Abaluck et al. (2018)'s heuristic is somewhat similar to mine, in part due to the nature of their empirical approach, they limit their study to individuals who the researchers were confident to end the year in either the initial coverage region or coverage gap.

This paper proposes an alternative "in-between" response that a beneficiary may use to make the prescription purchasing decision. She may use a heuristic mental shortcut to calculate her objective expected year-end marginal price. Rather than comparing the net present value of the costs and benefits to conceive of an optimal expected year-end price, she may estimate a perceived marginal price based on her beliefs of the population objective probability of ending the year in any contract region. Using the objective population probability of being in any contract region, she can then infer the price in that region. In the two region case, $Pr_w(\text{in Region 2 in W}|Z_w) = 1 - Pr_w(\text{in Region 1 in W}|Z_w)$.

If a beneficiary were to make a decision to spend on prescriptions in the last period of the year, no-matter her method of evaluating her "in-between" response, her perceived marginal price is just her marginal spot price. There is no uncertainty as to what her marginal cost is. However, at earlier parts of the year $w < W$, her beliefs on her probability of ending the year in any particular region and the associated expected marginal price can impact her spending patterns.

Consider the example probabilities of ending the year in Region 2 presented in Figure 1.3. In the series of Panel A graphs, suppose it is early in the year, the beneficiary who has \$0 cumulative total spending believes that there is only a 20%

Figure 1.3: Example probabilities of ending the year in Region 2, perceived marginal price, and predicted spending



Note: The example assumes an individual faces the contract in Figure 1.2 Panel B.

chance of ending the year in Region 2 (and necessarily a $1 - 0.2 = 0.8$ chance of ending the year in Region 1). As her cumulative total expenditure Z_w approaches the threshold L, the probability of ending the year in Region 2 approaches 1. Even before she reaches $Z_w = L$, because there are still many weeks left in the year, she anticipates an almost 100% chance of ending the year in Region 2.

Suppose she employs the heuristic with no discounting to generate her marginal cost, then her objective expected marginal price (HMP) is below.

$$HMP_w = c_L Pr_w(\text{in Region 1 in W } |Z_w) + c_H Pr_w(\text{in Region 2 in W}|Z_w)$$

Her overall expected marginal price would then be the function depicted in in Figure 1.3 A. Her perceived marginal price transitions from $0.8c_L + 0.2c_H$ to c_H as the cumulative total expenditure increases. If demand is a monotonically decreasing and continuous function of price, this implies that her spending should be a monotonically decreasing function of her perceived marginal price and should also

adjust far prior to the limit L.

Suppose Panel B represents a week at the end of the year. Towards the end of the year, the beneficiary's uncertainty around ending the year in Region 2 decreases, because there is less time left in the year for health shocks as compared to Panel A. Any uncertainty only remains if she is in a narrow range of cumulative total expenditure values just prior to L. As she approaches the end of the year, her probability of ending the year in Region 2 approaches a piecewise formula, where this probability is 0 if $Z_w < L$ and 1 otherwise. Similarly, if demand is a monotonically decreasing function of price, then her spending should also approach a piecewise formula. If she has spent very little, she may be certain to end the year in Region 1 and subsequently uses the c_L rate, while if she has crossed the limit L, she should spend according to the c_H rate, i.e. $s_w = s(c_L)$ if $Z_w < L$ and $s_w = s(c_H)$ otherwise. In aggregate data, it is natural to see how these expectations could generate spending that may appear to be discontinuous at the limit L, especially if there are few individuals observed in the transition region.

1.2.1 Discounting

While the examples in Figure 1.3 did not include any explicit discontinuities in the probabilities or expected marginal prices, an alternative scenario could exist. For example, suppose that in a period w early in the year beneficiaries believe that there is a 20% chance of ending the year in Region 2 for all $Z_w < L$ and by necessity the probability is 1 if $Z_w \geq L$. The explanation for why a beneficiary near the limit may not upwardly revise her 20% probability of ending the year in Region 2 may include her inattention to being so close to Region 2, or completely ignoring her potential future health shocks (having a zero discount factor).

This paper argues that non-zero geometric discounting and present bias models should not generate such discontinuities. Consider the adaptation to the heuristic model that allows for present bias. Assume the decision in the heuristic approach only considers today's medical impact (or benefit of the drugs) compared to the total cost today multiplied by the future perceived marginal price. The beneficiary would necessarily be making her spending decisions by discounting her entire heuristic marginal price $PMP_w = \beta\delta^{W-w} * HMP_w$, so she would purchase her prescriptions if

$$B(s|X) > \beta\delta^{W-w} HMP(Z_w, w|X)s$$

where β is the "present bias" discounting factor that represents the difference between the present t and all future outcomes and δ^{W-w} is the geometric or standard discounting factor that is the product of discounting in every week from the current week w to the end of the year W.

Assume a beneficiary has Panel A probability beliefs that generate a continuous heuristic marginal cost. Denote $\beta'_w = \beta\delta^{W-w}$ for a given week w. Then, introducing a $\beta > 0$ and $\delta > 0$ implies $\beta'_w > 0$ and will not generate a significant discontinuity in her perceived marginal prices. Instead her perceived marginal price would fall somewhere between $\beta'_w(0.8c_L + 0.2c_H)$ and $\beta'_w c_H$ in week w.

Further, introducing present bias as a explanation of suboptimal behavior observed in the empirical data is not necessarily appropriate. Present bias models of behavior only generate suboptimal behavior when decisions are made between "the present" and "the future" and not when considering different time points within the future. The assumption that the benefits of prescription coverage are incurred "today" and in the present, while prevalent in the literature is not necessarily accurate. When beneficiaries fill prescriptions, they are often not for immediate

consumption with 30 and 90 day supplies. Further, even if consumption was immediate, as Baicker et al. (2015) explain, the effects of many prescriptions such as statins to treat high cholesterol have far delayed benefits rather than any immediate symptomatic changes.

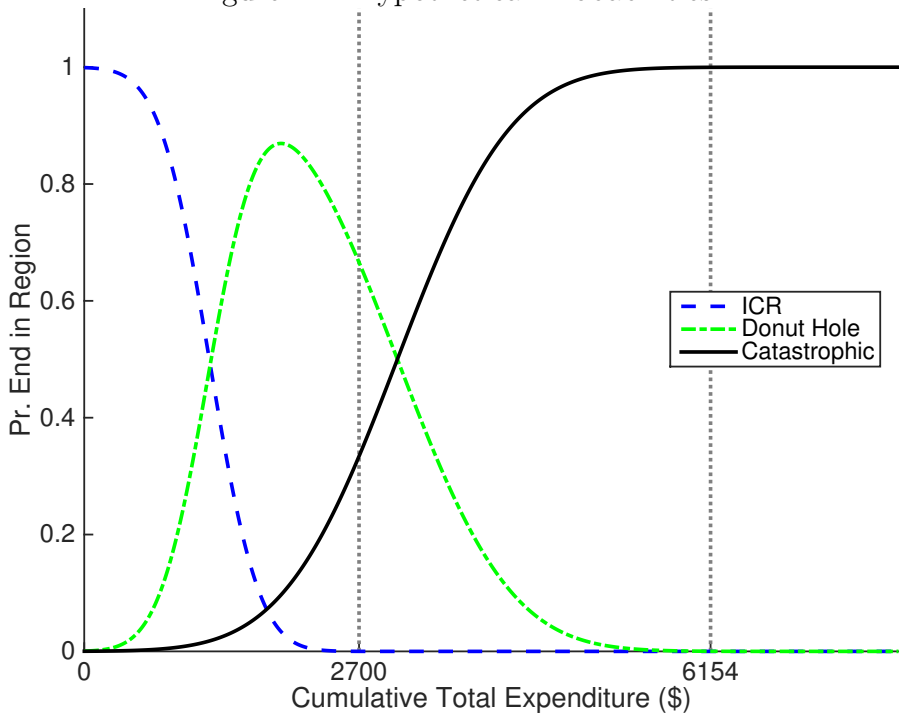
1.2.2 Intuition with three regions

The three-region setting is slightly more complex than the two-region setting, but the intuitions are similar. Assume a beneficiary has the example Medicare Part D contract and region limits displayed in Figure 1.1. Figure 1.4 displays hypothetical probabilities of ending the year in the three regions: the initial coverage region (ICR), the coverage gap (aka the donut hole), and the catastrophic region.

Similar to the probability of ending the year in Region 1 in the 2-region example, the probability of ending the year in the ICR is high at low cumulative spending values and low at high spending values. If it is earlier in the year, the probability of ending the year in the ICR can approach zero far prior to the subject crossing the $2,700 spending threshold. The catastrophic region is the terminal state and the probability of ending in that state is similar to the probability of ending the year in Region 2 displayed in Figure 1.3. It is the intermediate state of the donut hole which differs, and the probability of ending the year in the donut hole peaks when the probability of ending the year in the other two regions are at their lowest. As shown in this example, the peak of the probability of ending the year in the donut hole can occur prior to an individual even entering the donut hole depending on their end-of-year expectations.

Figure 1.5 presents the difference between the response of an agent A who re-

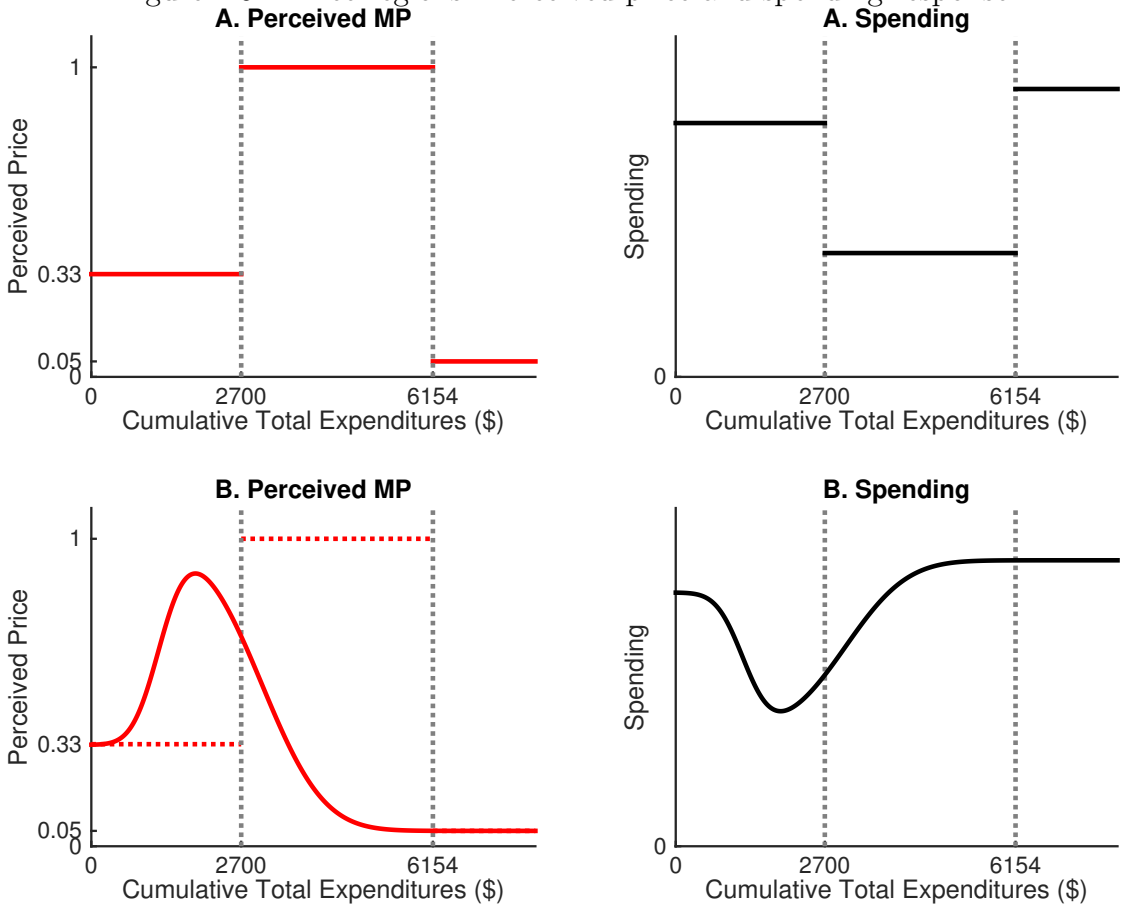
Figure 1.4: Hypothetical Probabilities



Note: An example graph of the probability of ending the year in any of three regions: initial coverage region (ICR), donut hole, or a catastrophic region. The region limits are based off of the 2009 example Medicare Part D contract from Figure 1.1. The probabilities are a stylized example.

sponds only to the spot price in panel A with the response of an agent B who responds using the Heuristic Marginal Price presented in panel B. The subject in panel A has a perceived marginal price that is equal to the Medicare Part D contract, while the panel B agent has a perceived marginal price that is equal to contract prices weighted by the probability of ending the year in each region. Assuming the agent's spending is still a continuous function of her perceived marginal price, then agent A's spending would be discontinuous at the boundaries of the Medicare Part D contract regions, while agent B's spending would change continuously. As the agents approach the end of the year, the probabilities should change and approach 1 if they are in the region and 0 otherwise, and agent B's perceived marginal price and spending should approach agent A's.

Figure 1.5: Three regions: Perceived price and spending response



Note: Assuming the hypothetical probabilities of ending the year in each region displayed in Figure 1.4, these panels show stylized examples of perceived marginal price and spending of hypothetical agents who respond to the spot price (A) and use the heuristic (B).

Section 1.4 applies this intuition to the data, and specifically presents the objective probabilities of ending the year in each of the spending regions and the implied perceived marginal price based on the observed data from Medicare Part D.

1.3 Background on Medicare Part D and data

Before discussing the data that is used in this study, this section covers the institutional details about the Medicare Part D program and the specific plan types that are part of the program.

1.3.1 Background

In the United States, Medicare is a health insurance program for the elderly that covered approximately 46 to 51 million individuals from 2009 to 2012.[6] It is structured in four parts: A, B, C, and D. Parts A and B include hospital and medical insurance for in- and outpatient care. Patients who are enrolled or eligible for Parts A and B, are also eligible to enroll in Medicare Part D, which provides insurance for prescription drug purchases, covering mostly self-administered drugs. Unlike Parts A and B, which are administered by the government, the Part D plans are administered by private insurers who are subject to the rules and regulations laid out by the government. Part C, also known as Medicare Advantage, is also administered by private insurers and is an all-inclusive alternative to Parts A, B, and D. While enrollment in Medicare is voluntary, individuals face significant penalties within the program if they choose not to sign up when first eligible (usually at 65) or have creditable (similar) health insurance coverage.

This paper focuses on beneficiary purchasing behavior of enrollees in Medicare Part D with stand-alone Prescription Drug Plans (PDP). In 2009, there were almost 18 million enrollees, with almost 20 million by 2012. The program began with

[6]Program statistics from the Centers for Medicare & Medicaid Services' Statistical Supplement https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Archives/MMSS/index.html

the Medicare Modernization Act of 2003, which was enacted in 2006. The government regulates a "standard" plan with a baseline minimum amount of coverage, and private insurers can offer a variety of plans that on an actuarially equivalent basis meet or exceed the generosity of the standard plan. This results in a significant variety in prescription coverage plans with 1,689 stand-alone PDP plans in 2009 ("The Medicare Part D", 2016).

In the context of studying beneficiary behavior when faced with nonlinear prices, the main advantage of studying Medicare Part D is the highly nonlinear structure of the coverage regions in the stand-alone Prescription Drug Plans (PDPs). This paper will focus on plans that do not have a deductible because the majority of beneficiaries do not choose plans with deductibles, but this section will discuss the full variety of plan structures. The 2009 standard Part D Plan included a deductible of $295, an initial coverage limit (ICL) of $2,700, and an out-of-pocket threshold (OOPT) of $4,350. The specific deductibles, ICLs, and OOPTs differ every year, but the structure of the standard plans and thus plans in the market are similar. The example plan in Figure 1.1 is actuarially equivalent to the standard plan displayed in the Appendix Figure A.1.

In the standard plan in 2009, patients are responsible for 100% of the cost of prescriptions until their current year cumulative spending reaches the deductible amount, after which they are in the initial coverage region (ICR). Note that both patient out-of-pocket (OOP) and "total" spending (the total amount spent through a combination of patient, insurance company, and drug companies) are the same up to the deductible. Patients in the initial coverage region (ICR) are then responsible for 25% or less of the total price of prescription purchases. If the cumulative total spending amount reaches the ICL amount, patients enter the phase often referred

to as the coverage gap or "donut hole" where they are again responsible for 100% of the total spending amount. If patients' prescription needs are so high that within one plan-year the sum of their OOP payments surpasses the OOPT, patients enter the "catastrophic" phase and are responsible for paying a greatly reduced share of the total costs of drugs. Specifically, they pay either the maximum of 5% of the total price of prescriptions or a $2.40 and $6.00 copay for a one-month supply of generic and branded drugs respectively. The pricing schedule resets at the end of the calendar year, and at the beginning of the next year beneficiaries begin anew with a total cumulative spending of 0 and the associated marginal costs.

The standard plan up until the catastrophic region has an exact mapping of out-of-pocket and total payments—an ICL of $2,700 corresponds exactly to cumulative OOP costs of $896.25, and an OOPT of $4,350 corresponds to $6,153.75 in total expenditures. In practice, an exact mapping between the two cumulative spending measures is difficult to ascertain in all plans. Plans only have to meet (or exceed) the coinsurance generosity of the standard plan on an actuarially equivalent basis through coinsurance, copays, or a combination of the two. Because of the actuarially equivalent clause, insurance companies have tremendous flexibility in structuring plans. Often, the cost sharing can be specific to drug tier or whether it is branded or generic. Regulation just requires that on average, the plan is expected to be similar to or more generous than the standard plan.

The inclusion of the Part D coverage gap or "donut hole" (and a main component of the plan's nonlinear structure) has been widely criticized and analyzed in the healthcare literature. The coverage gap was initially included as a cost-saving measure for the government similar to a deductible but positioned in the middle of the patient benefit schedule (Baker, 2006). The health policy literature has both

criticized the arbitrary location of the donut hole and the impact it has on drug adherence.

Every year, the standard plan's spending limits are updated to adjust for rising costs. And under the Patient Protection and Affordable Care Act (ACA), the government began to fill in the coverage gap and will continue to increase the plan benefits in this region through 2020. The phase out began in 2010 when the standard plan included an automatic $250 rebate for beneficiaries who reached the coverage gap. Further coverage in the donut hole increased in 2011 and 2012, when instead of a rebate, standard plans included a 50% discount on brand-name prescriptions that was paid by the drug manufacturer. This means that in 2011 and 2012, patients in the donut hole only paid 50% of the cost of branded drugs. In most cases, while not paid by the beneficiary, the 50% covered by drug manufacturers did contribute towards a beneficiary's cumulative out-of-pocket expenditures. Because of this set up, the discount did not significantly change the cumulative total expenditure amount it took for patients to get out of the donut hole. Table 1.1 shows the changes in the standard plan from 2009-2012.

This paper will focus on plans that have the government-defined ICL and OOPT, but not plans with the government-defined deductible. Enrollees have choices over a wide variety of Part D plans, with a majority of patients opting for plans with more generous plan benefits than the standard plans including no deductible plans. In 2006, fewer than 10% of beneficiaries were in plans with the standard design (Abaluck and Gruber, 2016), and my data supports this finding as well. Even prior to 2010 when the ACA started phasing out the coverage gap, many plans offered some type of gap coverage, though these were typically on generic prescriptions. See Section 1.3.2 for a deeper discussion of plan types.

Table 1.1: Medicare Part D benefit parameters for defined standard benefit 2009-2012

Plan Characteristics	2009	2010	2011	2012
Deductible	295	310	310	320
Initial Coverage Limit (ICL)	2700	2830	2840	2930
Out-of-Pocket Threshold (OOPT)	4350	4550	4550	4700
Total Expenditure equivalent OOPT	6153.8	6440	6447.5	6657.5
Rebate (1)		250		
Brand discount(2)			50%	50%
Generic copay (3)	2.40	2.50	2.50	2.60
Branded copay (3)	6.00	6.30	6.30	6.50

Note:

(1) The rebate begins when patients reach their out-of-pocket threshold (OOPT).

(2) The brand discount only applies when patients are in the coverage gap, i.e. when their cumulative total spending is above the ICL, and their cumulative non-insurer spending (patient payments, any subsidies, brand discounts paid by the drug manufacturers) is below the OOPT.

(3) In the catastrophic region, beneficiaries pay the maximum of the copay or 5% the total cost of the prescription.

For the purpose of understanding beneficiary behavior in the face of nonlinear contracts, there are other advantages to studying Part D plans instead of other nonlinear employer-sponsored health insurance plans. These advantages include the high frequency of claims and large percentages of beneficiaries experiencing different coverage phases year-over-year. Hoadley et al. (2011) indicate that 16% of Medicare beneficiaries ended the year in the coverage gap, with 3% of beneficiaries reaching the gap and passing it to end the year in the catastrophic region. Across 2008-2009, almost 30% of patients experienced the gap. Further, they document that reaching the coverage gap is persistent, as 71 percent of enrollees who reached the gap in 2008 did so again in 2009. This recurrence of reaching the coverage gap is due to the fact that many patients take medications for chronic conditions rather than for acute, short-term medical needs.

Another advantage of studying Part D is that many of the medications patients take in Part D are for chronic conditions, so enrollees in the nonlinear Part D setting may have a better ability to forecast their yearly spending on prescription drugs than in other types of health insurance. In a MedPac report on Medicare Part D, they state that the list of top 15 therapeutic classes of drugs by spending and volume has remained relatively consistent since 2007. The values from 2013 indicate that drugs in the diabetic, asthma/COPD antihyperlipidemics,[7] antipsychotics, antihypertensive,[8] and peptic ulcer therapeutic classes are responsible for approximately 40% of drug spending (MedPAC, 2016). These drugs are all used to treat ongoing chronic conditions.

Further, in studying beneficiary behavior, it is also an advantage that Part D claims only cover self-administered drugs and do not cover drugs administered at the doctor's office or in the hospital. Unlike hospital claims, there is an increased likelihood that filling prescriptions are decisions made by beneficiaries rather than decisions made directly by a medical professional. However, it is still a concern for older patients that drug purchases could be done by a proxy.

One of the challenges of the Medicare Part D data for studying spending in the non-linear pricing schedule across years is that the pricing schedules change over years. As mentioned, each year the standard plan adjusts, and it is highly likely that the individual private plans adjust. Patients also have the choice to switch plans to different insurers or to Medicare Part C at the end of each year. However, while plans change every year, the schedule remains similar with mostly minor increases in the exact limits of each coverage region. The literature also documents a significant amount of inertia and inattention in patient choice of

[7]Used to treat high cholesterol
[8]Used to treat high blood pressure.

plans, indicating that approximately 10% of Part D patients switch their plans between every two years (Abaluck and Gruber, 2016; Abaluck et al., 2018; Ho et al., 2017).

1.3.2 Data description

The primary dataset includes the prescription drug and medical claims from a random 5% subsample of Medicare beneficiaries enrolled in Part D in the years between 2009 and 2012, with approximately 2 million individuals per year.[9] The data come from the Center for Medicare and Medicaid Services. Broadly, the data cover the beneficiary demographics, their Part D prescription claims, and the plan characteristics of the specific Part D plan that each beneficiary chose. The prescription drug claims include the exact drug purchased, days supply, purchase date, the proportion paid by both the patients and insurance companies, and the benefit phase each claim occurs in. The plan characteristics supplement contract information with full details on the plan premiums and the exact cost-sharing characteristics: deductibles, coinsurance, copays for specific drug types and tiers.

Basic demographic information of the beneficiaries (gender, age, race) along with hospitalization and doctor claims information from Medicare Part A and B, which are used to determine patient health conditions, are also observed. With these data, I use the CMS-provided risk model to calculate a "risk score" or summary estimate of the expected average drug spending implied by patients' demographics and health conditions.[10] In order to normalize the risk score comparison,

[9]This analysis does not consider patients enrolled in Medicare Advantage (Part C), because while the Medicare Part D claims data include Part C prescription claims, the dataset does not include the doctor and hospitalization claims, which are used to control for heterogeneity.

[10]CMS use Hierarchical Conditional Codes (HCC) and RxHCC to adjust the reimbursement payments to insurance companies that offer plans in Medicare Advantage and other pro-

I use the risk and demographics score translation from 2011. Individuals who have a combined risk and demographic weight of 1 have an average prescription reimbursement liability of a typical Medicare Part D enrollee in 2011.

The analysis sample for this paper is constructed by keeping only individuals in the Medicare system who had Medicare Part D from 2009 through 2012 with specific beneficiary and chosen plan characteristics. To list the beneficiary characteristic restrictions, the sample contains beneficiaries who are 65 or older, are enrolled in Medicare PDPs from 2009-2012 through the Old Age and Survivors Insurance (OASI), and are not enrolled for disability insurance or other qualifiers for Medicare. The sample also excludes individuals who are dual eligible for Medicaid financial assistance or receive other types of low-income subsidies (LIS) for premiums or cost-sharing. These individuals are excluded because they face very low cost-sharing and minor changes in their marginal costs. Even individuals who only receive premium subsidies are omitted, because they are more likely to be lower income and are more likely to be influenced by budget constraints. Further, the analysis of the paper also excludes individuals whose Medicare Part B claims indicate they were in long-term care institutions (LTI) such as nursing homes in the prior year. Including also a restriction for individuals who were in Medicare Part D for all twelve months of each year, this leaves approximately 300,000 beneficiaries in each of the 2009-2012 years respectively. This is the "full" sample and is used as a comparison group to the analysis sample.[11]

There is a substantial amount of plan variety in this data set that is subject

grams. The HCC and RxHCC scores are used to reimburse the plans for managing patients with illnesses with expected increased medical and prescription medication costs, respectively. http://setma.com/EPM-Tools/tutorial-hcc-rxhcc-risk

[11]The sample exclusion restrictions are further discussed in Appendix Section A.1 and listed in Appendix Table A.1 with the percentage of the entire Medicare/Medicaid 5% sample they encompass. Note that not all individuals have Medicare Part D. See Appendix Table A.3 for summary statistics before plan choice restrictions.

to further restrictions to construct the analysis sample. The plan variety and the beneficiary's opportunity to switch plans between years pose a challenge for studying beneficiary behavior. There may be an endogenous relationship between a beneficiary's choice of plans and her spending patterns. That plan variety can be seen in Table 1.2. Across the four years, very few enrollees chose plans with deductibles; 65-75% of enrollees have plans without deductibles. Overall only 15-20% of enrollees have plans with the standard government-defined plan limits in the deductible, initial coverage region, and out-of-pocket thresholds. However, a majority of patients still have the same ICL and OOPT limits with 76% without a deductible. Despite the pervasiveness of the ICL and OOPT spending limits, less than 2% of beneficiaries have plans that use the exact 25% coinsurance rate suggested by the government for the initial coverage region. Because many beneficiaries do have plans with the Medicare-defined spending limits for their ICL and OOPT, this paper will focus on the beneficiary's response to approaching these limits rather than her response to the specific coinsurance rates.

Table 1.2: Summary of full sample of Medicare beneficiary plans 2009-2012

	2009	2010	2011	2012
Deductible: None	0.76	0.68	0.63	0.67
Deductible: Other	0.04	0.17	0.19	0.15
Has Standard Deductible	0.20	0.15	0.18	0.18
ICR: Standard Coinsurance	0.01	0.01	0.01	0.00
ICR: Cost Share Tiers	0.99	0.99	0.99	1.00
ICL: Standard	0.99	0.93	1	1
OOPT: Standard	1	1	1	1
Standard Plan Limits	0.20	0.15	0.18	0.18
No Deductible, Standard ICL & OOPT	0.75	0.61	0.63	0.67
Main 4 Year Sample	0.34	0.32	0.31	0.29
Observations	291,550	304,477	317,670	333,309

Note: The full sample includes individuals in each year who satisfy the criteria for the "Full 12 Month Sample" from Table A.1.

The analysis sample or baseline "No Deductible" sample then focuses on the balanced panel of individuals who qualify under the prior restrictions discussed, had the government prescribed limits for the ICL and OOPT, and who did not have deductibles. Individuals were also omitted if they were observed to have either zero spending in any year or had claims in every week of the year. The final sample contains 89,354 beneficiaries or about 30% of the full sample. Part of the reason for this restriction is to help standardize the spending limits for the analysis in later sections. This restriction has the negative effect of decreasing the sample size and reducing the generalizability of these results. Further, because these individuals choose plans without deductibles, the selected sample may be simultaneously more risky, more risk averse, and richer than an average Medicare Part D enrollee. Also because the sample requires the beneficiary have the same plan structure in all four years, they have higher inertia and may have higher costs of switching.[12]

The dataset of 2009-2012 claims differs from the data used in the papers previously mentioned. In all cases, this sample draws from a later sample of individuals with Medicare Part D than individuals who are in Joyce et al. (2013)'s 2006, Dalton et al. (2018)'s 2008, Einav et al. (2015)'s 2007-2009, and Abaluck et al. (2018)'s 2006-2009 sample. My panel is a longer sample than most other papers and covers 2008 medical history and four years of claims. Because this sample involved individuals who retained similar plan structures through all four years, they are mechanically more likely to have experience with Medicare Part D[13], their plan structure, and their prescription needs than individuals described in these other

[12]An additional "Standard" sample is also created from the 9,178 individuals who signed up for plans with the government-defined deductible, ICL, and OOPT, for whom summary statistics are presented in Appendix Section A.1.

[13]Remember also Medicare Part D began in 2006, so papers using earlier samples may study have less experienced subjects.

papers. These differences possibly translate to different findings in the empirical section.

Table 1.3 displays the demographics of the baseline sample of individuals without deductibles. The average age of the population in 2009 is about 75, which is slightly older than the average Medicaid 5% sample population, but is consistent with the ages of beneficiaries enrolled in Medicare Part D. The vast majority (approximately 95%) of beneficiaries are white. The sum of the 2011 RxHCC and demographic weight is less than one, indicating that the average person in this sample is healthier than the average Medicare Part D participant and reflecting the fact that disability, LIS, and LTI individuals were not included. However, they are sicker than the average individual in the full sample, which may reflect both the higher age and the optional nature of joining Medicare Part D for prescription purchases.

A majority of individuals have chronic conditions with 66% and 73% of individuals having hypertension and high cholesterol in 2008. Also, almost a quarter of the sample has diabetes and 10% has had cancer treatment of some kind in 2008. These are all conditions that often require constant prescription refills and spending using the Medicare Part D benefit. The Kaiser foundation documents that patients who took drugs to treat some of these conditions are far more likely to reach the coverage gap and catastrophic regions (Hoadley et al., 2011). They observed that the average Part D enrollee's probability of reaching either of the two regions as 19% in 2009 but 56% for patients on breast cancer treatment drugs, 40% for those taking oral anti-diabetics, 32% for those on statins etc. Also, as the beneficiaries progress through the years, they get older and sicker. The increase in sickness is reflected in the claims data with an increasing trend in spending amount

Table 1.3: Demographics of baseline No Deductible sample of Medicare beneficiaries in 2009-2012

	Mean	SD
Age at End of 2009	74.94	6.58
Start Medicare	1999.11	6.55
Female	0.65	0.48
Race: White	0.95	0.21
Race: Black	0.02	0.15
Race: Other	0.01	0.11
Race: Asian	0.01	0.08
Race: Hispanic	0.00	0.06
Observations	89,354	

	2009		2010		2011		2012	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2011 RxHCC weight	0.48	0.29	0.51	0.29	0.53	0.30	0.55	0.31
2011 RxHCC demo. weight	0.42	0.01	0.42	0.01	0.42	0.01	0.42	0.01
Diabetes	0.24	0.43	0.26	0.44	0.27	0.44	0.28	0.45
Hypertension	0.67	0.47	0.69	0.46	0.70	0.46	0.71	0.46
Has Cancer	0.10	0.30	0.11	0.31	0.11	0.31	0.12	0.32
High Cholesterol	0.73	0.44	0.76	0.43	0.76	0.42	0.77	0.42

Note: The baseline No Deductible sample includes individuals in each year who satisfy the criteria from Table A.1 and also were in a plan with standard ICL and OOPT limits with no deductible from 2009-2012. Risk scores are normalized to 2011 RXHCC scores for consistency across years.

and frequency through 2009-2012.

Even though the majority of beneficiary plans do not follow the standard Medicare-defined coinsurance amounts, the coinsurance levels between the different coverage regions are on average still economically and significantly different from each other. Table 1.4 displays the average coinsurance amount that beneficiaries in the baseline No Deductible sample face. Because the plans patients choose do not have deductibles, their effective coinsurance rate in the initial coverage region (ICR) is higher than the standard plan at 39% to 41% of the total cost of care. The coinsurance rate in the donut hole in 2009 and 2010 prior to the ACA legislation to filling in the donut hole was not quite 100% but still significantly high at approximately 92%. The beneficiary responsible portion of the coinsurance rate during the coverage gap in 2011-2012 was significantly lower at approximately 54-

55% with the addition of the 50% discount on branded drugs. While the difference between the ICR and Coverage Gap coinsurance rates in the first two years of the sample is higher than the latter two years, there is still a difference in the latter two years. This means that beneficiaries should qualitatively still respond to these plan characteristics in the way laid out in Section 1.2. It is expected that any beneficiary response to the coverage gap in 2009-2010 may be muted in 2011-2012, because the marginal price difference between the two regions is smaller. The average coinsurance levels for patients with standard plan limits follow closer to the government recommended plan and are shown in the Appendix Table A.4.

Table 1.4: Average (person-week) coinsurance by insurance region 2009-2012

	2009		2010		2011		2012	
	Percent	Count	Percent	Count	Percent	Count	Percent	Count
ICR	38.63	4,285,602	40.69	4,300,522	39.75	4,277,143	40.71	4,293,291
Coverage Gap	92.23	319,117	91.55	306,331	55.08	325,492	54.28	309,185
Catastrophic	5.87	32,100	5.80	30,742	5.95	36,657	5.97	38,708

Note: Table is generated from the baseline No Deductible sample. The coinsurance rates are averaged over the amount the patient pays (does not include the drug manufacture discounts in 2011 and 2012) divided by the total expenditure cost in the person-week observation where spending occurs. This rate is effectively weighted by the time individuals spend in each phase. The count reflects the fact that there are more person-week observations in the ICR region than others. These sums do not reflect the counterfactual coinsurance rates that beneficiaries with low spending would have faced if they had reached higher spending. While the data contain the actual structure of the beneficiary plans with exact coinsurance and copay rates for drug tiers, it is difficult to summarize in a specific coinsurance rate without knowing the mix of drugs that patients may consume.

The baseline sample is used to study beneficiary behavior as they cross spending phases, and the sample does include individuals who are likely to reach the coverage gap and beyond. Table 1.5 shows that while the majority of the beneficiaries without deductibles end the year in the initial coverage region, over 21% of enrollees end the year past the coverage gap with approximately 2-3% reaching the catastrophic coverage region. While heterogeneity among beneficiaries mean that many would not have realistic expectations of reaching the higher spending coverage regions, there is certainly a significant subset of enrollees who might expect to

end the year in these regions.

Table 1.5: Proportion of beneficiaries in each insurance region at the end-of-the-year

	2009	2010	2011	2012
ICR	76.47	77.70	76.93	78.81
Gap	21.03	19.97	20.28	18.46
Catastrophic	2.49	2.32	2.79	2.73
Observations	89,354	89,354	89,354	89,354

Note: Table is generated from the baseline No Deductible sample. The proportion of beneficiaries that end the year in each phase is averaged over the individual beneficiary.

The sample beneficiary's average raw claim occurrence probability (i.e. the probability of submitting a claim) in each of the Medicare Part D insurance coverage regions is depicted in Table 1.6. Across the four years, while the probability of ever making a prescription claim in a week is 32-34%, the raw probabilities do differ significantly within the insurance regions. Of the beneficiaries who are in the initial coverage region, their average probability of spending is in the 31-33% range. The weekly claim probability for the observations in the coverage gap is higher at 43-47% and highest in the catastrophic region at 56-57%. The overall average is very similar to the average in the ICR region, since the majority of the individual-week observations occur in the ICR. This pattern can also be seen in Figure 1.6, which displays both the frequency of claims and the mean claim occurrence within $50 bins in 2009 as a function of the cumulative total expenditures. This image illustrates the raw probabilities of beneficiary's spending.

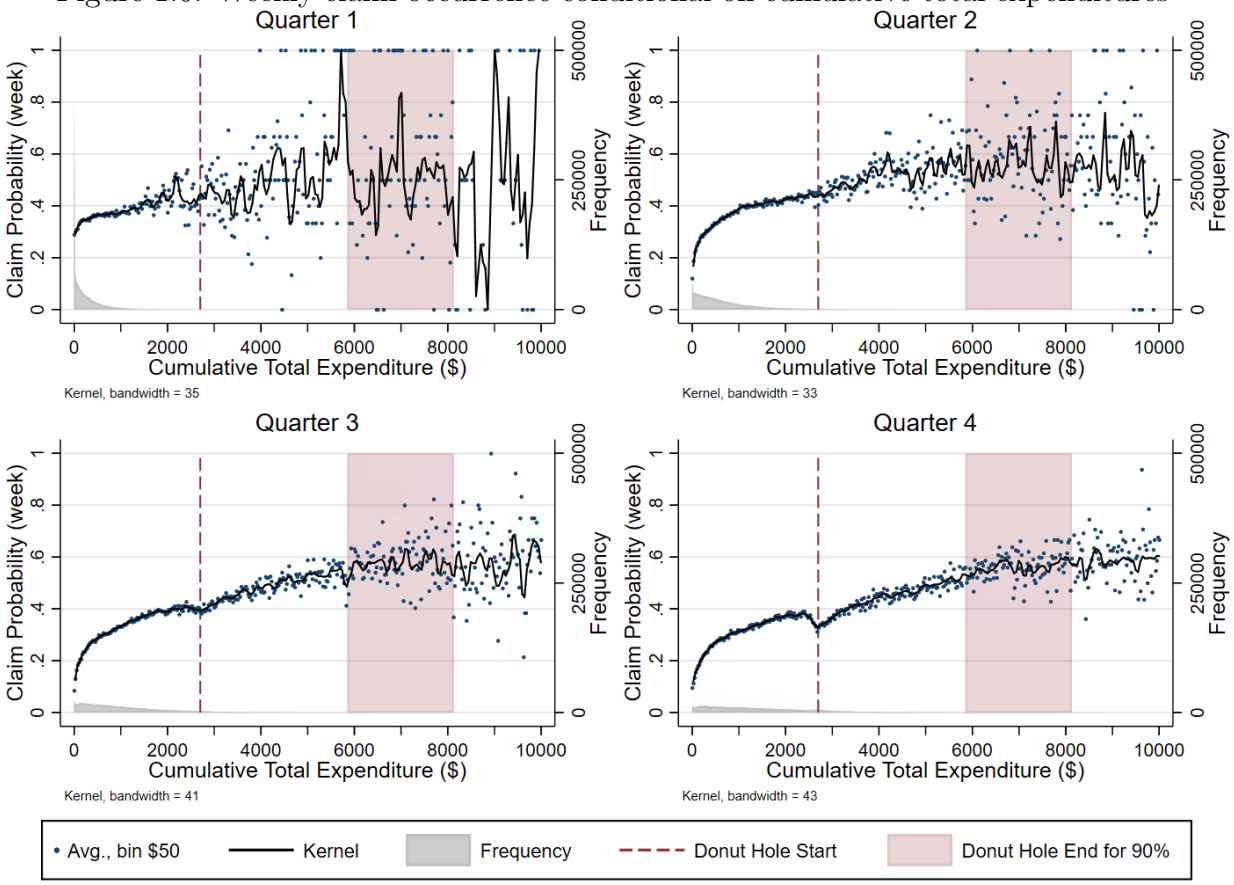
Given the number of individuals who end the benefit year in each region and the large number of person-week observations in the ICR relative to the catastrophic region, it makes sense that the ICR probability is closer to the total average probability. The differences in the probabilities across regions is a possible indication of heterogeneity in the probability of spending, where individuals with

Table 1.6: Average probability (%) of weekly spending in coverage regions

	2009	2010	2011	2012
ICR	31.48	31.90	32.15	32.71
Coverage Gap	42.65	43.63	45.01	46.93
Catastrophic	56.22	56.84	55.87	56.84
All weeks	32.44	32.86	33.26	33.87

Note: Table is generated from the baseline No Deductible sample. Table displays the raw probability of spending in a week in each coverage region and year. The coverage regions are the initial coverage region (ICR), the coverage gap (aka the donut hole), and the catastrophic region. The average is of the probability of a person-week observation having a claim and is low because there are more observations in the ICR as seen in Table 1.4.

Figure 1.6: Weekly claim occurrence conditional on cumulative total expenditures



Note: Using a $50 dollar bin, the points are the average probability of spending on prescription purchases in a week conditional on the quarter of the year and the cumulative total expenditure. Einav et al. (2015) produced very similar graphs of the probability of a prescription purchase in a month rather than the week level. This image is only of 2009 claims.

higher probabilities of claims, who may also have higher average spending amounts, are more likely to have observations in the coverage gap and catastrophic regions.

Einav et al. (2015) uses a similar graph to Figure 1.6 of the monthly probability of spending to help illustrate the empirical patterns that beneficiaries engage in at the coverage gap. However, in order to fully understand beneficiary behavior at and before the kink, researchers must take into consideration the significant amounts of heterogeneity in prescription needs that exist in the Medicare Part D population. Beneficiaries who are more likely to spend (and spend more), are also more likely to be observed in higher cumulative spending bins, while those who spend less are observed at the lower cumulative total expenditure levels. The approach that is taken in this paper to handle this heterogeneity is through the use of fixed effects in Section 1.5.

The limitations of these data include the fact that they do not capture prescription purchases outside of Part D such as large retailer generics since those purchases are outside the scope of the Medicare system. Further, this paper has limited data on beneficiary incomes. One potential explanation for non-standard behavior could be patients reaching budget constraints, and this paper is not able to directly measure individual liquid wealth.

1.4 Heuristic approach applied to Medicare Part D

This section applies the heuristic approach proposed in Section 1.2 to the empirical data. The goal is to recover the average objective expected end-of-year prices beneficiaries should expect as a function of week and current cumulative spending. One drawback of the heuristic method is that it requires a strong assumption that a beneficiary has access to the data to know her objective probabilities; however, this heuristic is still useful as a tool to understand how optimal agents should

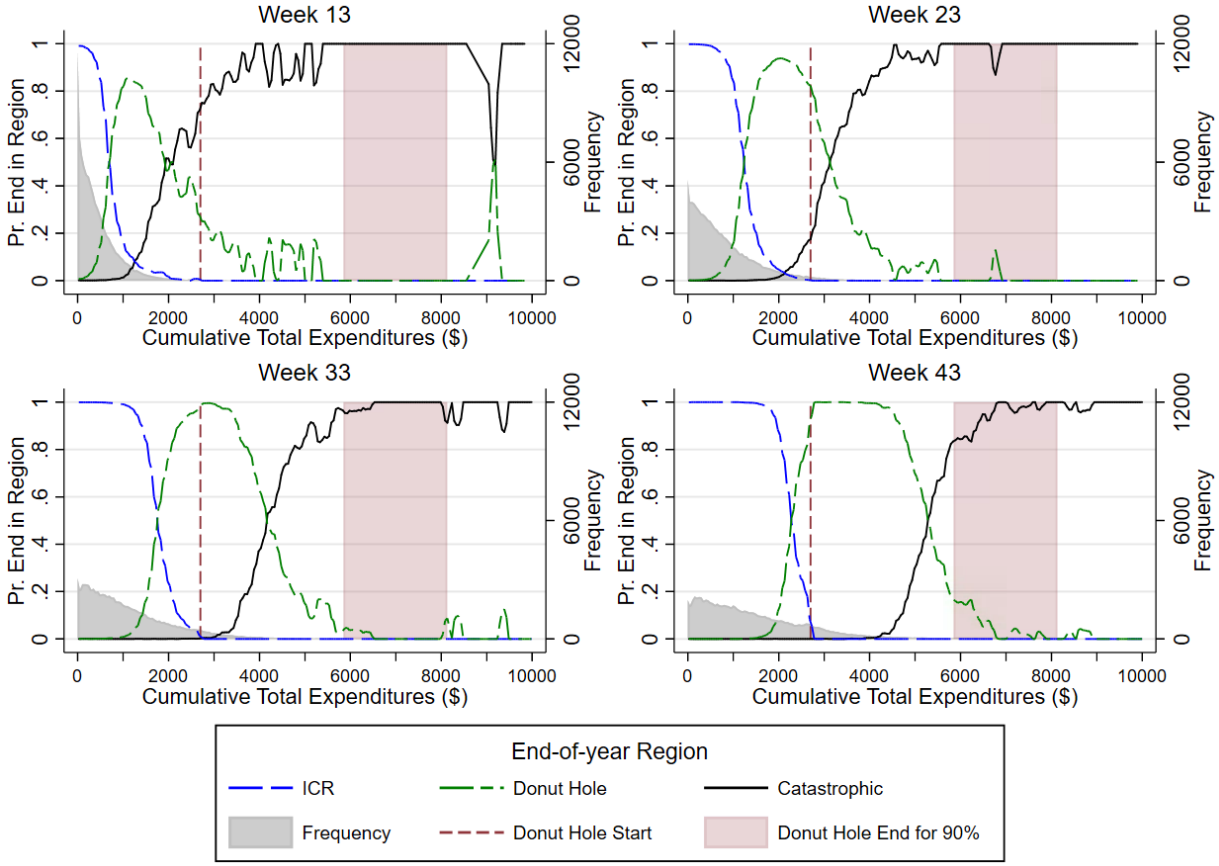
behave. Further, if the stylized facts about beneficiaries sharply reducing their spending when approaching the coverage gap earlier in the year are generally true, then under this heuristic this behavior would only be explained by discontinuities in beneficiary's subjective end-of-year probabilities, i.e. if they fail to update their beliefs or receive health shocks that lead to surprises.

There are more regions and more noise in these estimates than the simulated example, but basic interpretations hold. First, the probability distributions of ending in each of the contract regions based on the time of the year and beneficiary year-to-date spending are constructed. While these graphs are constructed at the cross-individual level, they should still provide insight for individuals to construct their internal beliefs.

For the baseline No Deductible sample, Figure 1.7 displays the raw probability in weeks 13, 23, 33, and 43 of ending the year in each coverage region conditional on their cumulative total expenditures (through week 12, 22, 32, and 42 respectively). The heuristic expected marginal price (HMP) is constructed using these probabilities as discussed in Section 1.2, and it is displayed in Figure 1.8. Both probabilities and marginal costs are calculated from beneficiaries whose weekly cumulative total expenditure amounts fall in \$50 bins in the x-axis. The red line represents the initial coverage limit and the boundary between the initial coverage region of low coinsurance and the coverage gap. There are significantly more differences in the bin means for the probability ending the year in the catastrophic and initial coverage gap conditional on the cumulative total spending. This is due in large part because the out-of-pocket threshold (i.e. the boundary between the coverage gap and catastrophic region) translates to different cumulative total expenditures for different insurance plans and drug consumption patterns. The

noise in probabilities is exacerbated in the week 13 panel because there are few individuals who have accumulated high spending levels that early in the year.

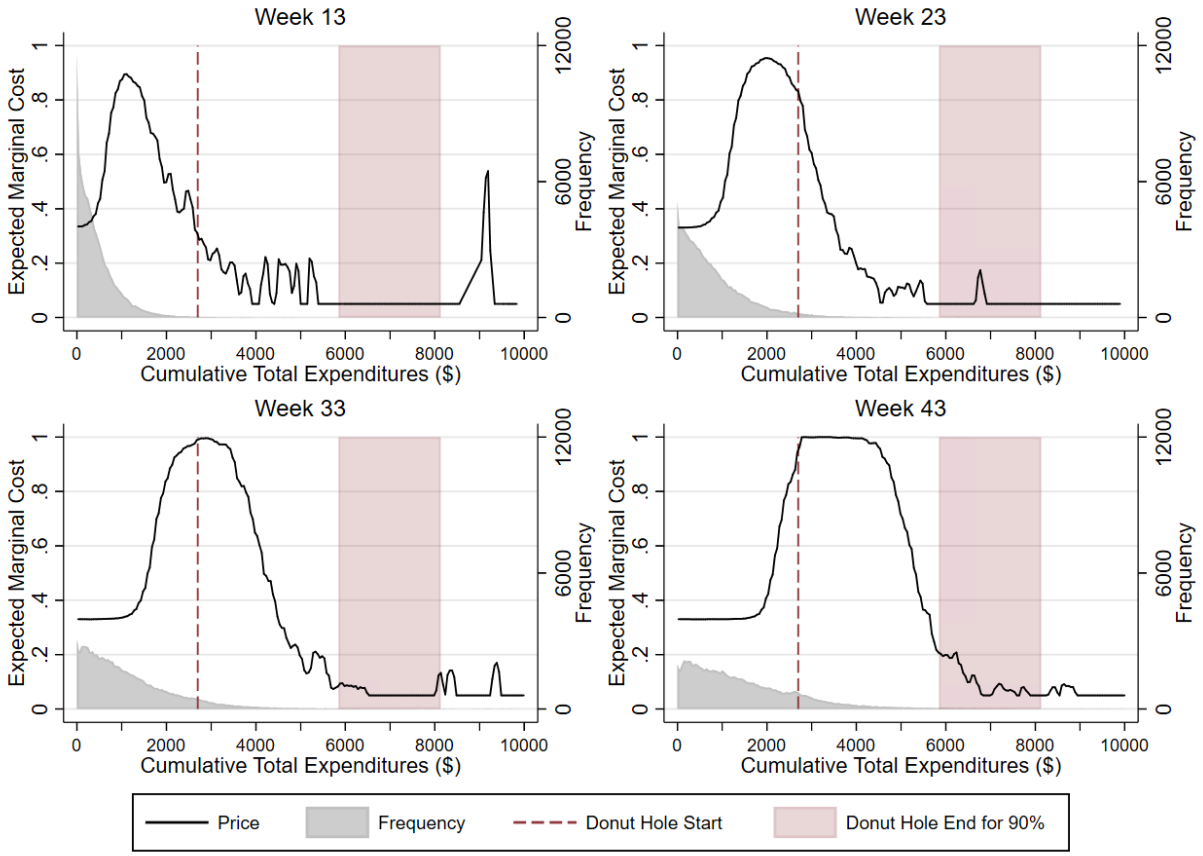
Figure 1.7: Distribution of the probability of reaching each coverage region in 2009 conditional on cumulative total spending



Note: The graph depicts an Epanechnikov kernel-weighted polynomial of the probability of a beneficiary being in coverage region r at the end of the year given week w and within a \$50 bin of the cumulative total spending Z_w. This makes up the distribution $F_r(Z_w, w)$. The figure is generated from the baseline No Deductible sample. The bin size = 50 and was chosen for illustrative purposes. Similar images for 2010-2012 are included in the Appendix. Because the out-of-pocket threshold limit for entering the catastrophic coverage region does not on aggregate map to a specific cumulative total amount, the average probability of ending in those phases as a function of the cumulative total amount has a wide dispersion of points.

While much of the prior literature has focused on the behavior directly at the initial coverage limit, the effect of the non-linear pricing structure could be evident far prior to the coverage gap earlier in the year. In week 13, if a beneficiary's spending is between \$500 and \$1,500, she is most likely to end the year in the

Figure 1.8: Heuristic expected marginal price in 2009



Note: Depicts the expected marginal price based off of the objective distributions of the probability of ending the year in each coverage region depicted in Figure 1.7. The heuristic expected marginal price is $HMP(Z_w, w)) = F_{ICR}(Z_w|w) * MC(ICR) + F_{Gap}(Z_w|w) * MC(Gap) + F_{Cat}(Z_w|w) * MC(Cat)$, where MC(ICR) = .33, MC(Gap)=1, and MC(C=.05) the government standard plan amounts.

coverage gap, and this translates to her highest HMP being in that region. If a beneficiary's spending is already over $1,500, she is most likely to end the year in the catastrophic zone, even though she is still far away from even the transition from the ICR to the coverage gap. If a beneficiary uses the HMP as her perceived marginal price, she should increase her spending as the HMP decreases, which begins well before the 2009 ICL of $2,700.

As time progresses, the probability of ending the year in any of the regions also shift to higher cumulative totals and tend closer to 0 and 1, and the beneficiary's

highest heuristic expected marginal price more closely resembles the plan spot prices. Over time the cumulative total spending level at which she experiences her highest HMP moves closer to the discontinuity between the initial coverage and the coverage gap regions. While the distribution of the probability of ending the year in the coverage gap was a maximum of 80% in week 13, there is an increasing group of individuals who have spent around $2,000 by week 20 and $2,700 in week 33 who are certain to end the year in the coverage gap. Thus the highest HMP, which should correlate with a beneficiary's lowest levels of spending move to about $2000 in week 23 and to the ICL or $2,700 in week 33. In week 43, or approximately 2 months before the end of the year, there already exists a discontinuity in the probability of ending the year in the coverage gap at the initial coverage limit. These graphs are not generated based on the behavior of necessarily standard forward-looking agents, so the HMP may approach the spot price earlier than for standard rational agents.

In translating the beneficiary's perceived marginal price to their spending patterns, if the demand function is smooth and quantity demanded is decreasing with the marginal price of prescription purchases, broad predictions can be made. The expectation is that within any time period the cumulative total expenditure with the lowest heuristic expected marginal prices in a time period should correspond with the cumulative total expenditure amount that has the highest level of spending. Similarly, within a time period, the cumulative total expenditure with the highest heuristic expected marginal price should correspond with the lowest levels of spending. If the demand function is smooth and the transitions in the HMP are smooth (as they are for the majority of the periods), then the spending should also be smooth.

The relationship between the HMP and spending is not necessarily expected to be one-to-one and would not be as dramatic as a simple "flip" of the HMP curve, but the expectation is that the location of spending changes should correspond with the peaks and troughs of the HMP. In fact, because prescription purchases often have immediate and significant health benefits, these drugs may be relatively inelastic goods and respond little if at all to the marginal price changes. Using the end-of-year purchases, Einav et al. (2018) measures the elasticities of different drug classes and find an overall elasticity of -0.037 so that a one percent increase in out-of-pocket cost leads to a 0.037 percent decrease in the probability of filling a claim. Thus, the expectation is that the HMP would result in small changes in the probability of filling a claim as well.

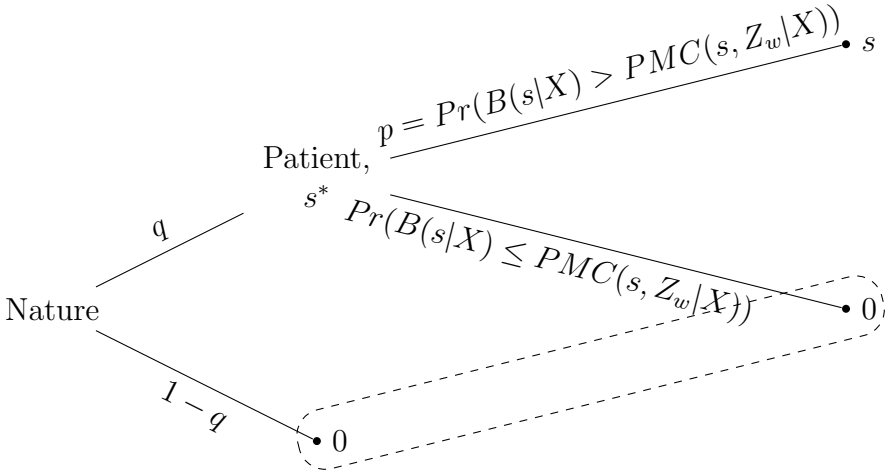
The next section will discuss the empirical approach to estimate a graphical representation of beneficiary spending patterns. It will also discuss how consistent behavior is with the simple heuristic model of spending.

1.5 Estimation model and results

This section details an empirical model of prescription purchasing behavior under Medicare Part D that ascribes prescription spending to both opportunities to spend and then beneficiary decisions to spend. The model emphasizes the effects of past cumulative spending on weekly prescription purchases. Ultimately the estimation shown in this paper lumps both the opportunity to spend and the decision to spend as one process, but the model is laid out to provide intuition on the beneficiary's underlying motivations. The intent of the estimation is to document the patterns of consumer behavior that emerge as patients progress through their insurance plans

and not to capture the full dynamics of the consumer prescription choice problem. The beneficiary's prescription spending decision is aggregated and examined on the weekly level to reduce the size of the problem.

Figure 1.9: Spending Decision



The initial model shown in Figure 1.9 assumes beneficiaries have a choice whether to spend on prescriptions and a separate limited choice on the amount to spend. The only observables that are available in Medicare Part D data are prescription purchases and not the beneficiary's direct consumption of drugs. Because of the nature of chronic prescriptions, many are offered in 30 to 90 day refill amounts, which means that there is a periodicity to beneficiary's spending patterns that are not necessarily driven by a choice to spend or not. In order to capture some of this, the model assumes that the personal utility from medications should only apply if a patient has a medical event or shock that requires treatment (determined by Nature). Thus, the decision to use Medicare Part D to purchase prescriptions should only occur when those events arrive and a doctor has written the patient a prescription. These events can be temporary health shocks that require treatment such as antibiotics for pneumonia, or continuations of existing conditions that require a prescription refill. The probability of a medical event

occurring q depends upon observables X such as demographics (age, risk scores, historic Medicare Part D usage, etc.) and observable environment characteristics (e.g. time of the year). It may also depend upon unobservable patient characteristics and health shocks. However, essential to the model is the idea that the actual probability of a true medical event occurring should not change due to the arbitrary insurance coverage region (defined by the cumulative total spending Z_w up until week w) imposed by the patient's insurance plan coverage—that is the model assumes $E(q|X, Z_w) = E(q|X) \forall r$.

Once a patient receives a medical event, they have a choice whether to spend on prescriptions and a choice on the exact amount of out-of-pocket and total prescription spending. Within the second stage prescription filling decision, the patient has some flexibility in the total prescription costs and thus their out-of-pocket costs. They can choose branded or generic drugs, or they can potentially ask their doctor to prescribe alternative drugs in a class of drugs that treat their medical shock. I assume that beneficiary spending, conditional on receiving an event, falls in some truncated distribution. Willingness to consume given a medical shock is s, and the dollar amount of prescriptions patients actually purchase is censored at 0, and the probability $Pr(s > 0) = qPr(B(s|X) > PMC(s, Z_w|X))$. The total payment amount s depends on the patient's observable and unobservable characteristics X, but also depends on the region r of the beneficiary's insurance plan and the distance in spending to them. Changes in a patient's total spending between insurance regions is meant to capture changes in the patient's expectations of or response to out-of-pocket marginal price changes.

From what has been laid out so far, it is clear why a censored regression spending model such as a Tobit would be inappropriate. There are legitimate zeros when

beneficiaries do not have prescriptions to refill or health events, where the benefit of any drug spending is minuscule and the cost would require a doctor's visit for a new prescription.

Rather than estimating the more complicated model of both the choice to spend and spending amount, this paper focuses on the empirical probability of observing non-zero spending in a week. This probability is represented by $1 - p' = 1 - q + qP(s \leq 0)$ the probability of observing zero spending either because the beneficiary did not have a claim, or because she had a claim and she chose not to fill the claim.

This paper takes a reduced-form fixed-effects approach that differs from the prior literature in order to control for the heterogeneity in the enrollees' spending patterns shown in Figure 1.6. Einav et al. (2015) and other papers take a structural approach that imposes substantial assumptions on beneficiary behavior. Along with a dynamic optimization model, Dalton et al. (2018) run simple linear regressions with fixed effects to examine the flat effect of being within $110 of and in the donut hole on individual's average weekly spending (and other measures of spending). This paper improves upon their reduced-form approach by applying it to a four-year panel of observed weekly spending patterns over a much larger number of beneficiaries to reduce the negative dynamic panel bias with fixed effects (see Section 1.5.3). Also, rather than using a single linear indicator of being near the coverage gap, this paper's analysis allows for a flexible cubic spline to characterize beneficiary's spending patterns. This is described in detail in Section 1.5.1.

Among the reduced-form literature, other papers have handled the heterogeneity in other ways. Kowalski (2016) uses a quantile regression with an instrumental variable to analyze the change in people's healthcare spending when their year-

end marginal prices change due to accidental (and assumed exogenous) injuries to family members. Since Medicare consists of only individual plans, this paper does not use a similar instrumental variable approach. Abaluck et al. (2018) effectively net out individual fixed effects by levering their panel data and observing the difference in individual spending due to plan changes between different years. However, their analysis is purposefully focused on spending for individuals who are unlikely to cross coverage regions and is thus limited at spending kinks. Joyce et al. (2013) compares the difference in the spending patterns of patients who have standard Medicare Part D Plans with plan non-linearities with the patients who receive low-income subsidies (LIS) and thus do not have significant coverage gaps. However, using the spending patterns of the LIS as a baseline comparison group for non-LIS patients may ascribe inherent differences between the groups to the plan coverage structure.

1.5.1 Estimation

In order to identify how the probability of claims occurring responds to the nonlinear marginal prices of the Medicare Part D nonlinear contract, this analysis takes a fixed-effects-regression approach with a dynamic panel.

Model 1 is represented by Equation 1.1, a linear probability model and the main estimation approach for this paper. Suppose Medicare Part D individual claims are grouped on a weekly level with spending s_{iyw} in year y and week w. Let the occurrence of spending $o_{iyw} = I(s_{iyw} > 0)$ be a binary variable. Let

$$o_{iyw} = \alpha_i + \boldsymbol{\gamma}\mathbf{X}_{iy} + f(Q_{iyw}, Z_{iyw}, \tilde{Z}_{iyw}) + \tau_y + \varepsilon_{it}. \tag{1.1}$$

The variable $Z_{iyw} = \sum_{u=1}^{w-1} s_{iyw}$ is a measure that represents the cumulative total expenditures within year y up until week w, and the variable $\tilde{Z}_{iyw} = \sum_{u=1}^{w-1} OOP_{iyw}$ is a measure that represents the cumulative total out-of-pocket expenditures in the same time frame. The variable $Q_w \in \{1, 2, 3, 4\}$ indicates whether a week is in the first, second, third, or fourth set of 13 consecutive weeks in a year (simplified to be called quarter variables). While the data is observed on a weekly level and the heuristic in Section 1.4 was also presented on a weekly level, this estimation aggregates the time fixed-effects on the quarter level. The variable α_i is an individual fixed effect that is constant across years and nests any gender, race, and age in 2009 information about the beneficiary. $\mathbf{X}_{iy}$ is the set of year-varying individual demographics that include beneficiary RxHCC 2011 demographic and risk scores are based off of their known health conditions from the prior year.[14] The purpose of the risk scores are to reimburse prescription spending, and as such they are an important measure to capture any between year changes in beneficiary's probabilities of spending. The variable τ_y is the year fixed effect y that is the same for all individuals.

The variables Z_{iyw}, $\tilde{Z}_{iyw}$, and Q_{iyw} enter the estimation through a flexible functional form:

$$f(Q_{iyw}, Z_{iyw}, \tilde{Z}_{iyw}) = \sum_{q=1}^{4} \sum_{r \in \mathcal{R}} I(Q_{iyw} = q) \left(\eta_{qr} + g(dZ^r_{iyw}) + \tilde{g}(d\tilde{Z}^r_{iyw}) \right) \qquad (1.2)$$

This function f is parameterized as a piecewise function of 12 restricted cubic splines for each of the two cumulative spending measures over the interaction of

[14] While the demographics score is a function of an individual's fixed characteristics, it is a non-linear function and thus still included in this analysis. Within a year, the demographics score plus the risk score is a relative measure of the riskiness of an individual compared to others in Medicare Part D. People who have a demographics plus risk score equal to 1 are considered to have an average reimbursement liability.

four quarters of the year and three separate spending regions: ICR, donut hole, and catastrophic region. The limits that define the spending regions, $\overline{ICL}_y$ and $\overline{OOPT}_y$, change every year. Hence, in order to study the response as beneficiaries reach these limits, define $dZ_{iyw} \equiv Z_{iyw} - \overline{ICL}_y$ and $d\tilde{Z}_{iyw} \equiv \tilde{Z}_{iyw} - \overline{OOPT}_y$. The cumulative spending measures are denoted within each region $r \in \mathcal{R}$ as dZ^r_{iyw} and $d\tilde{Z}^r_{iyw}$, where the regions are defined as functions of both the cumulative total spending and the cumulative out-of-pocket spending.

$$r = \begin{cases} ICR, & \text{if} \quad dZ < 0 \quad \& \quad d\tilde{Z} < 0 \\ DonutHole, & \text{if} \quad dZ \geq 0 \quad \& \quad d\tilde{Z} < 0 \\ Catastrophic, & \text{if} \quad dZ \geq 0 \quad \& \quad d\tilde{Z} \geq 0 \end{cases} \tag{1.3}$$

The functions g and $\tilde{g}$ denote the cubic splines for dZ and $d\tilde{Z}$ respectively.[15] Each cubic spline is a natural spline and has 3 knots located at the 10, 50, and 90 percentiles suggested by Harrell (2001).[16] The cumulative total and cumulative out-of-pocket expenditure measures both define the spending regions and allow us to understand how beneficiaries change their spending patterns as they approach the regions across the year. Table 1.7 displays the percentile values of the cumulative spending measures dZ and $d\tilde{Z}$ that are used to define the spline knots.

This functional form allows the slope (and form) of the relationship between the cumulative spending measures and the beneficiary weekly claim probability to vary separately in each of the 12 region-quarter grids. It assumes that the relationship in f is the same across all four years, with the year effects only altering the level of spending across all beneficiaries through τ_y in Equation 1.1. Alternative models

[15] The function differs for dZ and $d\tilde{Z}$ because of the exact placement of the knots and thus the spline function notation differs slightly. Let k_i, $i = 1, 2, 3$ be the knot values, then the $g(\mathcal{V})$ function is a linear regression of $V_1 = \mathcal{V}$ and $V_2 = \frac{(\mathcal{V}-k_1)^3_+ - (k_3-k_2)^{-1}\{(\mathcal{V}-k_2)^3_+(k_3-k_1) - (\mathcal{V}-k_3)^3_+(k_2-k_1)\}}{(k_3-k_1)^2}$.

[16] Section 1.5.4, discusses how this specification was selected for over alternative cubic splines with 3 knots and one with 4 knots.

Table 1.7: Percentile values

	dZ			$d\tilde{Z}$		
Percentile	ICR	Gap	Catastrophic	ICR	Gap	Catastrophic
5	-2,849	48	3,777	-4,674	2,976	11
10	-2,827	100	4,034	-4,593	-3,695	20
35	-2,565	432	5,258	-4,443	-3,848	84
50	-2,340	717	6,373	-4,342	-2,933	145
65	-2,565	1114	8,362	-4,233	-2,561	252
90	-981	2426	24,435	-3,882	-1,349	1,074
95	-577	2977	38,488	-3,720	-835	1,768

Note: Shows select percentile values of dZ and $d\tilde{Z}$ in each coverage region. These percentile values are used in determining the location of the knots in the estimation results.

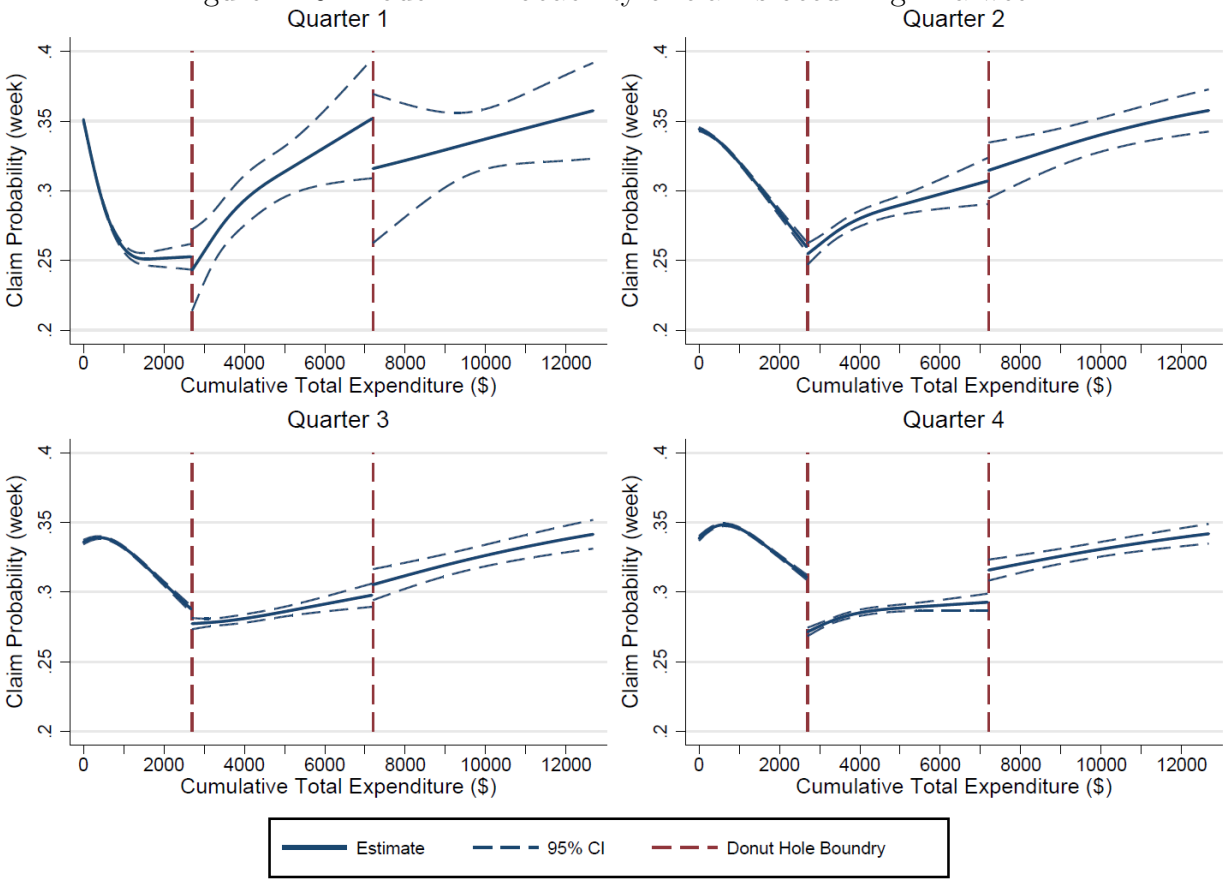
were considered such as ones that included an individual-year fixed effect for more flexibility. However, such models were rejected in order to mitigate the potential bias with dynamic panels and fixed effects. See the discussion in Section 1.5.3.

1.5.2 Baseline estimation results

Figure 1.10 and Table 1.8 present the results of Model 1's linear probability fixed-effects estimation using a cubic spline. Errors are clustered at the individual level. Rather than displaying the coefficient estimates of the piecewise f function, the figure displays the predicted values from those estimates that describe the effect the cumulative spending values on the probability of filling a prescription claim in a week. The remaining coefficient estimates for the beneficiary risk scores and year (γ and τ_y) are included in Table 1.8.

The main takeaway from this analysis is that patients do anticipate and respond to the pricing incentives in the Medicare Part D contract. This analysis indicates that on average patients decrease their spending patterns far in advance of entering

Figure 1.10: Model 1: Probability of claims occurring in a week



Note: Displays the predicted values of the claims occurrence probability $\hat{o}_{iyw}$ from the fixed-effects panel regression of Equation 1.1 on the beneficiaries in the No Deductible sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values. The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure, which increases with the cumulative total expenditure at the average coinsurance rates presented in Table 1.4. The probabilities are predicted assuming that beneficiary has the sample average risk and demographic scores from 2009, 0.4753 and 0.4196 respectively. Each line segment represents a prediction made assuming the year is 2009 with 2009 spending limits and across time average individual risk and demographics. Since the estimates are from a linear fixed-effects-regression, the predicted values are plotted assuming the mean fixed-effects coefficient, which is zero.

the donut hole and begin to increase it prior to exiting; however, these changes are smooth except in the end of the year. This analysis did not find the behavioral break at the entrance to the donut hole early in the year that previous literature has suggested.

Table 1.8: Model 1: Impact on the probability (%) of a claim in a week

Coefficient	Estimate	Std. Error
RxHCC Risk Weight	3.57	0.11
RxHCC Demographic Weight	-26.71	6.12
2010	0.30	0.05
2011	0.61	0.05
2012	1.42	0.08
N	18,585,632	

Note: The estimated coefficients on the risk scores $\mathbf{X}_{iy}$ and year-time dummies τ_y from the fixed-effects panel linear probability regression of Equation 1.1 on the beneficiaries in the No Deductible sample. All estimates in the table are significant at less than the .1% level.

The coefficients of individual risk and demographics scores are highly significant, but the scale of the effects may be economically small. On average for the predicted sample, the point estimates of the probability of a claim occurring in a week all fall in the range of around 25-35%. If these predicted values of weekly claim probabilities are extrapolated multiple weeks, this translates into an economic meaning of visiting the pharmacy once every 4 weeks versus once every 2.9 weeks. These differences would be higher for individuals with lower risk scores and the differences would be smaller for high-risk types.

Holding individuals and other characteristics constant, if a patient's risk score increases by one standard deviation (approximately 0.3), the probability of claims being observed in a week is expected to increase by approximately 1%. Riskier patients result in higher prescriptions as expected. Also, if a patient's demographic score increases by one standard deviation (approximately 0.01), the probability of observing at least one claim in the week is expected to drop by 0.25%. At first glance the negative coefficient on the demographics score appears counter intuitive; however, this can be explained since the demographics score is a nonlinear function of age, race, and other fixed characteristics. Holding all else constant, the demographics score actually decreases as age increases. The positive and signifi-

cant coefficients on the year fixed effects also capture both the effect of aging on the sample population and any year trends that lead to increases in medication purchasing frequency. Individuals who are older and in 2012 are 1.4% more likely to have claim in a week than in 2009. Because this is a fixed-effects-regression on a balanced panel in all four years, this regression does not differentiate the effects of aging versus the year fixed effects.

Figure 1.10 depicts in four panels the predicted probability of a claim occurring in a week in each of four quarters (13 week periods) in the year as the beneficiary's cumulative total spending (Z_{iyw}) changes.[17] Within each panel, the predicted probabilities are displayed as a piecewise function with the first segment representing the initial coverage region, the second the coverage gap, and the third the catastrophic region. Further these estimates are predicted assuming other coefficients are held constant with year 2009 fixed effects, average demographics and risk scores for the sample, 2009 ICL and OOPT limits, and average coinsurance rates in each spending region. Because the prediction assumes fixed coinsurance rates, the cumulative total is a one-to-one mapping of the cumulative out-of-pocket expenditures. The example out-of-pocket total of $4,350 translates to a cumulative total expenditure amount of $7,200.

Focusing first on the Quarter 1 panel, the probability of a claim occurring in a week are tightly estimated for low cumulative expenditures and decreases from 35% to 25% prior to spending even $1,000, indicating that high spenders may anticipate entering the donut hole. The point estimates as individuals cross from the ICR to the coverage gap do not indicate a discontinuity and the probability of claim occurrences increase prior to entering the catastrophic region. The point estimate at the entrance to the catastrophic region indicates that individuals decrease their

[17]The cumulative out-of-pocket expenditures also change here but this is not displayed.

spending. The confidence interval around these estimates do become progressively wider as the cumulative total expenditure values increase, because there are few individual-week observations at such high spending levels early in the year. Even with the wide confidence intervals, the claim occurrence probability is lower for individuals who just entered the donut hole than for individuals who have spent a thousand more. The predicted values show a drop in claim occurrence frequency when exiting the donut hole into the catastrophic region, however, the confidence intervals around these predictions are wide at approximately 10 percentage points and are not significant.

The behavior observed in Quarter 2 and 3 are similar to Quarter 1. There is a significant decrease and then an increase in the claims occurrence probability in the ICR and donut hole, respectively, as the cumulative total expenditure increases. These changes are smooth, and there is no evidence for a discontinuity in the spending measure at either the entrance to the donut hole nor the start of the catastrophic region. The predicted values are generated off of more precise estimates in the later quarters with a maximum confidence interval around the entrance to the catastrophic region of 5% and 2.5% in the second and third panels respectively. The predicted values in Quarter 2 and 3 differ from Quarter 1 in that the decrease in the cumulative total spending in the ICR is not as steep, and the lowest probability occurs at a higher cumulative total expenditure in the later periods.

In Quarter 4, the probability of spending is still in the 25% to 35% range for cumulative total expenditures below $12,000, but the predicted spending values differ at specific cumulative total spending and are estimated with very tight confidence intervals. They indicate sharp discontinuities in spending at the entrance to the

donut hole with a 4% drop in the probability of claims occurring in a week, and they indicate a 3% increase in claims occurrences upon entering the catastrophic region. The slope of the claim occurrence probabilities in the ICR and donut hole regions are a much lower magnitude and less economically significant even as they are estimated with more precision as time progresses. In addition, the slope within the ICR region appears to be convex.

This pattern supports the theory and simulations from Section 1.2, which predict that discontinuities in the probability of spending are most likely in the last time period because there is less ambiguity as to the beneficiary's end-of-year marginal costs. The lower magnitudes of the changes in the claim occurrence probability within each coinsurance region also reflect the theory that the perceived marginal cost should approach the actual coinsurance rates which are flat within a region. The convexity in the ICR that begins in Quarter 3 and is more evident in Quarter 4 was not predicted from the heuristic marginal price applied to data, but it is consistent with theory. At the end of the year, beneficiaries would have more information on their end-of-year region, and in the last months of the year, enrollees in the low values of the ICR may decrease their perceived marginal price as the probability of entering the donut hole decreases (and then increase their spending).

1.5.3 Bias in a dynamic panel with fixed effects

The estimates of fixed-effects models applied to a dynamic panel are known to be biased if the number of time periods (T) is small and the cross sectional size of

the panel (N) is large (Nickell, 1981).[18] This model is affected by Nickell bias because the cumulative total and out-of-pocket expenditures Z_{iyw} and $\tilde{Z}_{iyw}$ are both functions of the lagged dependent variable, the claim observation o_{iyw}. The paper does take precautions to reduce the influence of this bias on the analysis. First, the analysis spans beneficiary behavior over four years, or $T = 208$ weeks, a longer time span than normally cited in the literature. As Nickell (1981) demonstrates with a simple lag, as $N \to \infty$, the inconsistency of the estimated lagged parameter is of the order $1/T$. So while the number of beneficiaries in the No Deductible sample $N = 89,354$ is large, the potential bias with a larger T is greatly reduced.

The classic Nickell bias as it applies to the coefficient on the lag of the dependent variable is negative, and the coefficient on the cumulative spending measures in each region would be similarly negatively biased. To verify, the direction of the bias as it applies to the specific types of lags in this estimation are simulated with noise and presented in the Appendix Section A.3. Any bias in the estimates are more likely to be observed at low cumulative spending levels and leads to a more negative slope. The bias alone does not lead to discontinuities in the effect of the cumulative spending measures on the claim occurrence probability.

Other estimation approaches that would have circumvented the bias problem of the fixed-effects approach such as that presented by Anderson and Hsiao (1982) or Holtz-Eakin et al. (1988) (popularized and more commonly known as Arellano and Bond (1991)) have their own problems. These methodologies use a first-differences approach to net out the fixed-effect α_i and then use further lags of the lagged variable (in this case the Z_{iyw} and $\tilde{Z}_{iyw}$) as instruments in a 2SLS and GMM style

[18]Nickell (1981) highlighted that the time demeaning operation of fixed effects in a dynamic panel data model $y_{it} = \alpha_i + \beta y_{it-1} + \epsilon_{it}$ leads to a transformed regression model $y_{it} - \bar{y}_i = \beta(y_{it-1} - \bar{y}_{it-1}) + (\epsilon_{it} - \bar{\epsilon}_i)$ where the $\bar{y}_i, \bar{y}_{it-1}, \bar{\epsilon}_i$ indicate time averages. The error terms $(\epsilon_{it} - \bar{\epsilon}_i)$ and regressors $(y_{it-1} - \bar{y}_{it-1})$ are correlated even as $N \to \infty$.

estimation respectively. While not biased, Anderson and Hsiao (1982) does not use all available data and can result in imprecise estimates. Further, because of the large size of the data, the number of interactions of Z_{iyw} and $\tilde{Z}_{iyw}$ terms, and the long length of the panel, the Arellano-Bond methods are not computationally tractable in this specification.

1.5.4 Robustness

I consider additional models to address some of the concerns of model specification and find results that support the initial findings. Alternative splines of the cumulative total and cumulative out-of-pocket spending measures are included as a falsification test to verify that the functional form definition is not driving the result. I also consider an alternative spending measure (the count of the number of claims in a week) and present both the results of a linear and Poisson model as an alternative to the linear probability model.

Falsification test

In order to more fully verify the results of the estimation, this paper conducted falsification tests with different spline functions and discontinuities allowed at different points.

Model 1 used a cubic spline with 3 knots located at the 10, 50, and 90 percentile values for dZ, the cumulative total expenditures (centered on the ICL), and for $d\tilde{Z}$, the cumulative OOP expenditure (centered on the OOPT). One potential concern is that the cubic spline with 3 knots may over-smooth the beneficiary's response especially close to the region boundaries. Two alternative splines were tested in

models called False 1 and False 2. One included three knots located at the 5, 50, 95 percentile values for both regressors and another with four knots at located at the 5, 35, 65, and 95 percentile values for both regressors. The knot values are displayed in Table 1.7. Model 1 was selected over these other two models by both the AIC and BIC criteria.

Another approach to verifying the accuracy of Model 1 was taken by introducing additional potential discontinuities in the domain. In False 3, the following linear probability model was conducted where the initial coverage region is broken down into two separate regions $ICR1$ and $ICR2$. Then the cumulative measures that enter Equation 1.1 and 1.2 are dZ^r_{iyw} and $d\tilde{Z}^r_{iyw}$ where $r \in \mathcal{R}'$ such that

$$r = \begin{cases} ICR1, & \text{if} \quad dZ < -1000 \quad \& \quad d\tilde{Z} < 0 \\ ICR2, & \text{if} \quad -1000 \geq dZ < 0 \quad \& \quad d\tilde{Z} < 0 \\ DonutHole, & \text{if} \quad dZ \geq 0 \quad \& \quad d\tilde{Z} < 0 \\ Catastrophic, & \text{if} \quad dZ \geq 0 \quad \& \quad d\tilde{Z} \geq 0 \end{cases} \tag{1.4}$$

Equation 1.1 is estimated with the regions in $\mathcal{R}'$ and standard errors clustered at the individual level.

Table 1.9 displays the estimated coefficient values for the year fixed effects and risk scores for False 1-3. The predicted values of the claims occurrence as the cumulative total and out-of-pocket expenditures change are displayed in Figure 1.11 for False 3. The figures for False 1-2 are displayed in the Appendix Tables A.2 and A.3 respectively.

The results from these falsification tests are all very similar to the results from

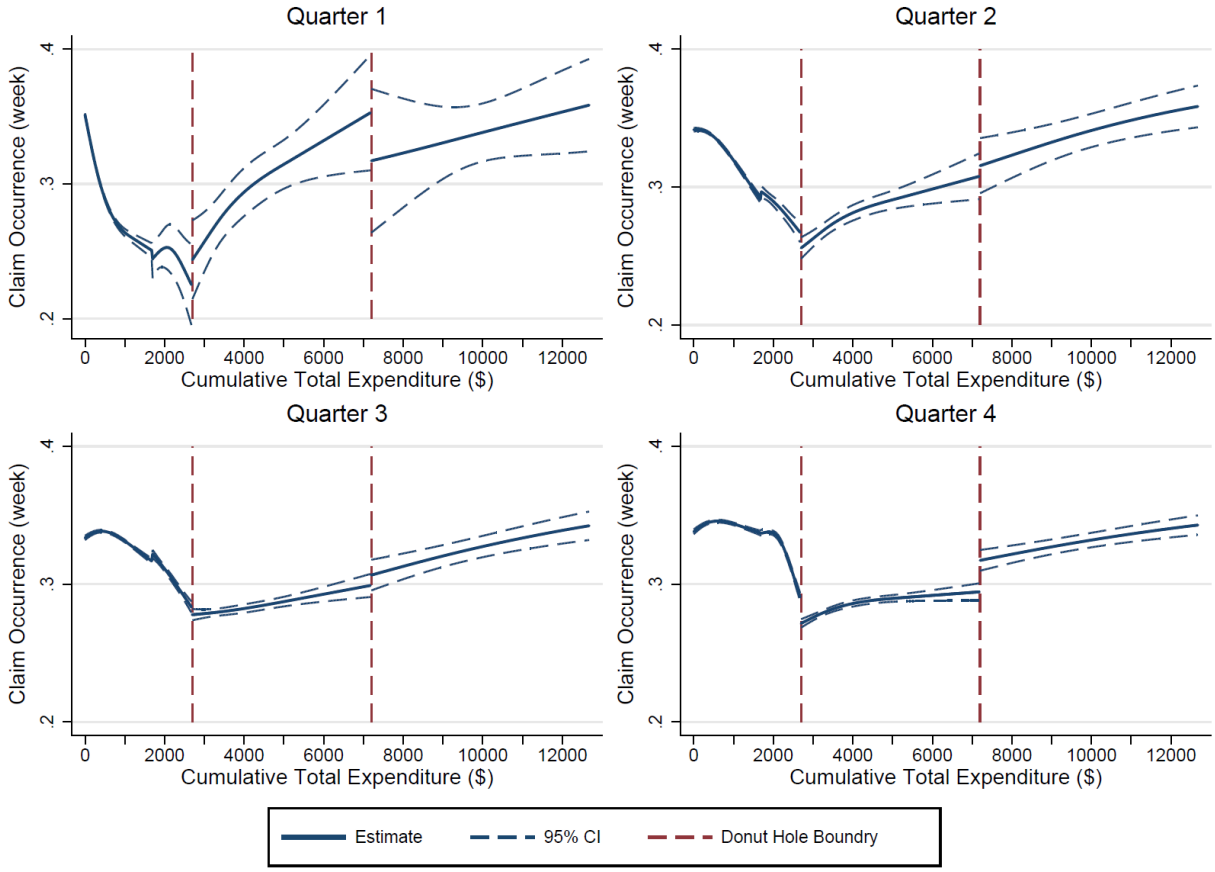
Table 1.9: Falsification 1, 2, 3: Impact on the claims occurrence probability (%)

	False 1		False 2		False 3	
Coefficient	Estimate	SE	Estimate	SE	Estimate	SE
RxHCC risk weight	3.57	0.11	3.57	0.11	3.58	0.11
RxHCC demo. weight	-26.63	6.12	-26.73	6.12	-26.58	6.12
2010	0.32	0.05	0.31	0.05	0.23	0.05
2011	0.63	0.05	0.61	0.05	0.53	0.05
2012	1.41	0.08	1.43	0.08	1.27	0.08
N	18,585,632					

Note: The estimated coefficients on the risk scores $\mathbf{X}_{iy}$ and year-time dummies τ_y from the fixed-effects panel linear probability regression of Equation 1.5 under different spline and region assumptions. In False 1 and 2, the spline of the cumulative total and cumulative out-of-pocket expenditures are defined by a cubic spline with four and three knots at the 5, 35, 65, 95 and 5, 50, 95 percentiles respectively. The percentile values are shown in Table 1.7. In False 3, the spline is defined as in Model 1, but it introduced an additional potential discontinuity at \$1000 less than the initial coverage limit (ICL) in the estimation. All estimates in the table are significant at less than the 0.1% level.

Model 1 and did not have a large impact on the coefficient estimates or the predicted claims occurrence values as the cumulative expenditures changed. In False 3, while we allowed for a discontinuity at ICL-\$1000, neither a statistically significant nor economically significant discontinuity was estimated at this point. One minor but noticeable difference between these models and Model 1 is that in Quarter 1, the point estimates of the predicted values of the claims occurrence at the high values of the ICR are slightly higher than in Model 1, but they still fall within the confidence interval of the original estimates. Also from False 3, we observe that the predicted values in $ICR2$ do have a larger confidence interval than the predictions over the same domain in Model 1. This pattern does highlight the low number of observations of such high cumulative spending in the first quarter of the year.

Figure 1.11: False 3: Probability of claims Occurring in a week



Note: Displays the predicted values of the claim occurrence probability $\hat{o}_{iyw}$ in a week from the fixed-effects panel regression of Equation 1.5 under the assumptions of False 3 on the beneficiaries in the No Deductible sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values. The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. See Figure 1.10 for the exact values used to generate the prediction.

Fill count

The original dependent variable in Model 1 is the binary variable of the occurrence of any claim in a week and was chosen to highlight the extensive margin of the beneficiary's prescription purchasing decision. However, beneficiaries often have multiple health conditions and multiple prescription claims that they have the choice to fill in a week. Hence, I also study the results using an alternative

consumption measure of the number of prescriptions filled in a week.

Alternative specifications use the number of claims observed in a week n_{iyw} as the dependent variable in both a linear regression in Model 2 and one that assumes a Poisson regression in Model 3. The fixed-effects linear and Poisson regressions face the same critique due to dynamic panel bias as the prior Model 1 linear regression.

The linear model is described in Equation 1.5 while the Poisson model in described in Equation 1.6 with a log-linear regression model where λ represents the mean number of claims.

$$n_{iyw} = \alpha_i + \boldsymbol{\gamma}\mathbf{X}_{iy} + f(Q_{iyw}, Z_{iyw}, \tilde{Z}_{iyw}) + \tau_y + \varepsilon_{it} \tag{1.5}$$

$$\ln \mathbb{E}(n_{iyw}) = \ln \lambda_{iyw} = \alpha_i + \gamma\mathbf{X}_{iy} + f(Q_{iyw}, R_{iyw}, Z_{iyw}) + \tau_y \tag{1.6}$$

For both models, the f retains the same functional form as in Equation 1.2 and the $\mathbf{X}$ variables are the same as in Model 1.

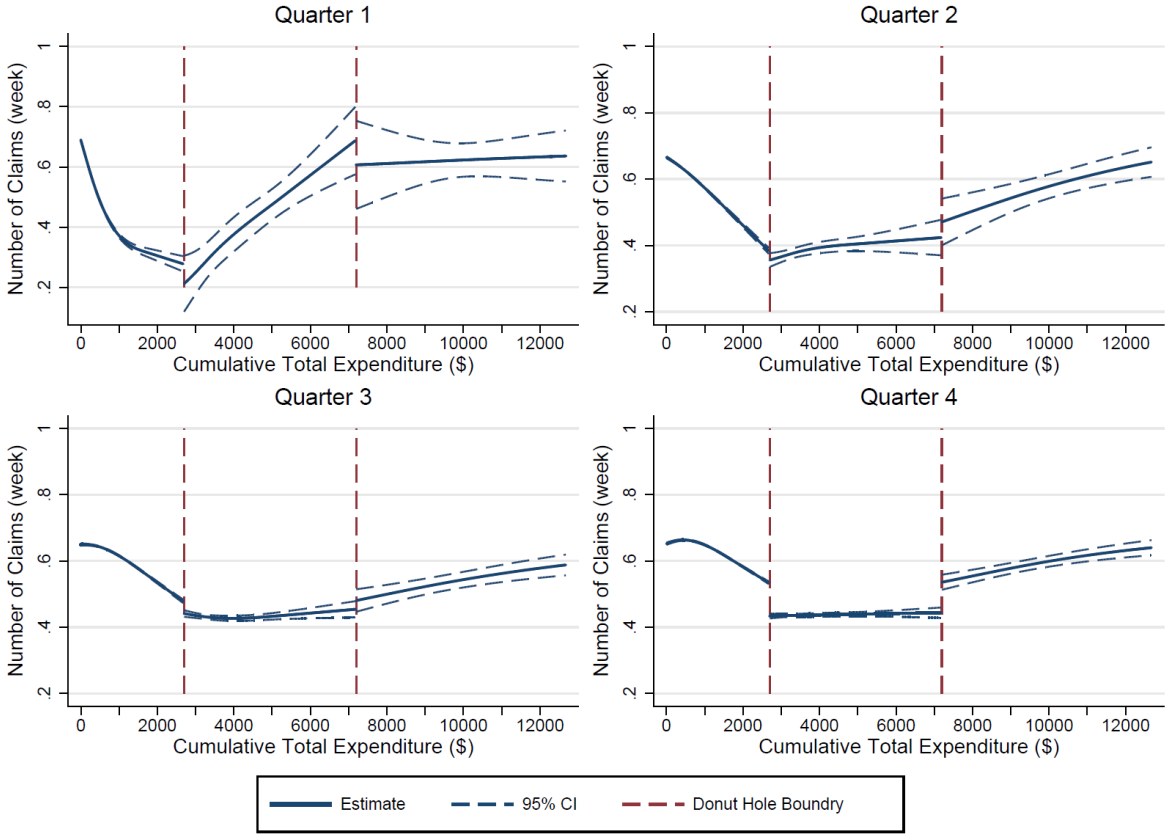
Table 1.10: Model 2 and 3 estimates: Impact on the number of claims in a week

	Linear Estimates (x100)		Poisson Estimates	
Coefficient	Estimate	Std. Error	Estimate	Std. Error
RxHCC Risk Weight	8.37	0.26	0.12	0.004
RxHCC Demographic Weight	-34.31	13.6	-0.70	0.222
2010	-0.41	0.1	-0.0017	0.001
2011	0.25	0.12	0.0103	0.002
2012	1.11	0.02	0.0221	0.003
N		18,585,632		

Note: The estimated coefficients on the risk scores $\mathbf{X}_{iy}$ and year-time dummies τ_y from the fixed-effects panel linear probability regression of Equation 1.5 and the Poisson regression of Equation 1.6 on the No Deductible sample. All estimates in the table are significant at less than the 5% level. While the estimates from both models are presented in the same table, the interpretation for the Poisson estimates is multiplicative of the exponential of the estimate.

Qualitatively, the estimates are very similar to those estimated from Model

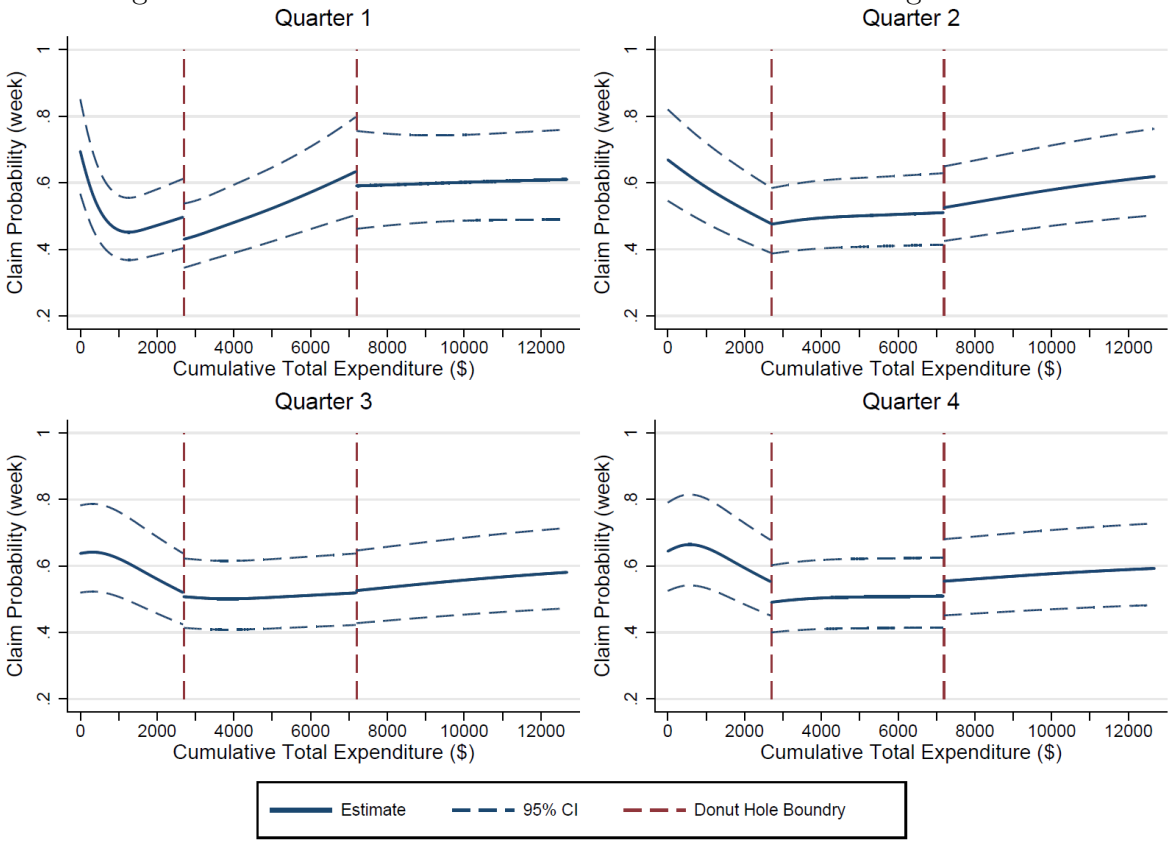
Figure 1.12: Model 2: Incidence of claims occurring in a week



Note: Displays the predicted values of the average number of claims in a week $\hat{n}_{iyw}$ from the fixed-effects panel regression of Equation 1.5 on the beneficiaries in the No Deductible sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values. The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. See Figure 1.10 for the exact values used to generate the prediction.

1 where the dependent variable was the occurrence of claims. In Models 2 and 3, the level of claims incidence is much higher, which reflect the fact that many beneficiaries file multiple claims in a week. The point estimates in both Models 2 and 3 are very similar and range from approximately 0.3 to 0.7, translating to a claim occurring every three weeks to every 10 days. The estimates from Model 3 are slightly less extreme than Model 2, particularly in Quarter 1. The confidence intervals in Models 1 and 2 are similarly scaled and precise (except in the high values of Quarter 1), but the confidence intervals from Model 3's Poisson

Figure 1.13: Model 3 Poisson: Incidence of claims occurring in a week



Note: Displays the predicted values of the average number of claims in a week $\hat{n}_{iyw}$ from the fixed-effects panel regression of Equation 1.6 on the beneficiaries in the No Deductible sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values. The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. The values used are similar to those from Figure 1.10; however, it differs in that the fixed effects used to generate these values is not zero. Unlike the linear model, the mean fixed-effects coefficient in the Poisson model is not zero, so the predicted values are scaled by the exponential of the mean fixed effect estimated $\alpha_i = -0.145$.

regression are much wider than those from the linear case averaging about a 0.2 band around the estimates.

The results from Models 2 and 3 support the findings from Model 1, even though a different spending measure is used. While there are changes in the average number of claims in each week as the cumulative spending measures increase, these changes are smooth except potentially in Quarter 4. Claims decrease prior

to entering the donut hole and increase prior to entering the catastrophic spending region, though both changes have lower magnitudes in the later quarters of the year.

1.5.5 Discussion

The conclusion from the estimation of Models 1-3 is that on average, individuals behave more optimally than some of the literature has found, because on average enrollees anticipated and responded to the pricing structure in Medicare Part D. This analysis is aggregated over many heterogeneous actors, and while the fixed-effects approach helps to control for individual unobserved heterogeneity, it may still complicate the analysis. That is to say, not that any one individual behaves optimally, but there does not appear to be a systematic decrease in spending in this sample prior to entering the coverage gap. Further, it is also important to understand why these results differ from some of the previous findings.

The results are aggregated over various heterogeneous variables. While the analysis included year fixed effects to control for differences in the levels of spending across year, slopes were not allowed to vary by year. It is possible that the slopes of the beneficiary claims response within the coverage gap differ across years, because the beneficiary plans differ across the years. The marginal costs in the 2011 and 2012 coverage gap are significantly closer to the ICR coinsurance levels, and thus this should result in fewer changes in spending. The expectation then by the estimation presented in Equation 1.1 and 1.2 across all four years is that the predicted f is "too flat" to describe 2009-2010 and not flat enough to describe 2011-12.

Further, in analyzing the f function estimates, it is important to remember that, while the fixed effects may "net out" the heterogeneity in the probability of spending between individuals, it does not control for the heterogeneity in health shocks or expectations that may exist at any point in time and cumulative expenditure level. This means that there are potentially heterogeneous responses to the nonlinear contract. For example, in any quarter, but particularly earlier in the year, the estimates for the change in the claims occurrence probability include both individuals who do not expect to end the year in the donut hole and decrease their spending as the probability of that outcome increases, and individuals who expect to end the year in the catastrophic region and may increase their spending as this outcome becomes more certain. This paper's conclusions on not finding a discontinuity at the region boundaries is still reasonable even in light of this potential heterogeneity. There is not a reasonable prior to think that individuals would have a discontinuous increase in their claims occurrence that perfectly offset another set of individuals discontinuous decrease.

Another source that could increase the heterogeneity in the response to the cumulative spending measures is the experience and knowledge individuals have with nonlinear pricing structures in general and each year's Medicare Part D plan specifically. Individuals with more experience (and no surprises) would be expected to have more constant spending patterns that result in a flatter overall f response curve. Further work should be done to understand how experience impacts the beneficiary response to the contract features.

Experience, sample selection, and estimation methods could be why this paper's estimates on finding a discontinuity at the donut hole differed from Dalton et al. (2018). Individuals in their dataset come from one year of observation in 2008,

are not on traditional Medicare, are known to be richer, and have higher spending than Medicare patients. The sample in this paper joined prior to 2008, and we observe four years of their choices. The fact that this paper's sample is from later years also increases the probability that they either had more personal experience with or opportunities to learn about Medicare Part D which started in 2006.

1.6 Conclusion

This paper builds on the existing literature on beneficiaries' dynamic response as they approach the many discontinuities in the Medicare Part D pricing structure. Throughout all health insurance, the government and insurers have significant control over the cost-sharing features that are responsible for non-linear pricing, and with the rise in health-care costs, these institutions are more likely to use them as cost-control measures. Unfortunately, the effect of these cost-control measures on beneficiary behavior and health is not fully understood. While the literature has identified sharp drops in spending particularly at the Medicare Part D coverage gap, trying to explain this behavior using time-discounting models has resulted in discounting estimates far lower than the broader economics literature.

The first main contribution of this paper is its discussion of an expected price model that uses the objective probability distributions of beneficiaries' end-of-year prices given their spending probabilities. The key takeaway from this heuristic model is that if beneficiaries know the objective probabilities of ending the year in each region for the population, the expected marginal price that a beneficiary responds to can be constructed and be used to make purchasing decisions. Because the end-of-year region probabilities have relatively smooth transitions (in all but

approximately the last 10 weeks of the year), beneficiaries' marginal price and then spending should also be smooth (except for the last weeks). Even if beneficiaries have inaccurate beliefs on their objective end-of-year probabilities or are present-biased, as long as they update those beliefs in each time period, a heuristic marginal price would not generate sharp spending changes unless there were sharp changes in probabilities.

A second significant contribution of this paper is the graphical representation of beneficiaries' claims rates conditioning on the cumulative sums of their total spending. Using separate linear probability and Poisson regressions with individual fixed effects to control for heterogeneity in base levels of spending, this paper illustrates beneficiary claims rates in a way that allows direct visual comparison with their heuristic marginal spending.

The estimation finds that there are significant changes to beneficiaries' claims rates, some of which are consistent with the predictions of the heuristic marginal price, but that those changes may not be economically significant. The lowest amount of claim rates in each quarter of the year broadly matches the cumulative total expenditures values that produced the lowest expected marginal prices. The changes in claims rates for a predicted beneficiary with average risk and demographic scores indicate that the economic magnitude of the claims occurrences are on the scale of filing claims every three weeks versus filing claims every four weeks. Further work to understand the welfare consequences of these reductions would be to analyze whether the claims changes were by discontinuing drugs entirely or just small delays in going to the pharmacy for refills.

This finding in the paper differs from empirical results in the prior literature. This difference may be due in part to the sample selection. Because this sample

involved individuals who retained similar plan structures through all four years, they are mechanically more likely to have experience with Medicare Part D, their plan structure, and their prescription needs than individuals in the papers in the prior literature. Further, individuals who do not switch between plans with different limits may be different onto themselves as people who have high inertia and do not switch plans or just happen to have expected spending amounts that align well with their chosen plans. Further work could explore whether experience with Medicare Part D promotes individuals to exhibit more optimal behavior, because that would imply that informational and educational programs could promote that behavior. A more detailed subsample analysis may be warranted.

The role of heterogeneous types of responses to the coverage gap should also be considered in future work. While this paper controlled for heterogeneous levels of claim rates for individual beneficiaries, it is likely that due to random health shocks or experience with the Medicare Part D pricing schedules, individuals may separately increase or decrease their claims in a predictable way that adds noise to this paper's estimates.

Another challenge to this paper's research question was the large variety of plans and coinsurance rates offered from 2009-2012 and the policy changes introduced by the Affordable Care Act to fill in the donut hole. The lower coinsurance rates within the donut hole in 2011 and 2012 could be partially responsible for the more stable spending estimates that are found throughout this paper. An obvious related research project would be to study the actual impact of the Affordable Care Act's policy of filling in the coverage gap and subsequent health outcomes to determine whether it truly impacted beneficiary spending rates. The plan variety that was a challenge for the analysis in this paper, could be a boon for follow-up

research projects.

CHAPTER 2

THE NATURE OF RISK PREFERENCES: EVIDENCE FROM EXPERIMENTAL INSURANCE CHOICES (WITH LEVON BARSEGHYAN AND TED O'DONOGHUE)

2.1 Introduction

The question of how individuals make choices under risk is central to most fields in economics. Recent research has made significant progress estimating risk preferences using field data. (For a survey see Barseghyan et al. (2018)). Of course, field data has its own limitations. In particular, there are two main issues. First, it is often impossible to know how accurately individuals can predict/estimate probability distributions in a given field context. Second, the estimated risk preferences derived from one market setting may not accurately predict household behavior in other markets. In fact, the issue of external validity is even broader, because ideally we aim to know whether risk preferences estimated in the field are systematically different from those estimated in the lab.

To address these issues, this paper leverages a controlled experimental setting, where we can manipulate subject's beliefs on probabilities within questions designed to mimic choices faced by households in typical property insurance markets. In particular, our benchmark experiment is designed to replicate deductible choice sets in three lines of property insurance: auto collision, auto comprehensive, and home all-perils. Each of our subjects faces deductible menus, prices, and (rounded) claim rates that are randomly drawn from the benchmark data used in the field study of Barseghyan et al. (2013) (hereafter referred to as BMOT).

This paper contributes to the literature on estimating risk preferences in the field by addressing some of the common assumptions made in field studies, specifically (i) the difference between subjective and objective beliefs and (ii) the role that ambiguity aversion might play in insurance choices. This paper also contributes to the experimental and behavioral literature on estimating risk preferences. By formulating our questions to be similar to the actual menus that households face in the real world, our experiment presents a real insurance choice and is less abstract than the binary or certainty-equivalent choices that are often used in the lab.

Section 2.2 describes our experimental design. Subjects are recruited both from Amazon Mechanical Turk and from a more traditional student population. Subjects are asked to make an insurance choice of a premium-deductible pair given a known loss probability. We have three between-subject treatments: Insurance, Ambiguity, and Abstract. The Insurance treatment is framed to closely resemble a real life insurance choice to study the difference between the field and the experimental context. The Ambiguity treatment is similarly framed as a real life insurance choice, but the subjects are given a range of possible loss probabilities without any information about which probability is likely to apply to understand the role that ambiguity aversion might play in their choices. Finally, the Abstract treatment is derived from a real life insurance choice, but it is framed as a generic lottery without any insurance context. The Abstract treatment is included to investigate whether the context of questions matters.

The motivation for the experiment is recent research that estimates risk preferences in the field such as BMOT, Snowberg and Wolfers (2010), and Sydnor (2010). They have empirically shown that rank-dependent probability weighting (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Quiggin, 1982) explains

consumer risk-taking decisions better than classic expected utility theory (EUT) or alternative models. However, in the applications of probability weighting to economic field data, these papers cannot distinguish between whether households overestimate the probability of a claim and whether they are indeed overweighting the probability of tail outcomes. Thus a key benefit of our experiment is to provide our subjects with the known loss probabilities or ranges of known loss probabilities.

Unlike most of the experimental literature on estimating risk preferences and probability weights, our experiment uses insurance language to provide four to six different risky options in each choice set. While some of the literature on estimating risk preferences does use insurance language (Bruhin et al., 2010), they elicit risk preferences through binary choices over risky gambles and certainty equivalents. Further, our experiment is concerned with how beneficiaries make choices over the low risks that occur with property insurance events, so our subjects are presented with choices where their risk exposure span 1-12% with more questions on lower probabilities. The majority of the experimental literature has elicited risk preferences over more widely spaced risks such as 5%, 10%, 25% and higher in Bruhin et al. (2010); 8.3%, 16.7% and higher in Prelec (1998); 1%, 5%, 10%, 15% and higher in Fox and Tversky (1998).

Section 1.3.2 includes summary statistics on the premiums and claim rates presented to subjects along with their deductible choices.

We evaluate the effects of our experiment and treatments in three ways. First, in Section 2.4, we take a traditional approach to evaluating a controlled experiment and compare how the three different treatments changed the premium-deductible choices made by our subjects. From our reduced form results, we find that the experimental subjects were significantly more likely to choose the least risky (lowest)

deductibles in the Ambiguity treatment and the most risky (highest) deductibles in the Insurance treatment.

Then, in Section 2.5.1 and Section 2.5.2, we build models that allow us to estimate our subjects' underlying homogeneous and heterogeneous risk preferences, respectively. Our models of risk preferences are built with both a standard risk aversion parameter and a nonparametric probability distortion function following BMOT. Because our experiment focuses on a dense set of claim probabilities in the range of 1-12%, our nonparametric probability distortion function provides better estimates of the shape of the true functional form than the experimental literature at low probabilities.

Our estimation of risk preferences reiterates the importance of probability distortions in explaining choices under risk. We recover substantial probability overweighting and negligible curvature in the utility. In our estimation of heterogeneous risk preferences, we also find substantially more heterogeneity in the probability distortion function than in the prior literature, concluding that a heterogeneous model is more suitable for our analysis.

The estimated risk preferences from the three treatments also support our reduced form results with a significantly and statistically higher probability distortion function estimated for the subjects in the Ambiguity treatment compared to the baseline Insurance treatment. The estimated probability distortion functions in the Abstract treatment were also higher than the Insurance treatment but not significant. The findings from the Ambiguity treatment imply that ambiguity aversion could have a substantial impact on the choices people made in our experiment and the lab, causing them to respond as if there was more risk.

Further, we show in Section 2.5.2, our estimates of risk preferences are consistent with predictions of a classic model of maxmin ambiguity aversion where subjects respond to the riskiest in a set of possibilities (Etner et al., 2012; Gilboa and Schmeidler, 1989).

In comparing our findings to the prior literature, our experimental estimates indicate there is substantially more probability distortions in our data than found in both the field data such as BMOT and Sydnor (2010) and prior experimental analysis (Prelec, 1998; Fox and Tversky, 1998; Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). The raw choices of our subjects, the reduced form results, and estimated heterogeneous risk preferences all indicate that there is substantially more heterogeneity in the premium-deductible choices and estimated probability distortions than found in field data such as BMOT. Our experimental estimates are most comparable to the risk preferences estimated by Bruhin et al. (2010), whose model of heterogeneous risk preferences allowed for multiple discrete "types" of subjects who acted according to EUT and CPT individuals with extreme probability distortions.

Section 2.6 concludes.

2.2 Experiment

The experimental laboratory setting is particularly suited to address common foundational assumptions made in studies estimating risk preference using field data. In field studies, researchers often assume subjects are fully informed agents and respond to their objective risks. Further, field studies on insurance commonly simplify away external influences on subjects choices, assuming that households' choice

of insurance plans are active choices that are made independently of other choices[1] and reflect true preferences as opposed to recommendations from insurance agents or family members. In the experimental setting, researchers can inform subjects of their risks directly and generally can exert greater control on subjects' tasks and perceptions.

2.2.1 Design

This paper designs a between-subject experiment where two groups of subjects were randomized into one of three treatments: Insurance, Ambiguity, and Abstract. Subjects were then presented with three insurance deductible-premium choices with known loss probabilities and a financially incentivized payoff. In each question, the choice can be decomposed into a lottery. Subjects are presented with the probability (or probability range) of a loss μ and a menu of premium-deductible pairs $\{(p_d, d) : d \in \mathcal{D}\}$. If subjects choose to pay premium p_d for deductible d, then they choose the lottery

$$L_d = (-p_d, 1-\mu; -p_d - d, \mu).$$

The experiment was designed with three distinguishing features that enable us to draw conclusions on how choices in the lab differ from the field. First, by using known probabilities, we distinguish between households' subjective beliefs on risk and their probability distortions.[2] Second, the experimental treatments

[1] The independence of choices is also known in the literature as narrow bracketing (Read et al., 1999)

[2] In our model and estimation, while we alleviate the issue of probability misperceptions, we continue to estimate a probability distortion function $\Omega(\mu)$ because we cannot distinguish between probability weighting and editing as proposed in Prospect Theory (Kahneman and Tversky, 1979;

differ in language in order to offer a direct comparison between the "insurance choice" offered in the Insurance treatment versus a general lottery in the Abstract treatment. The difference in language helps us draw conclusions on general differences between the lab and field. Last, in order to draw conclusions on the role that ambiguity aversion may play for individuals who are not aware of their risks, the Ambiguity treatment provides a range of loss probabilities as opposed to one discrete probability in the other two treatments. The treatments are further discussed in Section 2.2.3. Refer to the Appendix B.1 for additional experimental materials.

2.2.2 Subjects

Subjects for the experiment were recruited from two different sources and directed to our web-based experiment. On February 10-16 of 2014, we collected responses from 950 Amazon Mechanical Turk (MTurk) workers who were restricted to be located in the US, have greater than 1000 HITS, and have greater than 95% approval rating.[3] In addition, on April 16-24 of the same year, 689 undergraduate student subjects from Cornell University[4] were recruited via email solicitation.

The experimental survey presented to the MTurk and student subjects differed only in the payment schemes and some demographics asked of the students (gender and graduating class year). MTurk workers were paid 50 cents to complete the experiment and had a one-in-five chance of earning a bonus incentive payment in

Tversky and Kahneman, 1992).

[3]We chose the restrictions in an effort to guarantee quality work. With a high number of HIT completions, it is likely that these subjects are frequent MTurk workers, so our base and bonus payments would be considered generous to them.

[4]Student emails were acquired from the major lists in the Economics, Government, and Biology Departments, three departments with large student bodies.

the range of zero to five dollars; student subjects were not given a participation fee, and they had a one-in-twenty chance of a bonus incentive payment in the range of 0-$40 dollars depending on their answers and chance.

We chose to sample the two subject groups for separate reasons. Amazon Mechanical Turk has a large diverse set of workers who can quickly be recruited for low-cost experiments (Paolacci et al., 2010; Buhrmester et al., 2011; Berinsky et al., 2012; Mason and Suri, 2012). While evaluating studies that use Amazon Mechanical Turk samples for political science experiments, Berinsky et al. (2012) find that MTurk samples are often more representative than student and in-person convenience samples. Because of these features, many social scientists have turned to MTurk to run their experiments (Paolacci et al., 2010). In addition to the MTurk subjects, the student subjects serve as a more traditional subject pool for experiments.

While neither of these two experimental subject groups are expected to be representative of the households in BMOT or other insurance settings, we believe they offer insight into the decision making process.

2.2.3 Treatment design

Insurance treatment

In the baseline Insurance treatment, subjects were asked to choose a deductible and premium in each of three insurance scenarios: home (H), auto comprehensive (M), and auto collision (L), in random order. Figure 2.1 Panel A displays an example of the language used from the home scenario, and the language between

the scenarios only differs in the scenario given.

Within each insurance scenario j, subject i was presented with a probability of a loss μ_{ij} and a choice over deductible and premium menu pairs $\{(p_{d_{ij}}, d_{ij}) : d_{ij} \in \mathcal{D}_j\}$, where $p_{d_{ij}}$ is the premium associated with deductible d_{ij}, and $\mathcal{D}_j$ is the set of deductible options.[5]

The data for the claim rates and deductible-premium menus do differ between subjects and between insurance scenarios. For each subject, their claim rate and menu values were randomly selected and rounded from a random sample of 1,000 households present in the BMOT insurance data. For each of the 1000 households and three insurance contexts (1000×3), the data includes the estimated claim rates as calculated in BMOT's analysis in the range of 1-12% rounded to the nearest percent and the the insurance deductible-premium menus rounded to the nearest dollar terms. Note that this means $\mu_{ij} \in \{0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11, 0.12\}$. The auto collision insurance question offers deductibles $\mathcal{D}_L = \{100, 200, 250, 500, 1000\}$, auto comprehensive insurance offers deductibles $\mathcal{D}_M = \{50, 100, 200, 250, 500, 1000\}$, and home insurance offers deductibles $\mathcal{D}_H = \{100, 250, 500, 1000\}$ presented in dollar terms along with their respective prices.

Because these are deductible choices and not liability choices, subjects were not given the distribution of accident losses. Their exposure with insurance is only the deductible. Not purchasing insurance was not an option.

[5]These are deductible choices and not liability choices.

Ambiguity treatment

In order to investigate the possible role of ambiguity aversion in the decision making process over risky and ambiguous prospects, the Ambiguity treatment differs from the base Insurance treatment in that claim rate probabilities are specified within ranges. These ranges $\boldsymbol{\mu}_{ij}$ are deterministic sets centered around the claim rate μ_{ij} that subject i would have received in context j under the other treatments.

If a given $\mu_{ij} \geq 0.02$

$$\boldsymbol{\mu}_{ij} = \{\mu_{ij} - 0.02, \mu_{ij} - 0.01, \mu_{ij}, \mu_{ij} + 0.01, \mu_{ij} + 0.02\},$$

otherwise

$$\boldsymbol{\mu}_{ij} = \{\mu_{ij} - 0.01, \mu_{ij}, \mu_{ij} + 0.01\}.$$

Refer to Figure 2.1 Panel B for an example of the home insurance question from the Ambiguity treatment. We focus one of the treatments on ambiguity because the Ellsberg paradox illustrates that decision makers often prefer lotteries with known distributions rather than unknown ones (Ellsberg, 1961). In real life, it is possible that households have some sense of their true claim rate, be it a distribution or a range, but do not know the exact probability of a loss. If households do incorporate ambiguity aversion in their decision making process, it could explain some of the probability overweighting found in the field.

If subjects are indeed ambiguity averse, subjects in the Ambiguity treatment should choose less risky deductibles. In estimating their underlying risk preferences, conditioning on using μ_{ij} in the estimation, the ambiguity treatment should result in an estimate with a higher probability distortion function than other treatments.

Abstract treatment

The last treatment is referred to as the Abstract treatment. The language within this treatment differs from the previous two treatments in that the questions are stripped of the insurance context phrasing. Regardless of whether the household's claim rates and deductible-premium menus come from the home, auto collision, or auto comprehensive context, the questions presented to the subjects were identically worded. The purpose of this treatment is to be more closely comparable to past experimental work on choices over risky lotteries that employ abstract language. It also serves as a comparison to determine how the context of the questions can impact choices. Refer to Figure 2.1 Panel C for an example of the home insurance question from the Abstract treatment.

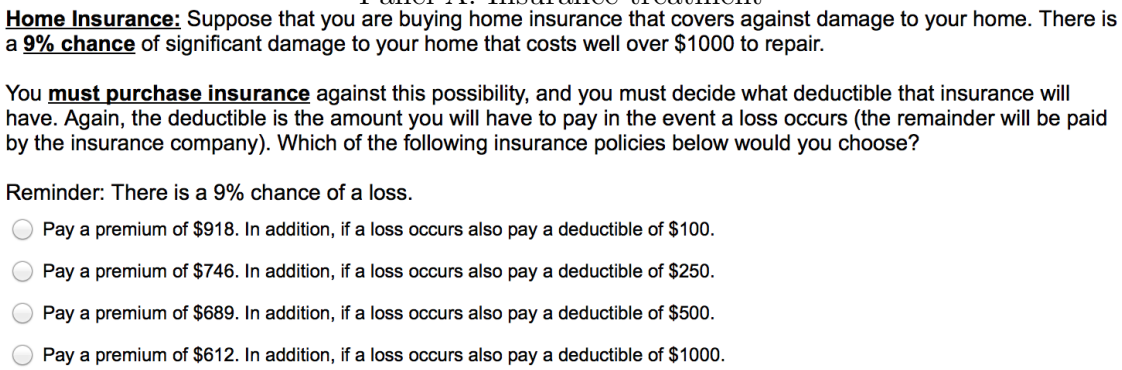
Incentives

For all treatments, subjects are incentivized to answer truthfully through a payment scheme that depends on the answers they choose. Before seeing the question, subjects are endowed with $2,000 laboratory dollars. They know that there is a chance (1-in-5 and 1-in-20 for MTurk and student subjects, respectively) that one of the three questions they receive is randomly selected to be played out, and subjects must pay the premiums and deductibles (if it applies) out of their laboratory dollars. The $2,000 laboratory dollar endowment is enough such that there is no scenario of a premium and deductible exceeding this amount.

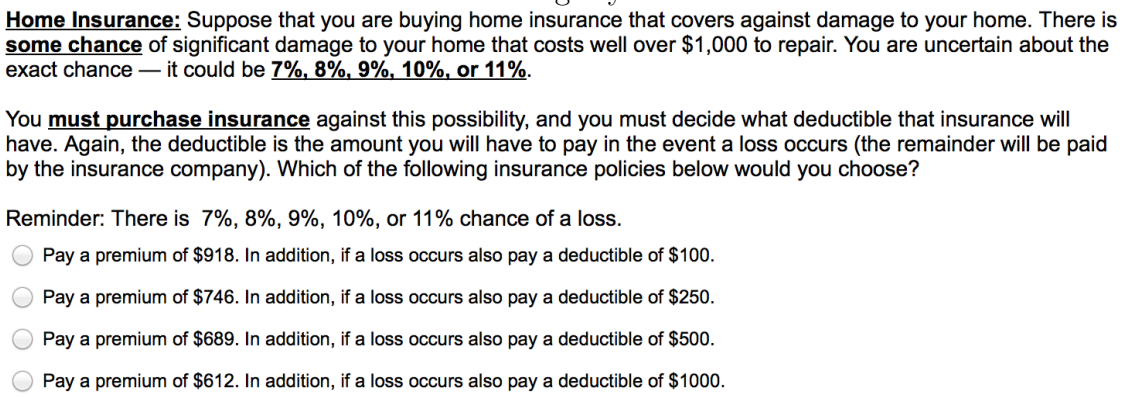
After completing the experiment, if subjects were selected for bonus payments, these subjects are issued a payment in actual dollars. They receive a conversion of their laboratory dollars to real dollars by dividing by 400 and 50 in the MTurk

Figure 2.1: Treatment designs in web experiment

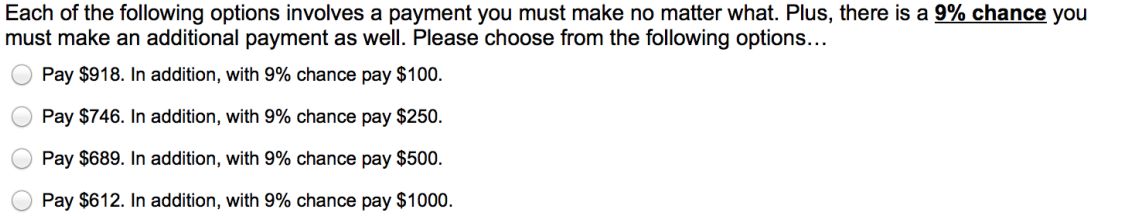
Panel A: Insurance treatment



Panel B: Ambiguity treatment



Panel C: Abstract treatment



Note: These questions all refer to the home insurance context, and similar language is used for the auto collision and auto comprehensive contexts.

and student subject pools, respectively. MTurk workers received the bonus through Amazon Mechanical Turk and students are emailed Amazon.com gift card. This payout scheme is consistent with the literature on narrow bracketing, because subjects only receive a payout from one of the three questions. It is in the subjects' best interests to optimize their choice at the individual question level.

Our experiment was designed to be short, consisting of only the three main questions so that subjects would be attentive throughout. We included a series of attention checks in order to alert the subjects to pay attention. Immediately following the consent form, subjects were asked an attention check question, and if subjects failed this attention check twice, they were excluded from our sample entirely.[6] Subjects were also asked a question regarding the payment scheme in order to verify that the subjects both understood and could compute their potential payoffs. If subjects failed this question, they were warned and could not proceed until it was correctly completed. We record the number of times a subject fails this question.

See Appendix B.1 for additional experimental materials.

2.3 Data Summary

Recall that the choice sets presented to subjects were a random draw from choice sets faced by field households in the BMOT data. Tables 2.1 and 2.2 describe the realized choice sets seen by the experimental subjects. Table 2.1 presents summary statistics for the realized premiums of the $500 deductible in the three insurance contexts.[7] The premiums of these insurance plans are quite large for home insurance and are lower for the auto collision and comprehensive insurance questions. Table 2.2 present statistics on the realized claim rate distribution presented to the

[6]This initial attention check question was pulled from Example 1 of Qualtrics's "4 Ways to Ensure Valid Responses for your Online Survey" (Smith, 2013).

[7]We only display the summary statistics for the price for the $500 deductible because the premium (pricing) structure for all deductibles in a coverage are linearly related. This means that the price to switch between deductibles in a set increases as the premium of the $500 deductible P_{500} increases. This pricing structure is explained in further detail in BMOT p.1505.

experimental subjects.[8] The observed claim rates in the data are similar to the estimated claim rate distributions presented in BMOT.[9]

Table 2.1: All Subjects: Summary of Premium Menus

	Mean	SD	1st percentile	99th percentile
Auto comprehensive premium for $500 deductible	100	49	25	278
Auto collision premium for $500 deductible	163	80	54	511
Home all perils premium for $500 deductible	566	182	198	1056
Cost of decreasing deductible from $500 to $250:				
Auto collision	49	25	15	156
Auto comprehensive	26	13	6	74
Home all perils	47	16	10	98
Savings from increasing deductible from $500 to $1,000:				
Auto collision	37	80	12	117
Auto comprehensive	20	49	5	55
Home all perils	62	182	15	130

Note: Amounts in dollars. Sample of 1639 subjects.

Table 2.2: All Subjects: Given Claim Probabilities

	Auto collision	Auto comprehensive	Home
Mean	6.7	2.1	7.3
SD	2.0	0.9	2.4

Note: Sample of 1639 subjects. Probabilities are presented as percentages.

Table 2.3 and Table 2.4 present the percentage of the subjects who chose the specific deductibles in the three insurance contexts (auto collision, auto comprehensive, and home insurance) between the subject groups and between treatments respectively. Table 2.5 presents the rank correlations between the subject's choices across the three insurance contexts for all subjects and within each treatment group.

[8]The average claim rate in the Ambiguity treatment is presented here.

[9]Because we replicated the field data, there are few subjects who drew probabilities in the high 10%-12% range. In later sections, we will not be able to say much regarding the probability distortions in this range due to the small sample size.

Table 2.3: Summary of deductible choices between subject groups

	All Subjects			MTurk			Student		
Deductible	Auto coll	Auto comp	Home	Auto coll	Auto comp	Home	Auto coll	Auto comp	Home
$50		11.3			12.8			9.3	
$100	16.4	13.6	11.2	17.5	12.3	12.6	14.8	15.4	9.1
$200	13.8	15.9		13.8	15.1		13.8	17.0	
$250	33.9	14.5	43.0	29.6	14.9	36.9	39.8	13.9	51.4
$500	21.5	19.4	25.9	22.5	16.9	27.3	20.2	22.8	23.9
$1000	14.5	25.3	20.0	16.6	27.9	23.2	11.5	21.6	15.5
N		1639			950			689	

Note: Values are percent of subjects.

Table 2.4: Summary of deductible choices between three treatments

	Abstract			Ambiguity			Insurance		
Deductible	Auto Comp	Auto Coll	Home	Auto Comp	Auto Coll	Home	Auto Comp	Auto Coll	Home
$50	8.0			12.3			13.7		
$100	11.1	12.4	8.3	17.1	18.9	12.7	12.6	17.6	12.4
$200	13.9	12.4		19.3	13.4		14.4	15.5	
$250	13.0	31.0	38.2	16.2	39.7	45.6	14.4	30.9	45.1
$500	23.7	27.6	26.9	17.8	18.8	29.6	16.7	18.3	21.2
$1000	30.2	16.5	26.5	17.3	9.2	12.1	28.2	17.6	21.2
N		556			544			539	

Note: Values are percent of subjects.

Table 2.5: Rank correlations between treatments

	All Subjects		Insurance		Ambiguity		Abstract	
	Comp	Coll	Comp	Coll	Comp	Coll	Comp	Coll
Coll	0.530		0.564		0.488		0.521	
Home	0.591	0.544	0.655	0.584	0.515	0.482	0.581	0.542
N	1639		539		544		556	

Note: Spearman rank correlation of deductible choices between all three questions. All correlations are significant at less than the 0.1% level.

Across the different insurance contexts, our subjects choose the whole spectrum of available deductibles. Compared to the choices made in the field data, there is more heterogeneity in the choices made in the laboratory setting. Higher and lower deductibles are chosen more often. For example, in the "All Subjects" panel of Table 2.3, 16.4%, 11.3%, and 11.2% of our 1639 subjects chose the smallest deductible in auto collision, auto comprehensive, and home insurance compared to

the 1.0%, 5.2%, and 0.1% of people who chose them in BMOT (Table 2). Similarly 14.5%, 25.3% and 20.0% chose the $1000 deductible in our sample for the three insurance categories while only 6.7%, 3.6% and 15.9% chose them in the field data.

While the experimental data has a wider dispersion of choices than the field, the within-subject choices were moderately consistent. The Spearman rank correlations of the deductible choices between insurance context pairs are all approximately 0.5 or higher. If a subject chooses a low deductible in one question, they are likely to choose low deductibles in all questions. The correlation between insurance choices and overall heterogeneity in choices suggest that there may be substantial between-subject heterogeneity.

These results preview differences in the estimation results on the choices our experimental subjects make compared to households in the empirical data. These differences may be attributed to the input of other people such as the insurance agent or family members, or it could be a symptom of lower stakes, salience, inattention, and random choice (though we took measures to prevent this) in the experimental subjects.

Table 2.3 and Table 2.4 also highlight small differences in choices between the subject groups and between treatments. Subjects in the Ambiguity treatment made more conservative deductible choices and were less likely to choose the highest $1000 deductible in all three contexts compared to the other two treatments. These differences will be further addressed in the experimental and risk preference results sections Section 2.4 and Section 2.5.2.

Refer to the Appendix Tables B.1 and B.2 for the statistics on within subject group treatments.

2.4 Results from a reduced-form analysis

We present the results of our experiment from two perspectives. In this section, we first discuss the effect that the treatments and menu variation had on the deductible-premium choices in each insurance context. Then, we build a model of our subjects' underlying risk preferences and analyze how the experimental treatments impact their estimated risk preferences in Sections 2.5, 2.5.1, and 2.5.2.

We conduct an ordered-probit analysis of the effect of the experimental treatments on the deductible choices in each insurance context. For each insurance context $j \in$\{Home (H), Collision (L), Comprehensive(M)\}, the dependent variable is the probability that subjects choose deductible d_j from the set of deductible-premium menu pairs $\mathcal{D}_j$. Among the deductibles-premium choices in each of the three insurance contexts, there are 4-6 choices that have an ordinal rank from most conservative (lower deductible amount, higher price) to more risky (higher deductible, lower price).

The ordered probit model that we run for the home insurance context (H) is below. The models for the other two insurance contexts are similar. Assume

$$\begin{aligned} d_H^* =& \mathbf{X'}_H \boldsymbol{\beta}_H + \epsilon_H \\ d_H =& \begin{cases} 100, \text{ if } y_H^* \leq \kappa_{H1} \\ 250, \text{ if } \kappa_{H1} < y_H^* \leq \kappa_{H2} \\ 500, \text{ if } \kappa_{H2} < y_H^* \leq \kappa_{H3} \\ 1000, \text{ if } \kappa_{H3} \leq y_H^* \end{cases} \end{aligned} \tag{2.1}$$

where $\mathbf{X}_H$ are the independent variables that pertain to this specific insurance context. These include the price for the \$500 deductible plan, the average claim rate that the subjects were given, the treatment group, and the subject pool. ϵ_H is

the error term for the regression of home insurance questions and is assumed to be normally distributed. Because we maintain the assumption of narrow bracketing, these regressions are run separately for each of the three contexts. We recover coefficients $\boldsymbol{\beta}_j$ for each insurance context j. We do not analyze the cross context effects.

The results of the ordered probit regression are presented in Table 2.6. They indicate that subjects in the Ambiguity and Abstract treatment groups were significantly more likely to choose lower deductibles (less risky) than in the base Insurance treatment. The marginal effects for the treatment effects and other coefficients are presented in Figure 2.2 and Figure 2.3 respectively.

Compared to the Abstract treatment, the Insurance treatment contained language that is much closer to insurance choices in the field, which subjects may have experienced before. The fact that choices made in the Insurance and Abstract treatment differed significantly highlights the fact that context matters.

Subjects in the Ambiguity treatment were 5-10% more likely to choose the \$1000 deductible than subjects in the insurance treatment. The coefficient on the Ambiguity treatment was also lower than the Abstract treatment coefficient at a 1% significance level in both auto contexts and a 5% level in the home insurance context.[10] Because the Ambiguity treatment group received a range of claim rate probabilities, the choice of lower deductibles could indicate that subjects are ambiguity averse. It is possible that given a range of claim rates, subjects could have perceived a higher subjective claim rate than the average. We discuss the ambiguity result further and test whether their underlying preferences are consistent with maxmin utility in Section 2.5.2.

[10] A one-sided chi-squared tests was used here. With a 2-sided test, the coefficient is different at 1% for comprehensive, 5% for collision, and 10% for home.

Table 2.6: Impact of treatment and subject groups on deductible choice

	Comp	Coll	Home
Price of \$500 Deductible (per \$100)	0.20***	0.22***	0.049***
	(0.057)	(0.035)	(0.015)
Claim rate (per %)	-0.13***	-0.053***	-0.044***
	(0.030)	(0.014)	(0.011)
Ambiguity treatment	-0.37***	-0.32***	-0.36***
	(0.064)	(0.064)	(0.066)
Abstract treatment	-0.17**	-0.17**	-0.24***
	(0.064)	(0.064)	(0.065)
Student	-0.024	-0.060	-0.15**
	(0.053)	(0.053)	(0.054)
κ_{J1}	-1.48***	-1.18***	-1.55***
	(0.092)	(0.11)	(0.13)
κ_{J2}	-0.94***	-0.71***	-0.21
	(0.089)	(0.11)	(0.13)
κ_{J3}	-0.49***	0.18	0.55***
	(0.088)	(0.10)	(0.13)
κ_{J4}	-0.11	0.90***	
	(0.088)	(0.11)	
κ_{J5}	0.43***		
	(0.088)		
N	1639	1639	1639

Standard errors in parentheses

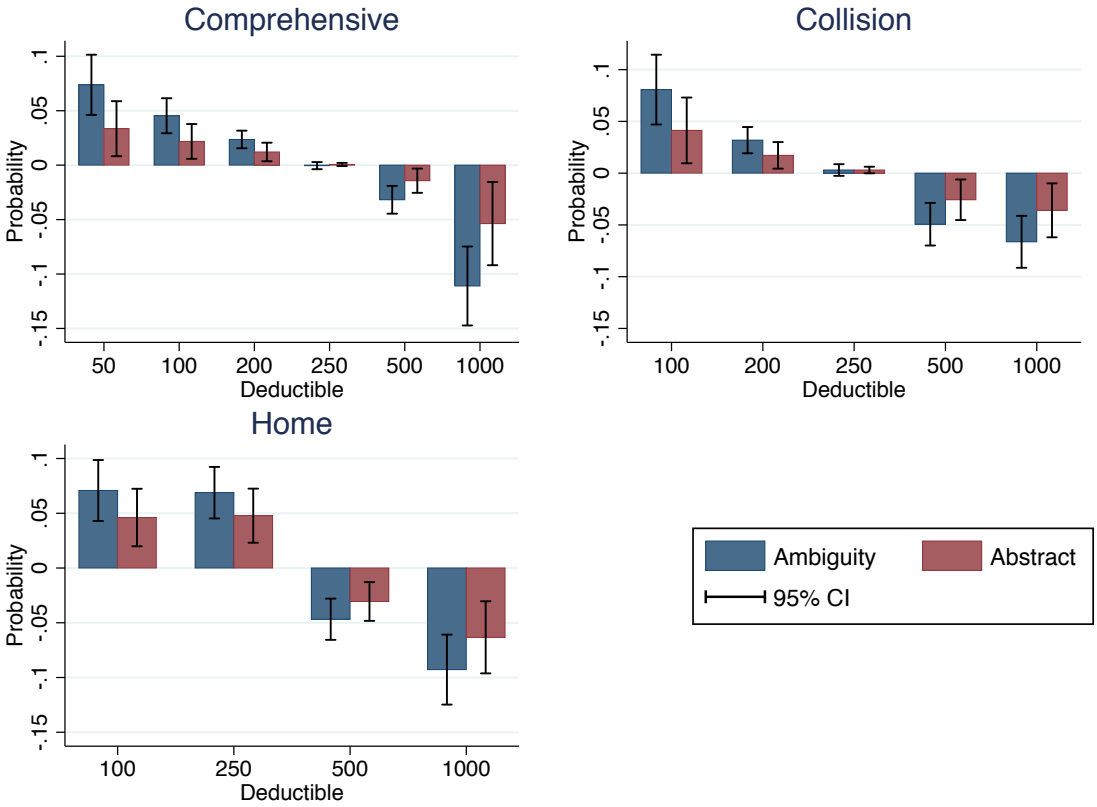
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The deductible prices, the claim rates, and the estimated coefficients are specific to each insurance context: home, auto collision, or auto comprehensive insurance.

In addition, our analysis controls for other sources of variation between subjects' choice sets that come from using actual data. Figure 2.3 presents the marginal effects predicted at the means of the \$500 priced deductible, the claim rate for each respective insurance domain, and being a student. Due to the premium structure of the field data, increases in p_{500} results in both more expensive deductibles and makes it more expensive to switch to a lower deductible, e.g. $p_{500} - p_{1000}$ is also increasing in p_{500}.[11] When the premium for the \$500 deductible is higher, subjects were significantly more likely to buy less insurance and choose higher deductibles

[11] See BMOT for more details on the pricing structure.

Figure 2.2: Marginal effects of treatment groups on deductible choices

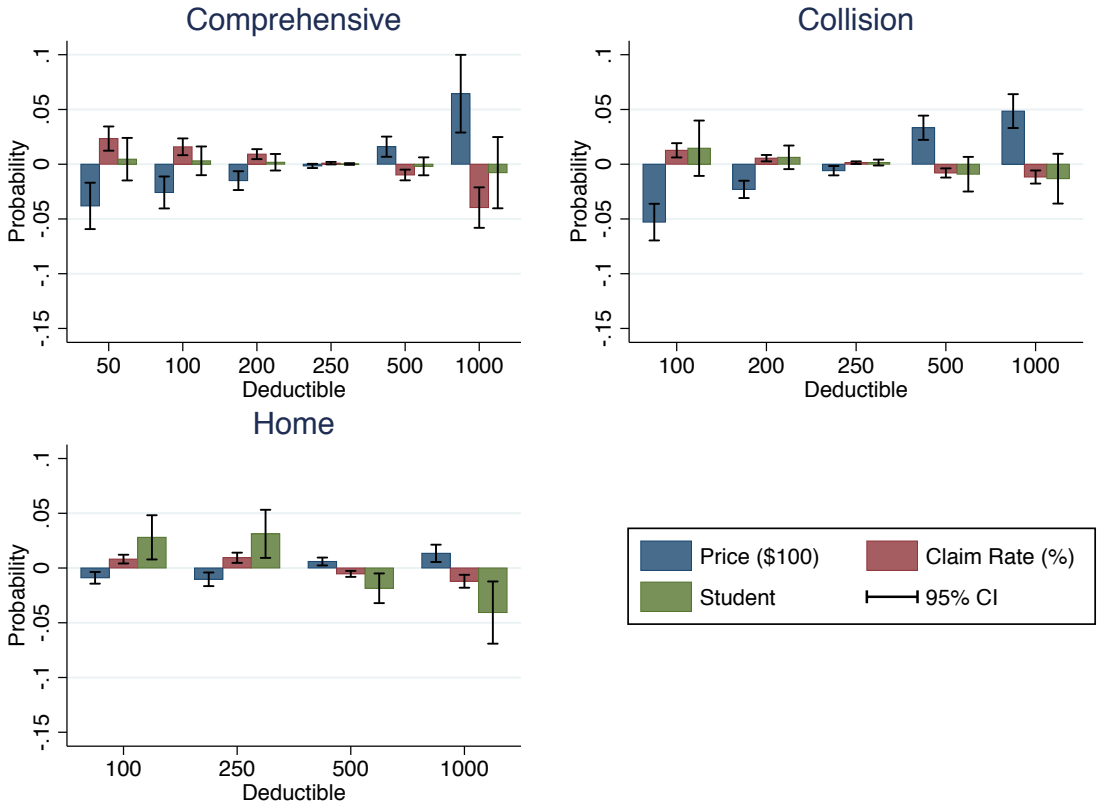


Note: Sample size 1639. The image presents the marginal effect on the probability of choosing one of the deductible choices when there is a change of 0 to 1 of the indicator of being in the treatment. The marginal effects are calculated after the ordered probit regression of Equation 2.1 for home and the other insurance domains. The marginal effects are assumed to be at means for other variables.

in all three contexts. This indicates that for a \$100 increase in the price of the \$500 deductible in comprehensive (about 2 standard deviations), the subjects are 6.4% more likely to choose a \$1,000 deductible. In auto collision and home all-perils insurance, \$100 is about 1.25 and 0.55 standard deviations of the \$500 deductible respectively. While a \$100 dollar increase in the \$500 deductible results in a smaller predicted changes in the probability of deductible choices than comprehensive, the change for all deductibles are statistically significant.

Similarly, as the risk or probability of an accident increases, subjects were sig-

Figure 2.3: Marginal effects of price, risk, and subject group on deductible choices



Note: Sample size 1639. The image presents the marginal effect on the probability of choosing deductibles due to a change in $100 on the price of the $500 dollar deductible, a percentage change in the claim rate, and the marginal effect for being a student in the comprehensive, collision, and home insurance domains. The marginal effects are calculated after the ordered probit regression of Equation 2.1 for home and the other insurance domains and are calculated at the means.

nificantly more likely to buy more insurance. A one-percentage point change is approximately a one standard deviation change in the comprehensive claim rate, a 0.5 standard deviation change in the collision claim rate and a 0.4 standard deviation change in the home claim rate. A one-percentage point increase in the claim rate decreases the marginal probability of choosing a $1,000 deductible by approximately 4%. Only the change in the probability of choosing the comprehensive deductible of $250 is not significant.

Students were more likely to choose lower deductibles than MTurk subjects,

but both the estimated coefficient and marginal effect are only significant in the context of home insurance choices. Student subjects were also not substantially likely to respond differently to our experimental treatments, thus the discussion of how their responses differed from MTurk subjects is included in Appendix B.3. Further, the effect of student demographics on their choices are also discussed in Appendix B.3.

2.5 Estimating Structural Models

Beyond understanding how choices differ due to our experimental treatments, we are interested in comparing how the underlying risk preferences of our experimental subjects differ from field subjects. Broadly, we follow the same modeling assumptions of BMOT, but the functional forms and distributional assumptions differ to fit the experimental data.

Subjects' decisions are modeled as choices over deductible-premium menus over risk. Our subjects are presented with the probability (or probability range) of a loss μ and a menu of premium-deductible pairs $\{(p_d, d) : d \in \mathcal{D}\}$. If subjects choose to pay premium p_d for deductible d, then they have chosen the lottery

$$L_d = (-p_d, 1-\mu; -p_d - d, \mu).$$

We assume that subjects choose the deductible-premium pairs to maximize their utility from the lottery $L_d = (-p_d, 1-\mu; -p_d - d, \mu)$. The utility from the

lottery is

$$U(L_d) = (1 - \Omega(\mu))u(w - p_d) + \Omega(\mu)u(w - p_d - d) \tag{2.2}$$

where $\Omega(\mu)$ is a probability distortion function that nests nonlinear probability weighting, Kőzesgi-Rabin loss aversion, and Gul disappointment aversion.[12] Following Cohen and Einav (2007), Barseghyan et al. (2011), and BMOT, the utility u is modeled parametrically with a normalized second-order Taylor approximation. Thus, Equation 2.2 becomes

$$U(L_d) = -[p_d + \Omega(\mu)d] - \frac{r}{2}[(1 - \Omega(\mu))(p_d)^2 + \Omega(\mu)(p_d + d)^2] \tag{2.3}$$

where $r = \frac{-u''(w)}{u'(w)}$ is a coefficient of absolute risk aversion (CARA).

Using this model, we employ an estimation strategy allowing for homogeneous risk preferences and for unobserved heterogeneous risk preferences of our experimental subjects. We estimate parameters for the coefficient of absolute risk aversion r, the probability distortion function $\Omega(\mu)$, and scale σ.

In order to account for households choosing different deductibles and making "inconsistent" choices, we use a random utility model (McFadden, 1974, 1981) with additively separable choice noise. Then the utility from choosing deductible $d \in \mathcal{D}$ is given by

$$\mathcal{U}(d) = U(L_d) + \varepsilon_d \tag{2.4}$$

where ε_d is i.i.d. with a type 1 extreme value distribution and scale parameter σ. Subjects choose deductible $d \in \mathcal{D}$ to maximize their utility $\mathcal{U}(d)$. As in BMOT,

[12] See BMOT for identification proofs.

the probability that a household choses deductible d in one insurance context is

$$Pr(d) = Pr(U(L_d) + \varepsilon_d > U(L_{d'}) + \varepsilon_{d'} \forall d' \neq d) = \frac{exp(U(L_d)/\sigma)}{\sum_{d' \in D} exp(U(L_{d'})/\sigma)}. \quad (2.5)$$

2.5.1 Analysis with homogeneous risk preferences

We first estimate models of homogeneous risk preferences, where all subjects are assumed to have the same risk aversion parameter r, probability distortion function $\Omega(\mu)$, and scale parameter σ, which is the same across all contexts to reduce the dimensionality of the estimation problem.[13] We make two functional form assumptions to estimating $\Omega(\mu)$ that are unique to our paper. For Model 1a, because the claim rates presented to the experimental subjects are discrete, we can estimate a nonparametric probability distortion function at every claim rate value.[14] In Model 1b, we reduce the dimensionality of the estimation problem and assert parametric assumptions on $\Omega(\mu)$ with a polynomial form.[15]

The $\Omega(\mu)$ function is then modeled in a nonparametric form in Model 1a

$$\Omega(\mu) = \begin{cases} \omega_{0.01}, & \text{if } \mu = 0.01 \\ ..., & \\ \omega_{0.12}, & \text{if } \mu = 0.12 \end{cases} \quad (2.6)$$

and with the polynomial form in Model 1b

$$\Omega(\mu) = a + b\mu + c\mu^2 \quad (2.7)$$

[13]This single σ assumption differs from BMOT, which estimated a different scale parameter σ_j for each insurance context j.

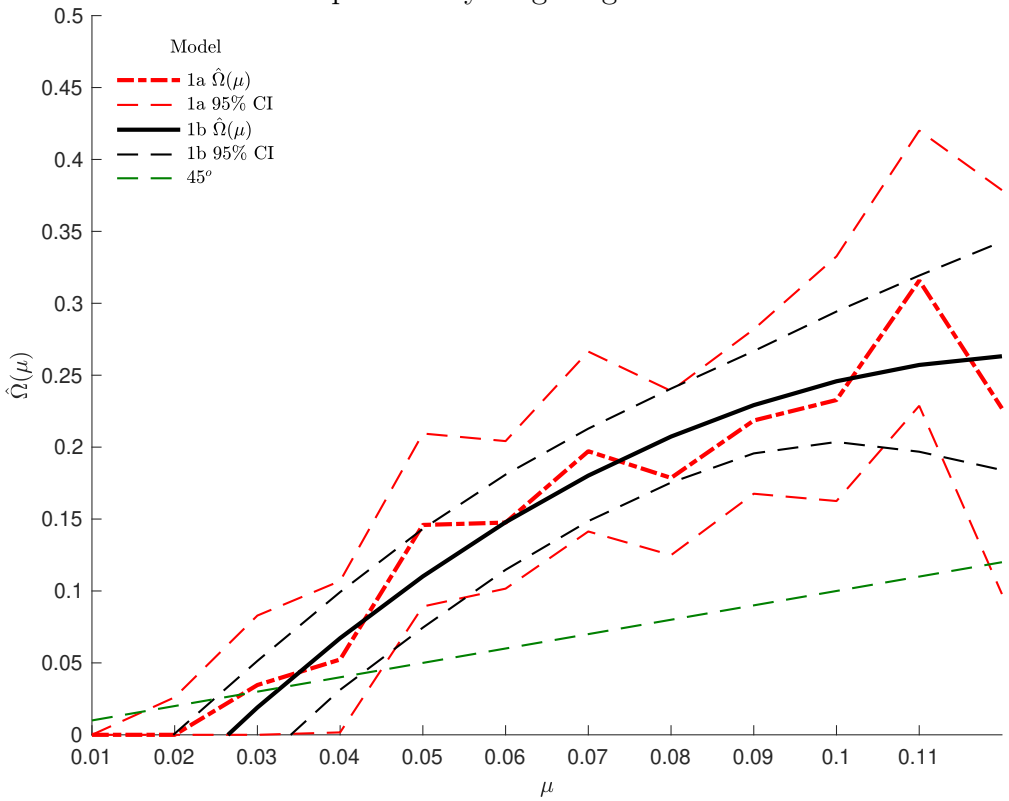
[14]This form differs from BMOT who estimated $\Omega(\mu)$ with three model variations: a quadratic expansion of $ln(\Omega(\mu))$, a quadratic expansion of $\Omega(\mu)$ and an 11-point cubic spline on the interval (0, 0.20).

[15]We estimate a Chebyshev polynomial expansion of $\Omega(\mu)$ and selected a quadratic on the basis of the Bayesian information criterion (BIC).

with the restriction that $\Omega(\mu) \in [0,1] \forall \mu$.

These models are estimated using maximum likelihood to recover Model 1a parameters $\theta \equiv (\omega_{0.01}, ..., \omega_{0.12}, r, \sigma)$ and Model 1b parameters $\hat{\theta} \equiv (a, b, c, r, \sigma)$ for each insurance context j and experimental treatment $\tau \in \{\text{Insurance, Ambiguity, Abstract}\}$. Figure 2.4 and Table 2.7 present the estimated $\hat{\Omega}(\mu)$, $\hat{r}$, and $\hat{\sigma}$ with 95% confidence interval bounds from the Insurance treatment. The results from the Abstract and Ambiguity treatment are similar and are presented in Appendix B.5.

Figure 2.4: Model 1a and 1b probability weighting functions in Insurance treatment



Note: This figure displays the estimated $\hat{\Omega}(\mu)$ from Model 1a and Model 1b across the 539 subjects in the Insurance treatment. The 95% confidence intervals from these estimated values are constructed with 1,000 bootstraps with a subsample of 80 percent of 539 with replacement. The bootstrap was conducted with a subsample due to estimated r, $\omega_{0.01}$ and $\omega_{0.02}$ being close to or at the lower bound of 0.

Our results are summarized in three key findings: the risk preference parameter

Table 2.7: Model 1a and 1b for Insurance

	Model 1a			Model 1b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	89.5	73.0	108.9	102.6	81.5	131.6
r	2e-13	2e-13	2e-13	4e-14	4e-14	4e-14

Note: 539 households in the Insurance treatment. See additional notes in Figure 2.4 on how estimates are constructed.

r is negligible; on average our subjects engage in substantial probability weighting to explain their premium-deductible choices; and there is substantial heterogeneity that is not explained by the model. In the Insurance treatment, the risk aversion parameter estimate is computationally zero, while the the majority of the estimated probability distortion function $\hat{\Omega}(\mu)$ lies above the 45 degree line. These estimates of $\hat{\Omega}(\mu)$ indicate our subjects engage in large and significant amounts of probability *over*weighting. At a 10% percent probability of a loss, the subjects behaved as if the probability was 23%. At very low probabilities, the estimate of $\hat{\Omega}(\mu)$ in both models indicate probability *under*weighting.

The estimated $\hat{\sigma}$ ranges from 89.5 to 102.6 in Model 1a and 1b, respectively. This $\hat{\sigma}$ estimate is substantially higher than those estimated from a similar model on the field data in BMOT Model 1 (a, b, and c)[16] and from subsequent models in this paper. The large σ indicates that there is a substantial amount of noise that is not yet explained by the homogenous model parameters.

While the homogeneous model indicates that subjects may be underweighting low probabilities, this finding is not robust to further models with heterogeneous risk preferences discussed below.

[16]The BMOT Model 1a/b/c differ from the homogenous model only in the functional forms assumed for $\Omega(\mu)$ and the fact that BMOT estimated a σ_j for each insurance context j. Given the similarities in the models, the difference in the σ estimates between our paper and BMOT are driven primarily by the underlying data differences between the field and experimental setting.

2.5.2 Analysis with heterogeneous risk preferences

We know from even the raw data in Section 2.3 that there is a substantial amount of heterogeneity in the dispersion of premium-deductible choices. This indicates that there may also be substantial heterogeneity in underlying preferences that the homogeneous preference model cannot capture.

We present results from six additional models that allow for individual unobserved heterogeneity in risk preferences. Model 2a and 2b allow for the most heterogeneity with both individually specific curvature of the utility function r_i and a scaled probability distortion function $\Omega_i(\mu)$ for each subject i. Model 3a and 3b are a simplification of Model 2a and 2b, respectively, and assume the same homogenous curvature in utility r for all individuals. Model 4a and 4b is further restricted to assume $r = 0$. In these models, we also recover estimates for the noise scale parameter σ that is the same across all individuals and insurance contexts.[17]

In the most general case of Model 2a and 2b, assume

$$r_i = r\xi_{r,i} \quad \text{and} \quad \Omega_i(\mu) = \Omega(\mu)\xi_{\Omega,i}$$

where

$$\begin{pmatrix} \xi_{r,i} \\ \xi_{\Omega,i} \end{pmatrix} \overset{iid}{\sim} Logit - normal \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Phi \right), \quad \text{with} \quad \Phi \equiv \begin{bmatrix} \Phi_r & \Phi_{r,\Omega} \\ \Phi_{r,\Omega} & \Phi_\Omega \end{bmatrix}$$

and $\Omega_i(\mu)$ is bound between 0 and 1, and $\Omega(\mu)$ is modeled in a nonparametric and polynomial form from Equation 2.6 and Equation 2.7 in Model 2a and 2b

[17]Notice Model 1 a and b are also nested in Model 2 a and b by assuming both r_i and $\Omega_i(\mu)$ are constant between all individuals.

respectively. We assume that choice noise is independent of any observed or unobserved heterogeneity in preferences and also independent between the experimental contexts.

We choose to model heterogeneity with a logit-normal error distribution, so $\xi_{r,i}$ and $\xi_{\Omega,i}$ are naturally constrained to $[0, 1]$, which also constrains Ω_i to $[0, 1]$ and r_i to [0, r].[18] Further, while r_i is not theoretically bound from above, it is bound in practice. Due to the nature of the choices, where the least risky deductible is either \$50 or \$100, what we can estimate for r_i is either in a bounded range or not identified. While the underlying model here is similar to that used in BMOT Model 3 and 4, they had assumed a log-normal error distribution for r_i and $\Omega_i(\mu)$. We prefer the logit-normal distribution in this case because the log-normal distribution is unbounded and not well behaved when r approaches 0 as it does in the experimental data. Another important benefit of using the logit-normal error distribution for $\xi_{\Omega,i}$ is that this functional form can allow for a single centered distribution or an approximation of a bimodal distribution of two types at the extremes, which allows our estimated probability distortion function to be comparable to the type estimation from Bruhin et al. (2010).

We estimate these models using maximum likelihood methods and recover Model 2a parameters $(\omega_{0.01}, ..., \omega_{0.12}, r, \sigma, \Phi_r, \Phi_\Omega, \Phi_{r,\Omega})$, Model 2b parameters $(a, b, c, r, \sigma, \Phi_r, \Phi_\Omega, \Phi_{r,\Omega})$, Model 3a parameters $(\omega_{0.01}, ..., \omega_{0.12}, r, \sigma, \Phi_\Omega)$, Model 3b parameters $(a, b, c, r, \sigma, \Phi_\Omega)$, Model 4a parameters $(\omega_{0.01}, ..., \omega_{0.12}, \sigma, \Phi_\Omega)$, and Model 4b parameters $(a, b, c, \sigma, \Phi_\Omega)$ for each experimental treatment τ. We approximate the likelihood function using a symmetric Gaussian-Hermite quadrature rule with 25 nodes in each dimension.[19] We use the estimates to assign fitted values of

[18] Ω is bound by [0,1]. Then, if $\xi_{\Omega,i}$ is also bound by 0 to 1, then the product $\Omega_i(\mu)$ is bound as well.

[19] For Model 2, because there are two dimensions of heterogeneity, in practice, after constructing

r_i and $\Omega_i(\mu)$ to each subject i.[20]

Table 2.8 summarizes the estimates for Model 2a, 2b, 3a, 3b, 4a, and 4b using only the responses from individuals randomized into the Insurance treatment. Figure 2.5 depicts the median estimate of $\hat{\Omega}_i(\mu)$ in the Insurance treatment for these models. Figure 2.6 depicts the scope of the heterogeneity in $\hat{\Omega}_i(\mu)$ using these six models.[21]

Table 2.8: Model 2, 3, and 4 for Insurance

	Model 2a			Model 2b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	33.66	28.13	39.17	34.66	29.27	41.46
r_i 10^{th} percentile	5.3e-33			8.5e-32		
$median(r_i)$	1.6e-10	1.6e-10	1.6e-10	1.4e-08	1.3e-08	1.5e-08
r_i 90^{th} percentile	3.2e-10			2.8e-08		
Φ_Ω	4.28	3.48	5.44	4.29	3.50	5.49
$\Phi_{\Omega,r}$	-156.77	-199.34	-114.10	-162.38	-210.26	-72.09
Implied corr$(\xi_{r,i}, \xi_{\Omega,i})$	-0.89	-0.93	-0.71	-0.90	-0.94	-0.38

	Model 3a			Model 3b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	33.79	28.11	40.52	34.85	29.33	43.11
r	1.6e-12	1.6e-12	1.6e-12	2.4e-11	2.4e-11	2.4e-11
Φ_Ω	4.11	3.40	38.30	4.16	3.40	37.76

	Model 4a			Model 4b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	33.79	28.49	39.99	34.85	29.46	42.08
Φ_Ω	4.11	3.41	7.51	4.16	3.46	35.58

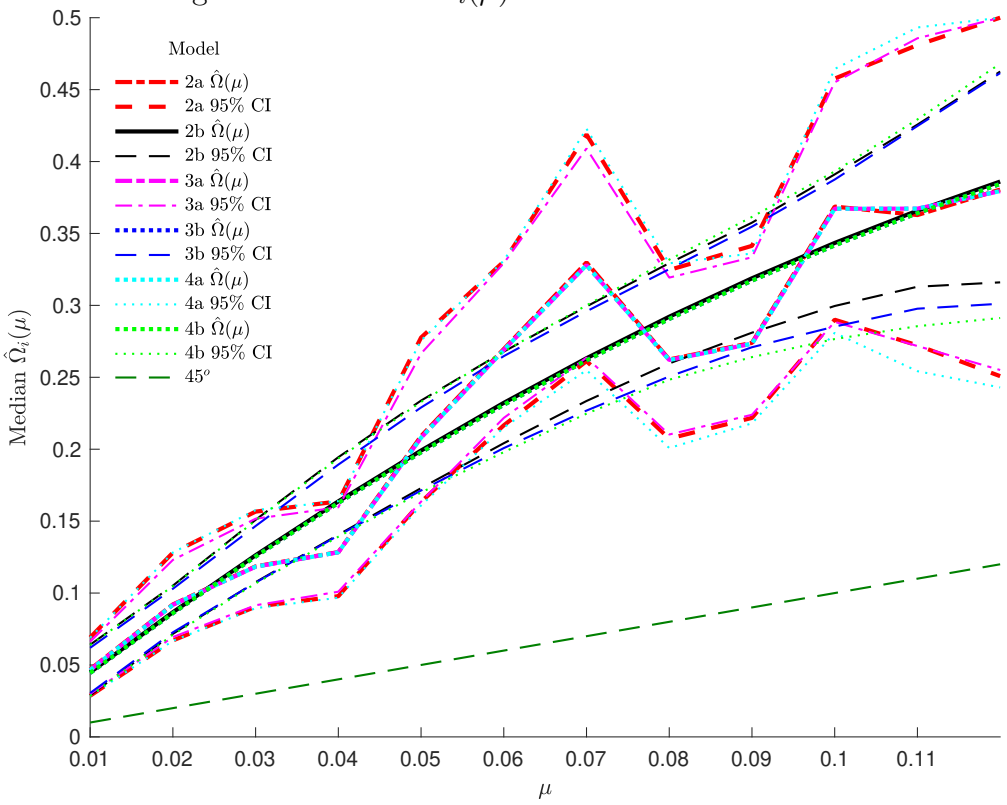
Note: 539 households in the Insurance treatment. The reported value is median $\hat{r}_i$ instead of $\hat{r}$, because r is the maximal value that the risk aversion parameter can take for $\xi_{r,i} = 1$. See additional notes in Figure 2.5 on how estimates are constructed.

$25^2 = 625$ pairs, we drop those whose associated weight is less than 10^{-6} and reweigh. This give us 177 weights.

[20]We choose quadrature and MLE methods to approximate the integral in the likelihood function as opposed to using the stochastic Markov-chain Monte Carlo (MCMC) estimation technique used in BMOT and Cohen and Einav (2007). This estimation method allows for faster computation and can be easily determined to converge, unlike MCMC.

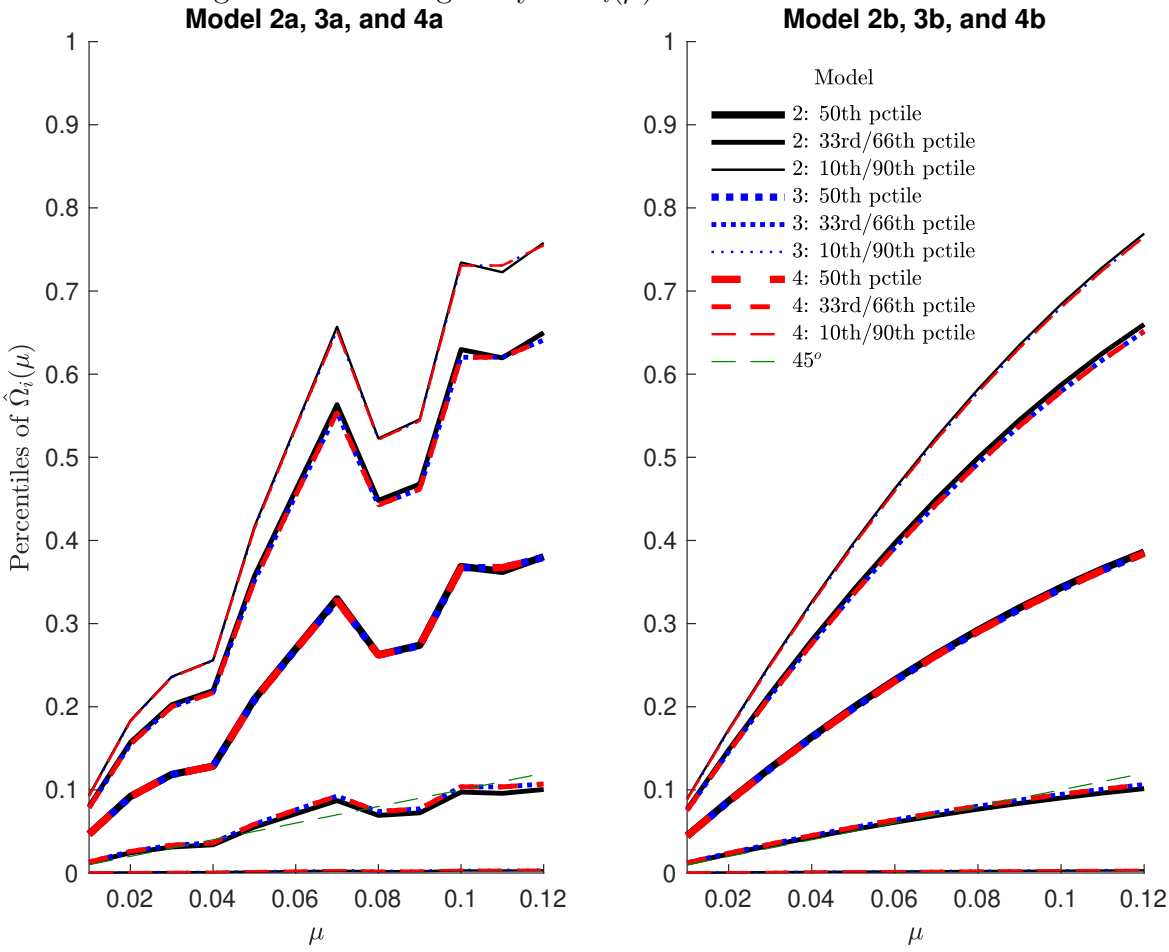
[21]We report median $\hat{\Omega}_i(\mu)$ rather than $\hat{\Omega}(\mu)$, since in Model 2a and 2b model form $\Omega_i(\mu) = \Omega(\mu)$ only if $\xi_{\Omega,i} = 1$ the maximal value given the logit-normal distribution.

Figure 2.5: Median $\Omega_i(\mu)$ in Insurance treatment



Note: This figure displays the estimated median $\Omega_i(\mu)$ from Model 2 a/b, 4 a/b, and 3 a/b across the 539 subjects in the Insurance treatment. Because the probability weighting function and r has a lower bound at 0, the 95% confidence intervals from the estimated values of Model 2a, 2b, 3a, and 3b are constructed with 1,000 bootstraps with random subsamples that are approximately 80% the size of the Insurance treatment subjects (431). Subsampling is used in these cases because the estimate of r approaches a lower bound in these models. The 95% confidence intervals for Model 4a and 4b are constructed from 1,000 bootstraps of N=539 drawn from the original sample with replacement.

Figure 2.6: Heterogeneity of $\Omega_i(\mu)$ in Insurance treatment



Note: Two panels that show the quantiles of the estimated $\Omega_i(\mu)$ given the estimated noise $\Phi_{\Omega,i}$ using subjects from the Insurance treatment. The quantiles of $\Omega(\mu)$ from Model 2a, 3a, and 4a, are displayed on the left, while the quantiles from Model 2b, 3b, and 4b are displayed in the right panel. See additional notes in Figure 2.5 on how estimates are constructed.

In the Insurance treatment of Model 2a and 2b, we find the risk aversion parameters $\hat{r}_i$ to be effectively zero even in the 90th highest percentile of subjects. The estimates of $\hat{r}_i$ in a narrow range close to zero motivates Model 3a, 3b, 4a, and 4b in which r is first assumed to be homogenous between subjects in the first two models and then assumed to be zero in the later two.

Given negligible $\hat{r}_i$ in Model 2, 3, and 4, we find the estimates of the median $\hat{\Omega}_i(\mu)$ are the same for the respective functional form of $\Omega(\mu)$, though the confidence interval bands differ slightly. We again find substantial probability overweighting. The median $\hat{\Omega}_i(\mu)$ under all models is significantly above the 45° line at all observed probabilities μ, even at very low probabilities. For $\mu = 0.01$, the median probability weighing is $\hat{\Omega}_i(0.01) = 0.05$, and at $\mu = 0.12$, the median is approximately $\hat{\Omega}_i(0.12) = 0.38$. The estimates of the median $\hat{\Omega}(\mu)$ are far tighter at low probabilities, because far more of our subjects received those values. The estimate of $\hat{\sigma}$ is in the 30-35 range in Model 2a/b, 3a/b, and 4a/b and is approximately one third of the $\hat{\sigma}$ estimated in Model 1a and 1b.

The experimental subjects exhibit a tremendous amount of heterogeneity in their risk preferences, but that heterogeneity is mostly a result of heterogeneity in probability distortions and not heterogeneity in the curvature ofutility. Figure 2.6 displays the heterogeneity in $\hat{\Omega}_i(\mu)$. Individuals in the 33rd percentile and below underweighted their probabilities, while those in the 90th percentile behaved as if losses were 70% likely at $\mu = 0.1$ in all models. This heterogeneity can be seen in the raw data where we observed many individuals choosing the \$100 and \$1,000 deductibles. From the rank correlations between the deductible choices, we know that the heterogeneity in choices is largely a result of the between-individual heterogeneity that was built into Model 2, 3, and 4 rather than just within-individual

heterogeneity.

The findings in these models are consistent with the field literature because choices are primarily explained through probability weighting rather than through curvature in the utility function through the risk aversion parameter r. Our estimation indicates that the experimental subjects in the Insurance treatment relied even less on curvature in the utility to explain their decision than field households did in BMOT, which reported $\hat{r} = 0.00064, \hat{r} = 0.00063, \hat{r} = 0.00049$ in their Model 1a, 1b, and 1c, respectively. Comparing the probability distortion functions, the median $\hat{\Omega}_i(\mu)$ estimates indicate less probability weighting of very low probabilities and substantially higher probability weighting at higher μ. BMOT estimated a mean probability distortion function $\tilde{\Omega}(0.01) = 0.07$ and $\tilde{\Omega}(0.1) = 0.15$.[22] While our estimates differ from those estimated in the field and experiments, they are closer to the estimates of probability weighting in prior experimental works such as (Bruhin et al., 2010). Unlike Bruhin et al. (2010), about a third of our subjects display probability underweighting, and this is potentially due to the difference of using certainty equivalent tasks in their experiment and a complicated menu of risky choices in ours.

The estimate of the median $\hat{\Omega}_i(\mu)$ in the heterogeneous models is higher than those from the homogeneous risk preference models (Model 1a and Model 1b) at all μ. Qualitatively, the most apparent difference is that Model 1a and 1b both recovered probability underweighting at low μ, while the median estimate from the heterogeneous preference models do not result in probability underweighting. When considering the underlying amount of heterogeneity in the choices subjects

[22] In BMOT's notation, their estimated mean probability distortion function would be $\hat{\Omega}(\mu)$. Because of the difference in the functional form and error distribution between our paper and BMOT, our $\hat{\Omega}(\mu)$ is the upper bound of the probability distortion function rather than the mean/median.

made and the distribution of $\hat{\Omega}_i(\mu)$, it is reasonable that the estimates from Model 1 differ from Model 2, 4, and 3 and that the later models, which allow for heterogeneous risk preferences, would be a better fit.

Further, the difference in the estimates from the homogeneous versus heterogeneous risk preferences models is in itself a difference between the field and our laboratory setting. The estimates from BMOT's homogeneous and heterogeneous models were comparable. This difference again highlights the heterogeneity of choices and preferences collected from the lab, because the heterogeneity in the distribution of $\Omega(\mu)$ results in different estimates from the homogeneous model and the median heterogeneous model. The field data in BMOT could have been generated by households with more similarities than individuals in the experimental setting.

Another key finding from our models of heterogeneous risk preferences is that, unlike the field, our estimates do not necessarily rule out Kőszegi and Rabin (2007) "rational expectations" loss aversion or Gul (1991) disappointment aversion. If we were to evaluate the Model 2b, 3b, or 4b at $\mu = 0$, our polynomial estimates of the median $\hat{\Omega}_i(0)$ would be approximately zero. In BMOT, their estimate of $\tilde{\Omega}(0)$ is significantly higher than zero in all of their models and approximately 0.06.

Treatment Effects

To evaluate the effect of the three treatments, we focus the discussion on the results of Model 4b. This model allows us to reduce the dimensionality of the estimation problem, while still capturing a reasonable approximation of the $\Omega(\mu)$ probability distortion function. The evidence from Model 2 and 3 that r is negligible for our subjects, allow us to use Model 4b to directly compare the differences in the median

$\hat{\Omega}_i(\mu)$ probability distortions between the treatment groups without concern for large changes in r_i.

Figure 2.7 displays the median $\hat{\Omega}_i(\mu)$ from the Ambiguity and Abstract treatments from Model 2b, 3b, and 4b, and Table 2.9 displays the corresponding parameter estimates. The results for Model 2a, 3a, and 4a for these treatments are similar and are included in Appendix B.6 along with details about heterogeneity in these models. Similar to the ordered probit evaluation, we initially assume that the subjects in the Ambiguity treatment consider the median value from the set of claims probabilities that are presented to them, i.e. $\mu = median(\boldsymbol{\mu})$.

Table 2.9: Model 2b and 4b for Ambiguity and Abstract

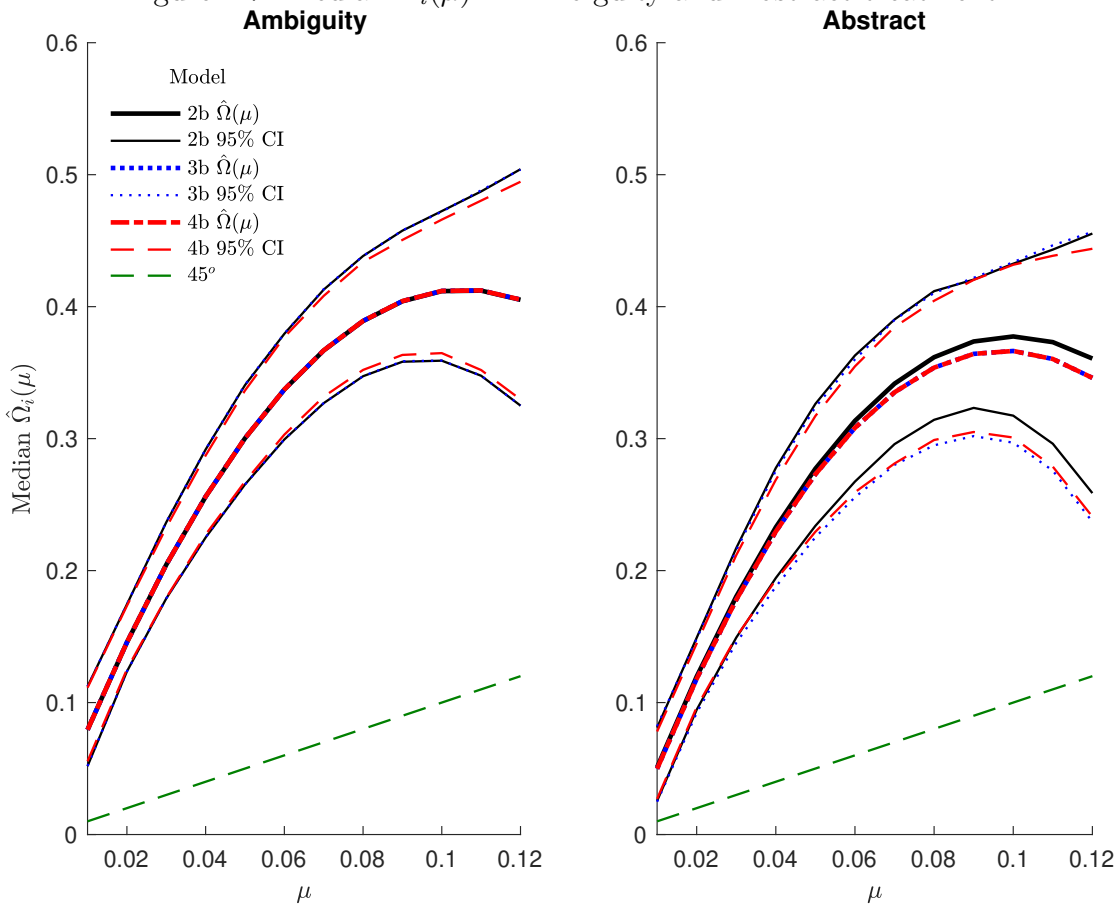
	Ambiguity: Model 2b			Abstract: Model 2b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	39.83	33.50	47.34	45.18	37.87	54.95
r_i 10^{th} percentile	2.2e-22			7.4e-23		
$median(r_i)$	1.0e-08	9.8e-09	1.2e-08	4.6e-15	4.6e-15	4.6e-15
r_i 90^{th} percentile	2.1e-08			9.2e-15		
Φ_Ω	2.61	2.24	3.08	4.99	3.90	16.16
$\Phi_{\Omega,r}$	-64.95	-74.55	-0.00	64.96	0.00	97.34
Implied corr$(\xi_{r,i}, \xi_{\Omega,i})$	-0.99	-0.99	-0.00	0.89	0.00	0.99

	Ambiguity: Model 3b			Abstract: Model 3b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	39.83	33.51	47.34	46.47	38.08	58.48
r	2.4e-11	2.4e-11	2.4e-11	2.7e-10	2.6e-10	2.7e-10
Φ_Ω	2.62	2.24	3.06	5.20	3.78	94.00

	Ambiguity: Model 4b			Abstract: Model 4b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	39.83	34.39	46.19	46.47	38.76	57.82
Φ_Ω	2.62	2.25	3.03	5.20	3.87	78.35

Note: Display select parameter estimates from Model 2b and 4b across the 544 subjects in the Ambiguity treatment and the 556 subjects in the Abstract treatment. The value we report is median $\hat{r}_i$ instead of $\hat{r}$, because r is the maximal value that the risk aversion parameter can take for $\xi_{r,i} = 1$. See additional notes in Figure 2.7 on how estimates are constructed. See Appendix Table B.7 and Appendix Table B.8 for Model 2a, 3a, and 4a results for the Ambiguity and Abstract treatments, respectively.

Figure 2.7: Median $\Omega_i(\mu)$ in Ambiguity and Abstract treatment



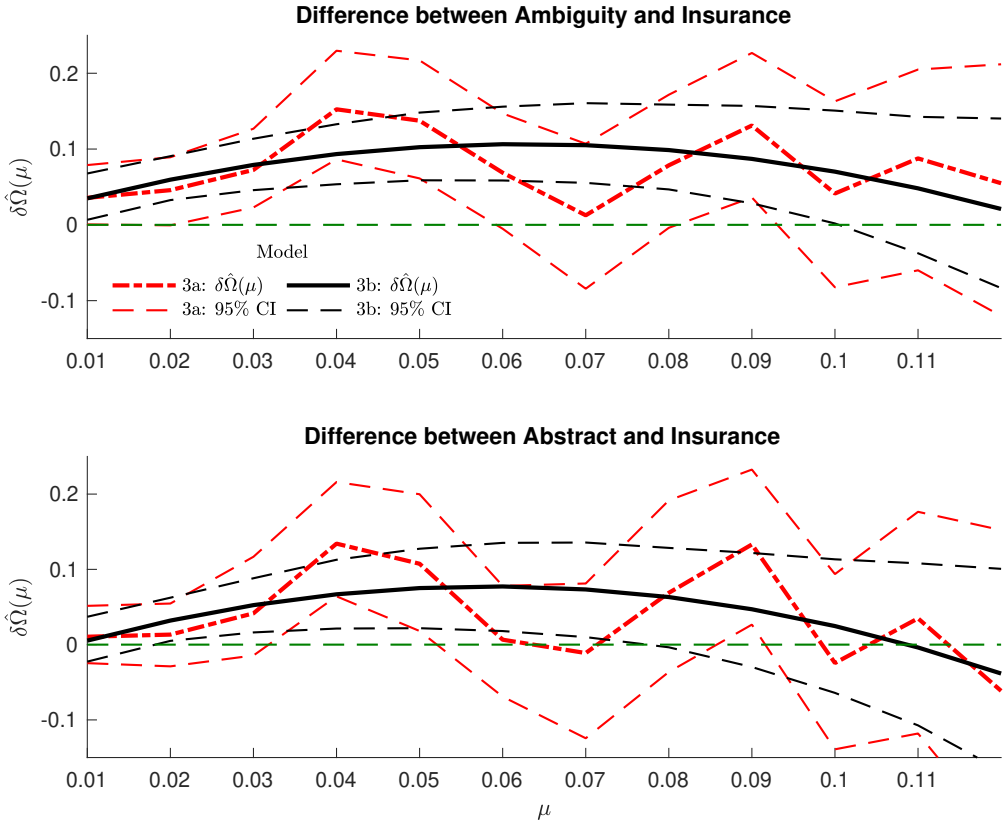
Note: Display the estimated median $\Omega_i(\mu)$ from Model 2b, 3b, and 4b across the 544 subjects in the Ambiguity treatment and the 556 subjects in the Abstract treatment. In the estimation for the Ambiguity model, assume that subjects consider the median value in the set of claim probability values presented to them $\mu = \boldsymbol{\mu}$. The 95% confidence intervals are constructed from 1,000 bootstraps drawn from each sample with replacement. See Appendix Figure B.6 and Figure B.7 for Model 2a, 3a, and 4a results.

Subjects in the Ambiguity and Abstract treatments are similar to those in the Insurance treatment in that the estimated median risk aversion parameter is small and the median probability distortion function exhibits substantial probability overweighting. While the risk aversion parameter is larger in the Ambiguity treatment, it is still only a fraction of the estimates found in the field. There are larger differences in the estimates for the median $\hat{\Omega}_i(\mu)$.

Figure 2.8 depicts how the estimated median probability distortion functions

in the Abstract and Ambiguity treatments differ from the baseline Insurance treatment in Model 4a and 4b. In both models, the treatment effect of the Ambiguity and Abstract treatments compared to the baseline Insurance treatment resulted in higher median estimated probability distortion functions, and the effect is greater in the Ambiguity treatment. In Model 4b, the estimate of the Ambiguity treatment effect is almost 5% at $\mu = 0.01$, peaks at $\mu = 0.06$ with $\hat{\Omega}^{\text{Ambiguity}}(0.06) - \hat{\Omega}^{\text{Insurance}}(0.06) = 0.11$, and is statistically significant at most claim probabilities from 1-10%.

Figure 2.8: Treatment differences with polynomial form and heterogeneous risk preferences



Note: Displays the difference in the Model 4a and 4b estimated median $\Omega_i(\mu)$ from the Ambiguity and Abstract treatments to the Insurance treatment. Let $\delta\hat{\Omega}(\mu)$ denote the difference of $\delta\hat{\Omega}_i(\mu) =$ median $\hat{\Omega}_i(\mu)^{\tau} -$ median $\hat{\Omega}_i(\mu)^{\text{Insurance}}$ where τ indicates the treatment of Ambiguity or Abstract in the top and bottom graphs respectively.

The median difference between the Abstract and Insurance treatments in Model

4b are not as high and are only statistically significant on the range of 2-7% claims probabilities. The Abstract treatment effect is zero at $\mu = 0$, peaks at $\mu = 0.06$ where $\hat{\Omega}^{\text{Abstract}}(0.06) - \hat{\Omega}^{\text{Insurance}}(0.06) = 0.076$. The Abstract treatment effect has a wider confidence interval than the Ambiguity treatment effect indicating that there is less consistency in the choices in the Abstract treatment. These results are consistent with our reduced-form analysis. The estimation results reveal that subjects behaved as if the claim probabilities were higher in the Ambiguity and Abstract treatments than in the Insurance treatments, choosing significantly less risky deductibles as a result.

Modeling ambiguity aversion

We next investigate whether the effects of the Ambiguity treatment are consistent with the predictions of a standard maxmin model of ambiguity aversion (Etner et al., 2012; Gilboa and Schmeidler, 1989).

The estimation in Section 2.5.2 assumes for the Ambiguity treatment that the subjective claim rate is the median of the set of claim rates that subjects received. However ambiguity aversion would predict that, when subjects are unsure of the distribution of a risk, they might be pessimistic and react as if the probability of a loss were higher than the median. Indeed, the estimates from the Ambiguity treatment are consistent such an effect.

We consider a model with a maxmin expected utility that incorporates the pessimism of ambiguity aversion (Etner et al., 2012; Gilboa and Schmeidler, 1989). This model would predict that individuals act according to the worst possible prior in a set of potential outcomes, which in this context means using the highest loss probability in the range that is given. In other words, the maxmin utility assumes

that Ambiguity treatment subjects behave as pessimistically as possible.

We test the maxmin model using the results from Model 4a and 4b by comparing the difference[23]

$$\Delta\Omega(\mu) = \hat{\Omega}_i^{\text{Ambiguity}}(\mu) - \hat{\Omega}_i^{\text{Insurance}}(\mu + 0.01) \quad \text{if} \quad \mu = 0.01$$
$$\text{and} \tag{2.8}$$
$$\Delta\Omega(\mu) = \hat{\Omega}_i^{\text{Ambiguity}}(\mu) - \hat{\Omega}_i^{\text{Insurance}}(\mu + 0.02) \quad \text{if} \quad \mu \geq 0.02.$$

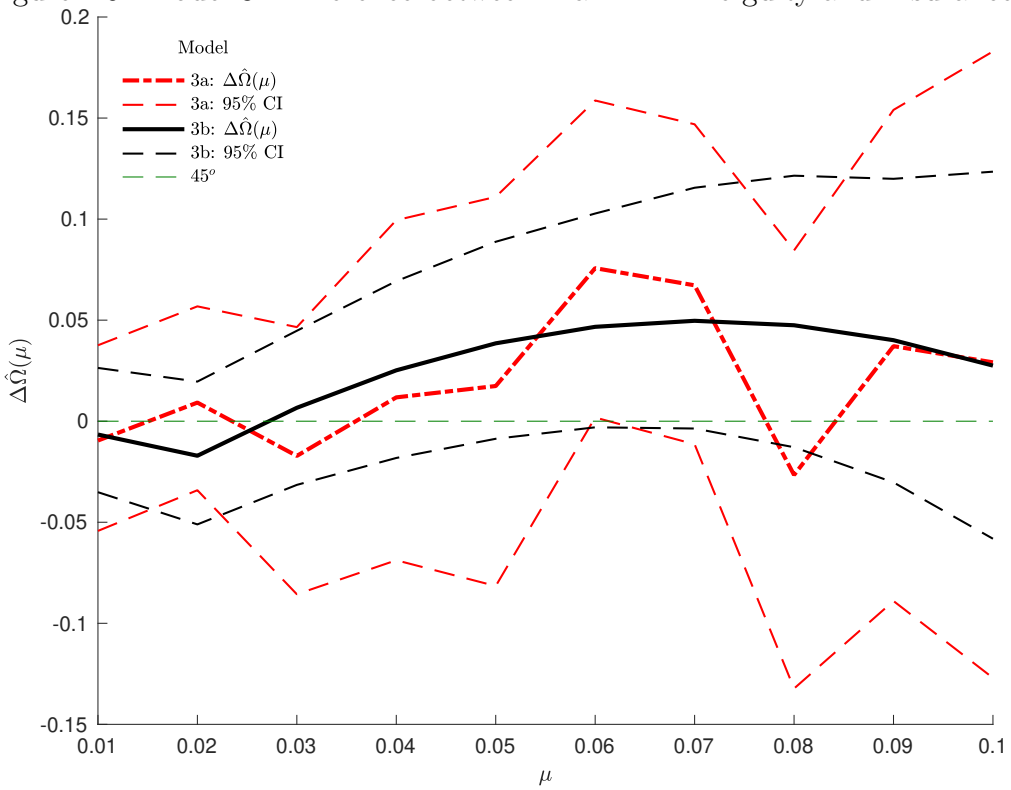
Figure 2.9 displays the Model 4a and 4b Ambiguity treatment effect assuming maxmin utility as a function of μ the median claim rate received. Similar to the prior analysis of the treatment effects, we focus our discussion on the results of Model 4b. We find that under the assumption of maxmin utility the Ambiguity treatment effect is lowest at $\Delta\hat{\Omega}(0.02) = -0.03$ and highest at $\Delta\hat{\Omega}(0.08) = 0.095$; however, the confidence bounds are wide and ranges from 10% to almost 20%.

This finding indicates that not only are our experimental subjects ambiguity averse, their response to ambiguity could be consistent with the maximin expected utility model. It is reasonable then, that estimating our Ambiguity treatment using the median claim probability, would generate more overweighting of probabilities than the Insurance treatment.

The findings from the Ambiguity treatment have important implications for choices made by households in the field when there is possibly both risk and ambiguity. If households in the field are unsure of but aware of the range of their objective claims probabilities, they may also have responded to the ambiguity by

[23]The difference is between the median estimated $\hat{\Omega}_i^{\text{Ambiguity}}(\mu)$ compared to a shifted version of the median estimate $\hat{\Omega}_i^{\text{Insurance}}(\mu)$.

Figure 2.9: Model 3: Difference between Maxmin Ambiguity and Insurance



Note: Displays the difference in the Model 4a and 4b estimated median $\hat{\Omega}_i(\mu)$ from the Ambiguity treatment to the Insurance treatment assuming maxmin utility as defined in Equation 2.8.

making less risky premium-deductible choices than the strict objective probabilities would predict. In other words, ambiguity aversion could contribute to some of the estimated probability distortions observed in BMOT.

2.6 Discussion and Conclusion

We conducted a novel experiment to understand how risk preferences differ between the laboratory and the field. Our experiment was designed to allow us to untangle the role of probability weighting, subjective beliefs, and ambiguity—a challenge for studying risk preferences in the field—and to be more realistic than prior experiments on risk preferences. Subjects made choices over multiple risky

options that closely resemble choices that people have to make in the real world. Unlike analysis of field data that must assume subjects' perceived probabilities of an insurance loss, in the experiment, an explicit claim rate probability or a range of probabilities are given to subjects and are known. And unlike the traditional experimental literature, our choices are not binary and were not certainty equivalents, instead they were modeled after the insurance data of Barseghyan et al. (2013) (BMOT).

Similar to the field literature on risk preferences, we find that the standard risk aversion measures using curvature in the utility function had a negligible impact on premium-deductible choices. Large and overweighting probability distortions play a prominent role in how subjects make decisions over risky prospects.

Unlike results from the field, results from our laboratory experiment suggest less curvature in utility and more between-subjects heterogeneity in choices and estimated risk preferences, particularly in probability distortions. While probability distortions are important in explaining subject choices, the shape of the probability distortion functions in the lab have different theoretical implications than those estimated in the field.

The probability distortion functions estimated in the lab exhibit less probability overweighting at small probabilities than in the field. BMOT found that their estimates from field data could not be explained by other models of risk averse behavior such as KR loss aversion and Gul disappointment aversion, but we were not able to reject those models due to the small probabilities. Our paper does not fully address mechanisms for these lower estimates, but they may be due to a combination of factors. It is possible that heterogenous subjects in the lab may not take the experimental insurance decision as seriously as a home or car owner would

treat the actual decision and thus be more prone to engaging in what Cumulative Prospect Theory called "editing", where subjects assume very low probabilities are zero. It may also be that there is more ambiguity in the real world which could contribute to households inflating the probability of rare events.

Further, the probability distortion functions estimated in the lab have a much steeper increase in the probability distortions with median estimates of $\hat{\Omega}_i(0.1)$ between 30% and 40%. These estimates are much higher than those found in the field and most experimental studies. Our estimates are comparable to the probability weighting estimates of the risky types found in Bruhin et al. (2010), but our subjects exhibited more probability underweighting which is potentially due to the real world multiple choice task as opposed to a certainty equivalent task in their experiments.

We also find that ambiguity aversion may also play a role in insurance decisions in the field. The subjects in our ambiguity treatment chose significantly lower deductibles, and their estimated risk preferences could be explained through the maxmin utility model.

Further research into how misperceptions may influence probability weighting could be conducted where higher amounts of noise in the claim rate probabilities could be given to subjects (i.e. more than $\mu_{ij} \pm 2\%$). In addition, within-subject responses to higher and lower degrees of claim rate ambiguity would also be important to capture in order to investigate individual responses to ambiguity.

CHAPTER 3

DIFFERENT CONTEXTS, DIFFERENT RISK PREFERENCES? (WITH LEVON BARSEGHYAN AND JOSHUA C. TEITELBAUM)

3.1 Introduction

Classical theories of risky choice posit that risk preferences are stable across decision contexts. The stability hypothesis reflects a basic tenet of rational choice theory known as invariance (Tversky and Kahneman, 1986) or context independence (Hausman, 2012). Context independence requires that preferences over options be invariant to the aspects of the choice situation other than the economic fundamentals, which in the case of risky options are the induced lotteries over outcomes.

Broadly speaking, the empirical literature on the stability hypothesis offers two main findings. On the one hand, studies that focus on the (strong) hypothesis of full stability—which usually take a structural approach and examine the within-person consistency of model-based estimates of risk aversion across domains—generally find that a person's risk aversion differs from one domain to the next, suggesting that risk preferences are not perfectly stable across contexts (e.g., Barseghyan et al., 2011). On the other hand, studies that focus on the (weak) hypothesis of some stability—which usually take a model-free approach and examine the within-person correlation of risk taking across domains—generally find that a person who takes on more risk in one context tends to do so in other contexts as well, suggesting that risk preferences have a stable component and are not entirely context dependent (e.g., Einav et al., 2012).

We provide new evidence on the stability hypothesis using data on households' coverage choices in five insurance contexts. A key feature of our data is that three contexts involve small-stakes choices while two involve large-stakes choices. The small-stakes choices are deductibles in three lines of property insurance: auto collision, auto comprehensive, and home all perils. The large-stakes choices are limits in two lines of liability insurance: auto single limit and home personal liability. We adopt the model-free approach of Einav et al. (2012) and assess the stability in ranking across the five contexts of each household's willingness to bear risk relative to its peers. Essentially, we rank the coverage options by risk within each context and compute the pairwise rank correlations among the households' choices across the five contexts. In our preferred baseline specification, we estimate the rank correlations controlling for variation across households in the price of coverage and the risk of loss in each context.

Consistent with prior results in the literature, we find that the households' small-stakes choices are positively rank correlated. We also find that their large-stakes choices are positively rank correlated. Strikingly, however, we find that the households' small-stakes choices are negatively rank correlated with their large-stakes choices. That is, we find that a household who takes on more risk than its peers in small-stakes contexts tends to take on less risk than its peers in large-stakes contexts, and vice versa, which does not support the stability hypothesis, even in its weak form. Moreover, we provide evidence that this result is not driven by heterogeneity in wealth or access to credit. As we argue below, this complicates seemingly ready explanations of our results.

3.2 Related Literature

There are several previous empirical investigations of the stability hypothesis. We highlight a few key studies in the economics literature,[1] giving separate treatment to studies that use data on market choices and those that rely on data from experiments and surveys.

3.2.1 Studies Using Market Data

In an early paper, Wolf and Pohlman (1983) compare the risk aversion of a dealer in U.S. government securities as first measured by his assessments of hypothetical wealth gambles and then estimated from his bid choices in Treasury bill auctions. The authors take a structural approach and assume the dealer is an expected utility (EU) maximizer. They find that "the dealer was substantially more risk averse in his bid choices than his assessments predicted" and conclude that people's "degree of risk aversion may depend on the specific context in which their choices are made" (p. 849).

Though pioneering, Wolf and Pohlman (1983) has two important limitations. First, it studies one person. Second, it compares hypothetical choices with market choices, which confounds the question of stability with that of external validity. Overcoming these limitations, Barseghyan et al. (2011) examine the deductible choices of 702 households across three insurance contexts: auto collision, auto comprehensive, and home all perils. Assuming that households are EU maximizers, the authors obtain three interval estimates of each household's risk aversion based

[1] Barseghyan et al. (2011) and Galizzi et al. (2016b) discuss additional studies. A separate literature investigates the consistency of risk preference measures obtained from different elicitation methods employed in experiments and surveys. For a summary, see Galizzi et al. (2016a).

on its three choices. They find that these intervals intersect—implying that the choices can be rationalized by the same degree of risk aversion—for only 23 percent of households, leading them to reject the hypothesis of full stability.

Rejecting the hypothesis of full stability does not imply that risk preferences have no stable component. Moreover, structural approaches to testing stability invariably comprise a joint test of the stability hypothesis and the assumptions of the structural model. With these points in mind, Einav et al. (2012) examine the workplace benefits choices made by 12,752 Alcoa employees in six contexts: health insurance, drug insurance, dental insurance, short-term disability insurance, long-term disability insurance, and 401(k) investments. The authors pursue a model-free approach (which we adopt here) in which they rank by risk the options within each context and assess the rank correlation of the employees' choices across the six contexts. They find that an employee's choice in each context is positively rank correlated with her choice in every other context, with stronger correlations across "closer" contexts (p. 2609), leading them to reject the hypothesis of no stability and conclude that risk preferences have a context-invariant component.[2]

In the wake of this methodological shift, Barseghyan et al. (2016) explore the connection between full stability under a structural approach and rank stability under a model-free approach. Using data on the deductible choices of 3,629 households across the three insurance contexts studied by Barseghyan et al. (2011), the authors document two findings: (i) the households' deductible choices are positively rank correlated, echoing the finding of Einav et al. (2012), and (ii) five in six households exhibit full stability under a rank-dependent EU model. They then show that the fully stable households drive the rank correlations.

[2]Einav et al. (2012) also pursue a structural approach that is conceptually similar to the approach in Barseghyan et al. (2011). Under this approach, they find that for roughly 30 percent of employees all six choices can be rationalized by the same degree of risk aversion.

Our paper builds directly on Einav et al. (2012). Like them, we take a model-free approach and examine rank stability across multiple contexts using data on market choices. The main distinction between our papers is the degree to which the stakes vary across contexts. In the contexts we study, the dollar values of the options range from the hundreds and thousands (in our small-stakes contexts) to the hundreds of thousands and millions (in our large-stakes contexts). As we discuss in Section 3.4.3, the dollar values of the options in Einav et al. (2012) range from the hundreds and thousands (in three contexts) to the tens of thousands (in the others). It is this distinction that reconciles our results. Both papers find evidence of rank stability across contexts involving stakes of the same or near orders of magnitude, while ours also finds evidence of rank instability across contexts involving stakes of remote orders of magnitude.[3]

3.2.2 Studies Using Nonmarket Data

Anderson and Mellor (2009) compare the responses of laboratory subjects to a series of hypothetical job gambles and a series of hypothetical inheritance gambles. The authors construct a categorical measure of the subjects' risk aversion based on the job gamble responses and then do the same for the inheritance gamble responses. They find that 34 percent of subjects exhibit the same degree of risk aversion across the two contexts and report a rank correlation of 0.175 between

[3]Collier et al. (2017) also study choices with remote stakes. Using data on households' deductibles and coverage limits in flood insurance, the authors structurally estimate the risk preferences implied by the two choices and find that they differ. Because they take a structural approach, their paper relies on stronger modeling assumptions than ours. Indeed, their estimation approach—parametric MLE of a random utility model—entails even stronger assumptions than the partial identification approach taken by Barseghyan et al. (2011). In addition, the two choices they study are made in the same context. We therefore view their paper as more in line with the related literature on how risk aversion varies with stake size (e.g., Binswanger, 1980; Kachelmeier and Shehata, 1992; Holt and Laury, 2002; Fehr-Duda et al., 2010).

the two measures.[4] Dohmen et al. (2011) use survey data from the German Socio-Economic Panel to compare respondents' self-reported willingness to take risks across five contexts: car driving, financial matters, sports and leisure, health, and career. The authors report that while the responses "are not perfectly correlated across contexts, ... the pairwise correlations are large, typically in the neighborhood of 0.5," which they argue "is suggestive of a stable, underlying risk trait" (p. 537). More recently, Ioannou and Sadeh (2016) compare the selections made by laboratory subjects from a set of real monetary gambles and a set of real "environmental" gambles (where the payoffs are numbers of bee-friendly plants). The authors find that subjects "exhibit a higher degree of risk aversion in the environmental domain relative to the monetary domain; that is, individuals tend to be more reluctant to take on large gambles with environmental outcomes than with monetary ones" (p. 31).[5] In addition to using nonmarket data, these studies differ from ours in that they either lack meaningful variation in stakes across contexts (e.g., Anderson and Mellor, 2009; Ioannou and Sadeh, 2016) or they study general domains of risky behavior in which the stakes are neither explicit nor well-defined (e.g., Barksy et al., 1997; Dohmen et al., 2011).

[4]The job and inheritance gamble questions are taken from the Heath and Retirement Study (HRS). Barksy et al. (1997) use the responses to the job gamble questions in the HRS to construct a measure of respondents' risk tolerance. They then present evidence that their measure predicts certain self-reported risky behaviors, "including smoking, drinking, not having insurance, choosing risky employment, and holding risky assets" (p. 551).

[5]In another incentivized experiment, Choi et al. (2007) test within-subject consistency (assuming maximization of a well-behaved concave utility function) across 50 risky portfolio choices. They find that while only 17 percent of subjects exhibit perfect consistency (app. C), a "significant majority" perform "only a bit worse" (pp. 1927-1928). More to the point, the authors report (without providing details) that "some subjects" exhibit a "switching" pattern—sometimes choosing extremely safe portfolios, sometimes choosing extremely risky portfolios, and sometimes chooisng intermediate portfolios—wherein their choices are "individually consistent" with risk averse utility maximization but "mutually inconsistent" with one another (pp. 1925 & 1936-1937).

3.3 Data and Sample

The source of our data is a large U.S. property and casualty insurance company. Our dataset contains annual information on more than 400,000 households who purchased auto or home insurance from the company between 1998 and 2007. The data contain all the information in the company's records regarding the households and their policies, including claims information.

We focus on three small-stakes choices and two large-stakes choices. The small-stakes choices are deductibles in three lines of property coverage: auto collision, auto comprehensive, and home all perils. Auto collision coverage pays for damage to the insured vehicle caused by a collision with another vehicle or object, without regard to fault. Auto comprehensive coverage pays for damage to the insured vehicle from all other causes, without regard to fault. Home all perils coverage pays for damage to the insured home from all causes, except those that are specifically excluded (e.g., flood). The deductible options range from \$100 to \$1,000 in auto collision, \$50 to \$1,000 in auto comprehensive, and \$100 to \$5,000 in home all perils. The mean increment between options is \$225 in auto collision, \$190 in auto comprehensive, and \$980 in home all perils.

The large-stakes choices are limits in two lines of liability coverage: auto single limit and home personal liability. Auto single limit coverage pays for bodily injury or property damage to others for which the insured driver is legally responsible. Home personal liability coverage pays for bodily injury or property damage to others for which the insured homeowner is legally responsible. The limit options range from \$60,000 to \$1,000,000 in auto single limit and \$100,000 to \$1,000,000 in home personal liability. The mean increment between options is \$188,000 in auto single limit and \$180,000 in home personal liability.

Our baseline sample comprises households who (i) purchased all three property coverages and both liability coverages and (ii) first purchased each coverage within any six-month window during the period from 2004 to 2007. The latter restriction helps avoid temporal issues, such as changes in household characteristics or the economic environment. We consider only the households' coverage choices at the time of first purchase. This helps ensure that we are working with active choices; one might worry that households renew their policies without actively reassessing their coverage options (Handel, 2013). These restrictions yield a baseline sample of 2,690 households.

For each household in our baseline sample, we observe its deductible or limit choice (as the case may be) in each coverage, as well as the pricing menu it faced in each coverage. According to conversations with the company and an independent agent who sells company policies, the choice environment is conducive to households making active and informed choices—there are no default choices, the pricing menu is available to a household when it makes a choice, and a household must choose a deductible or limit separately for each coverage.

Tables 3.1 and 3.2 provide descriptive statistics for the baseline sample. Table 3.1 reports demographic characteristics and claim frequencies. Table 3.2 summarizes the coverage choices and pricing menus.

Table 3.1: Demographics and Claims

	Mean	Std dev	5th pctl	95th pctl
Auto policies				
Driver 1 age (years)	57	15	32	80
Driver 1 female	0.38	0.49		
Driver 1 single	0.2	0.4		
Driver 1 married	0.58	0.49		
Driver 2 indicator	0.43	0.5		
Driver 3+ indicator	0.03	0.16		
Vehicle 1 age (years)	4.95	3.36	1	11
Vehicle 2 indicator	0.48	0.5		
Vehicle 3+ indicator	0.03	0.17		
Insurance score	788	106	602	957
Collision claims (per annum)	0.089	0.286	0.000	0.600
Comprehensive claims (per annum)	0.024	0.125	0.000	0.000
Single limit claims (per annum)	0.085	0.277	0.000	0.597
Home policies				
Home age (years)	44	31	2	105
Home value (thousands of dollars)	213	155	90	430
Insurance score	733	100	562	888
All perils claims (per annum)	0.058	0.192	0	0.451
Personal liability claims (per annum)	0.002	0.027	0.000	0.000

Notes: The table reports descriptive statistics for the baseline sample of 2,690 households. Insurance scores in auto and home are based on information contained in credit reports.

Table 3.2: Choices and Prices

	Share	Premium saving relative to safest option (dollars)			
	(percentage)	Mean	Std dev	5th pctl	95th pctl
Auto collision					
$100	1				
$200	15.2	40	23	15	84
$250	11.6	80	46	31	168
$500	63.8	134	77	52	281
$1,000	8.3	174	100	67	365
Auto comprehensive					
$50	5.1				
$100	4.7	45	32	15	93
$200	34.9	67	48	23	140
$250	11.2	74	53	26	155
$500	39.3	104	75	36	217
$1,000	4.8	127	91	43	264
Home all perils					
$100	0.3				
$250	22.3	186	156	83	403
$500	54.9	248	207	110	529
$1,000	21	330	275	146	694
$2,500	1.3	391	326	176	820
$5,000	0.3	463	386	206	1001
Auto single limit					
$60,000	0.2	109	46	55	200
$100,000	8.6	102	43	52	189
$200,000	0.7	78	33	40	143
$300,000	43.9	68	29	34	125
$500,000	43	57	24	29	106
$1,000,000	3.6				
Home personal liability					
$100,000	9.6	42			
$200,000	0.8	32			
$300,000	47.6	24			
$400,000	0.2	19			
$500,000	36.4	16			
$1,000,000	5.4				

Notes: The table summarizes the coverage choices and pricing menus for the baseline sample of 2,690 households. Share is the percentage of households who chose a given option (deductible or limit, as the case may be). The safest option is the lowest deductible in the property coverages and the highest limit in the liability coverages.

3.4 Methods and Results

3.4.1 Empirical Strategy and Baseline Results

We adopt the model-free approach of Einav et al. (2012) and assess the stability in ranking across contexts of each household's willingness to bear risk relative to its peers. To begin, we rank the options by risk within each context, ordering them from highest to lowest risk exposure. There are five or six options in each context (see Table 3.2). The safest option is the lowest deductible in property coverages and the highest limit in the liability coverages. We then compute the pairwise Spearman rank correlations in the households' choices across the five contexts.

Because these rank correlations do not control for potentially important covariates, we also examine the correlation structure of the residuals from a system of five equations:

$$\begin{bmatrix} y_i^{Auto\ collision} \\ y_i^{Auto\ comprehensive} \\ y_i^{Home\ all\ perils} \\ y_i^{Auto\ single\ limit} \\ y_i^{Home\ personal\ liability} \end{bmatrix} = \begin{bmatrix} \beta^{Auto\ collision} \\ \beta^{Auto\ comprehensive} \\ \beta^{Home\ all\ perils} \\ \beta^{Auto\ single\ limit} \\ \beta^{Home\ personal\ liability} \end{bmatrix} \cdot \mathbf{x}_i + \begin{bmatrix} \varepsilon_i^{Auto\ collision} \\ \varepsilon_i^{Auto\ comprehensive} \\ \varepsilon_i^{Home\ all\ perils} \\ \varepsilon_i^{Auto\ single\ limit} \\ \varepsilon_i^{Home\ personal\ liability} \end{bmatrix}, \tag{1}$$

where y_i^j denotes the rank-ordered choice of household i in context j, β^j is a vector of context-specific coefficients, $\mathbf{x}_i$ is a vector of household-specific covariates, and ε_i^j is a household- and context-specific residual. In theory, a household's choices depend not only on its risk preferences but also on the prices it faces and its risk profile. The baseline set of covariates ($\mathbf{x}_i$), therefore, includes controls for prices and risk. The price controls are log-transformed premiums for each coverage

assuming a $250 deductible or $200,000 limit, as the case may be.[6] The risk controls are expected annual claims under each coverage based on separate Poisson-gamma Bayesian credibility models. By construction, the risk controls take into account both the systematic and idiosyncratic components of a households' risk type. For further details, see Appendix C.1.

Following Einav et al. (2012), we estimate system (1) in two different ways. First, we treat it as a multivariate ordered probit regression model and estimate it by maximum likelihood.[7] Second, we treat it as a multivariate linear regression model and estimate it by least squares. Because the set of options in each context is discrete, the probit regression is our preferred specification.

Table 3.3 reports the baseline results. Panel A shows the Spearman rank correlations. Panels B and C display the estimated correlations from the probit and linear regressions, respectively. Each panel tells the same story. Across all panels, the correlation between each pair of small-stakes choices is positive, ranging from 0.26 to 0.70. Similarly, the correlation between the two large-stakes choices is positive, ranging from 0.44 to 0.57. By contrast, however, the correlation between every pairing of a small-stakes choice and a large-stakes choice is negative, ranging from -0.05 to -0.34. Overall, the baseline results suggest that the households exhibit a fairly stable degree of risk aversion relative to their peers across contexts that involve stakes of the same order of magnitude. At the same time, however, the results suggest that households who exhibit a higher degree of risk aversion than their peers in small-stakes contexts tend to exhibit a lower degree of risk aversion than their peers in large-stakes contexts, and vice versa.

[6] We do not include a price control for home personal liability because the premiums do not vary across households.

[7] We estimate the system by performing bivariate ordered probit regressions on every pair of equations. In each regression, we use the Huber/White/sandwich estimator to obtain robust standard errors.

Table 3.3: Baseline Results

	Auto collision	Auto comprehensive	Home all perils	Auto single limit
Panel A. Spearman rank correlations				
Auto comprehensive	0.617			
Home all perils	0.395	0.383		
Auto single limit	-0.129	-0.108	-0.224	
Home personal liability	-0.206	-0.219	-0.339	0.563
Panel B. Correlation estimates from probit regression				
Auto comprehensive	0.702			
Home all perils	0.398	0.338		
Auto single limit	-0.104	-0.056	-0.134	
Home personal liability	-0.149	-0.133	-0.205	0.574
Panel C. Correlation estimates from linear regression				
Auto comprehensive	0.552			
Home all perils	0.29	0.263		
Auto single limit	-0.077	-0.055	-0.113	
Home personal liability	-0.121	-0.114	-0.163	0.437

Notes: The table provides results for the baseline sample of 2,690 households. Each cell reports a pairwise correlation coefficient. For each correlation coefficient, the p-value associated with a test of whether the coefficient is different from zero is less than 0.01. The only exception is the correlation coefficient between auto comprehensive and auto single limit in panel B, for which the associated p-value is 0.023. The probit and linear regressions include controls for prices and risk.

3.4.2 Sensitivity Analysis

Umbrella Coverage

Twenty-six percent of the households in the baseline sample purchased umbrella liability coverage from the company to supplement their auto single limit and home personal liability coverages. The umbrella coverage options range from $1 million to $5 million in $1 million increments, and the premium associated with each coverage option is the same for all households.

The baseline results disregard the households' umbrella choices. To explore whether this biases our results, we treat households who purchased umbrella cov-

erage as having chosen a new "highest limit" option (i.e., a limit of unspecified amount greater than $1,000,000) in auto single limit and home personal liability, and we re-estimate (1) treating it as a system of ordered probits and including the baseline set of controls.[8]

Table 3.4, panel A reports the results, which tell the same story as the baseline results. Indeed, all but one of the pairwise correlations involving a liability insurance context are stronger than the corresponding baseline correlations. The only exception is the correlation between home all perils and home personal liability, which is slightly weaker than the corresponding baseline correlation.[9]

[8]We do not add a price control for umbrella coverage because the premiums do not vary across households.

[9]As a further check, we also re-estimate (1), again treating it as a system of ordered probits and including the baseline set of controls, on the subsample of 1,993 households who did not purchase umbrella coverage. Those results also tell the same story as the baseline results.

Table 3.4: Sensitivity Analysis

	Auto collision	Auto comprehensive	Home all perils	Auto single limit
Panel A. Correlation estimates accouting for umbrella choices				
Auto comprehensive	0.702			
Home all perils	0.398	0.338		
Auto single limit	-0.134	-0.103	-0.135	
Home personal liability	-0.165	-0.150	-0.176	0.842
Panel B. Correlation estimates with control for weath				
Auto comprehensive	0.703			
Home all perils	0.399	0.336		
Auto single limit	-0.106	-0.060	-0.144	
Home personal liability	-0.151	-0.138	-0.214	0.570
Panel C. Correlation estimates with controls for insurance scores				
Auto comprehensive	0.702			
Home all perils	0.398	0.338		
Auto single limit	-0.105	-0.061	-0.136	
Home personal liability	-0.148	-0.132	-0.204	0.576
Panel D. Correlation estimates with same-day choice window				
Auto comprehensive	0.707			
Home all perils	0.446	0.375		
Auto single limit	-0.137	-0.103	-0.157	
Home personal liability	-0.133	-0.159	-0.203	0.649

Notes: Panels A, B, and C provide results for the baseline sample of 2,690 households. Panel D provides results for the subsample of 1,694 households who purchased all five coverages on the same day. Each cell reports a pairwise correlation coefficient estimated from a system of ordered probits with controls for prices and risk. In panel A, households who purchased umbrella coverage are treated as having chosen a new "highest limit" option in both auto single limit and home personal liability. In panel B, the probit regression includes an additional control for wealth. In panel C, the probit regression includes additional controls for insurance scores in auto and home. For each correlation coefficient, the p-value associated with a test of whether the coefficient is different from zero is less than 0.01. The only exception is the correlation coefficient between auto comprehensive and auto single limit in panels B and C, for which the associated p-value is 0.015.

Wealth

Economists have long hypothesized that risk preferences depend on wealth (Friedman and Savage, 1948; Pratt, 1964; Arrow, 1971). The standard assumption is that absolute risk aversion is decreasing in wealth, which implies that, ceteris paribus, a household's willingness to pay for insurance decreases with its wealth. See, for

example, Pratt (1964, pp. 122-123): "Utility functions for which [the coefficient of absolute risk aversion] is decreasing are logical candidates to use when trying to describe the behavior of people who, one feels, might generally pay less for insurance against a given risk the greater their assets."

Our baseline analysis does not control for household wealth. To examine whether wealth effects may be driving our results, we add a control for wealth to the baseline set of controls and re-estimate (1) treating it as a system of ordered probits. We do not directly observe a household's wealth in our data, but we do observe a plausible proxy: the insured value of the dwelling covered by its home-owners policy ("home value"). Of course, we do not know the correlation between home value and wealth in our data. However, according to combined extract data (1989-2016) from the Survey of Consumer Finance, the correlation between home value and wealth is 0.47 (std. err. = 0.002).

Table 3.4, panel B reports the results. Each pairwise correlation is virtually identical to the corresponding baseline correlation. It thus appears that wealth effects are not driving our results.

Access to Credit

In theory, a household's ability to borrow after a loss event can affect its demand for insurance (Handel et al., 2015; Jaffe and Malani, 2017). To investigate whether differences in access to credit may be driving our results, we add controls for households' insurance scores in auto and home to the baseline set of controls and re-estimate (1) treating it as a system of ordered probits. Insurance scores are akin to credit scores. Both are derived using the same five categories of information contained in credit reports (payment history, level of indebtedness, length of credit

history, new credit and pursuit of new credit, and types of credit), though they differ somewhat in how they weight these categories (Morris et al., 2017). For this reason, we believe that insurance score, like credit score, is a good proxy for a household's access to credit.[10]

Table 3.4, panel C reports the results. Again, each pairwise correlation is virtually identical to the corresponding baseline correlation. This suggests that differences in access to credit are not driving our results.[11]

Choice Window

In the baseline sample, we restrict attention to households who, inter alia, purchased all five coverages within a six-month window. There are two opposing considerations in selecting a choice window. On the one hand, a narrower window helps to avoid what Einav et al. (2012, p. 2611) call "the problems of inferring preferences from 'stale' choices," which they note "could be particularly concerning if individuals might have made their choices ... at different points in time." On the other hand, a wider window helps to improve inference by increasing sample size.

We are not concerned that a six-month window is too narrow. Our baseline sample comprises 2,690 households, which we believe is sufficiently large to draw valid inferences. To address the concern that a six-month window may be too wide, we re-estimate (1) on the subsample of 1,694 households who purchased all five coverages on the same day. As before, we treat (1) as a system of ordered

[10] There is ample evidence that credit score is a good proxy for access to credit (e.g., Baker, 2017).

[11] As a further check, we re-estimate (1), again treating it as a system of ordered probits, with controls for wealth and insurance scores (and their interactions) added to the baseline set. Once again, the results tell the same story.

probits and include the baseline set of controls. Table 3.4, panel D reports the results. They tell the same story as the baseline results. Indeed, all but two of the pairwise correlations are stronger than the corresponding baseline correlations. The only exceptions are the pairwise correlations between auto collision and home personal liability and between home all perils and home personal liability, which are slightly weaker than baseline.

3.4.3 Comparison with Einav et al. (2012)

We close this section with a discussion comparing our results with those of Einav et al. (2012). Using data on the workplace benefits choices of 12,752 Alcoa employees, Einav et al. (2012) pursue the same model-free approach (which they develop) to assess the rank stability of the employees' risk preferences across six contexts: health insurance, drug insurance, dental insurance, short-term disability insurance, long-term disability insurance, and 401(k) investments. In their baseline analysis, where they control for variation in benefit menus, they find that an employee's choice in every context is positively rank correlated with its choice in every other context, implying that employees who exhibit a higher degree of risk aversion than their peers in one context tend also to do so in other contexts, and vice versa. They find very similar results when they add controls for risk. The strongest pairwise correlations are between short- and long-term disability insurance (0.76) and among health, drug, and dental insurance (ranging from 0.30 to 0.49). Somewhat weaker are the correlations across the disability and medical insurance contexts (ranging from 0.21 to 0.26). The weakest are between 401(k) investments and every other context (all below 0.05, including two that are slightly negative but not

statistically different from zero).[12]

In order to compare our results with those of Einav et al. (2012), we must classify their contexts according to the magnitude of the stakes involved, applying the same criteria that we use to classify our contexts. Recall that in our small-stakes contexts the values of the options and the inter-option increments range in the hundreds and thousands dollars, whereas in our large-stakes contexts the value of the options range in the hundreds of thousands and millions of dollars with inter-option increments that range in the hundreds of thousands dollars.

For the reasons we detail in Appendix C.2, we conclude that none of the contexts in Einav et al. (2012) involve large-stakes choices. Specifically, we conclude that three contexts—health, drug, and dental insurance—involve small-stakes choices. In two contexts—short-term disability insurance and 401(k) investments—we determine that the stakes range in the thousands and tens of thousands of dollars but not the hundreds of thousands of dollars, and so we classify them as moderate-stakes contexts. We also classify the remaining context—long-term disability insurance—as involving moderate-stakes choices, though the reasons are less straightforward.[13]

Given these classifications, we see that our results and those of Einav et al. (2012) complement one another. We both find a pattern of positive pairwise corre-

[12]The quoted results are from Table 3.3B, panel A in Einav et al. (2012), which reports correlation estimates from a system of ordered probits with controls for benefit menus and risk.

[13]We note that in an effort to establish the comparability of the choices they study, Einav et al. (2012, p. 2616) argue that "the incremental decisions across each domain are quite comparable in *expected* magnitude, ... ranging from several hundred to a few thousand dollars" (emphasis added). We do not disagree. But the fact remains that the choices in their first three contexts differ categorically from the choices in their last three contexts in terms of the absolute magnitude of the stakes involved. Morever, were we to classify choices according to the expected magnitudes of the options (or inter-option increments), this arguably would be inconsistent with taking a model-free approach, as it would presuppose a model that entails comparisons over expected values or utilities.

lations among small-stakes choices. To this common result, Einav et al. (2012) add two findings: patterns of positive (or at least non-negative) pairwise correlations among moderate-stake choices and across small- and moderate-stakes choices. We also add two findings. The first is a pattern of positive pairwise correlations among large-stakes choices, which taken together with the previous findings hints at a stable component of risk preferences that operates across contexts involving stakes of the same or near orders of magnitude. The second finding that we add to the mix is our main contribution: a pattern of negative pairwise correlations across small- and large-stakes choices, which hints at a lack of risk preference stability across contexts involving stakes of remote orders of magnitude.

There is another way to see how we build on Einav et al. (2012). Leaving 401(k) investments aside for the moment, Einav et al. (2012) find (i) moderately positive correlations between contexts involving stakes of the same order of magnitude (small/small or moderate/moderate) and (ii) weakly positive correlations between contexts involving stakes of adjacent orders of magnitude (small/moderate). We corroborate the first finding (for small/small) and extend it (to large/large) and progressively add a third: (iii) weakly negative correlations between contexts involving stakes of remote orders of magnitude (small/large). Returning to 401(k) investments, Einav et al. (2012) acknowledge that this context is "the most difficult to reconcile with any of the others" (p. 2636), and they attribute the difficulty to a difference in kind between investments and insurance. Our results suggest an alternative explanation: employees may perceive 401(k) investments as a borderline large-stakes context, particularly if they view their allocation choice as applying to more than just their current year's contributions. This could explain the extremely weak correlations (more or less zero) between 401(k) investments and every other context in Einav et al. (2012).

3.5 Discussion

We examine the hypothesis that risk preferences have a stable, context-invariant component using data on households' insurance choices. We study five insurance contexts, three involving small-stakes choices (deductibles) and two involving large-stakes choices (liability limits). Adopting the model-free approach of Einav et al. (2012), we assess the extent to which the households' choices display a stable ranking in their willingness to bear risk relative to their peers. While we find that the households' choices reflect a stable ranking in risk taking across the three small-stakes choices and across the two large-stakes choices, we also find that the households who take on more risk than their peers in small-stakes contexts tend to take on less risk than their peers in large-stakes contexts, and vice versa, which does not support the stability hypothesis.

What could explain our results? Three stories come readily to mind. None is unassailable, however, and so each leaves open questions for future research.

The first is a story about relative risk aversion. Suppose that rich households choose higher deductibles and higher liability limits than poor households. The intuition might be that rich households want insurance against large losses but not small losses (which they can self-insure at a lower cost), whereas poor households want insurance against small losses but not large losses (because you can't get blood from a stone). This pattern of choices, which could explain our results, could arise from a population of households with standard EU preferences and heterogeneous relative risk aversion. Standard EU preferences feature a concave utility function that is defined over wealth and exhibits DARA (Pratt, 1964; Arrow, 1971).[14] Concavity implies a positive willingness to pay for insurance. Let π

[14]In this paragraph, DARA stands for decreasing absolute risk aversion, and IRRA, CRRA,

denote this willingness. DARA implies that π decreases with wealth, which could account for rich households choosing higher deductibles than poor households. If the utility function also exhibits IRRA/CRRA/DRRA,[15] then, ceteris paribus, π is increasing/constant/decreasing in stakes (Menezes and Hanson, 1970; Zeckhauser and Keeler, 1970). Thus, the right kind of heterogeneity in relative risk aversion (e.g., rich households have IRRA and poor households have CRRA) could account for rich households also choosing higher liability limits than poor households.

This story, while plausible, has at least two important counterpoints. The first is our analysis in Section 3.4.2, which casts doubt on the possibility that wealth differences are behind our results. The second is the Rabin (2000) critique, which contends that EU theory is not a plausible model of risk aversion across small- and large-stakes gambles.[16]

A second story features consumption commitments (i.e., spending obligations that are costly to adjust). Suppose that some households have consumption commitments while others do not. Chetty and Szeidl (2007) show that, within an EU framework, consumption commitments can induce non-concavities in the utility function (cf. Friedman and Savage, 1948; Markowitz, 1952) that increase risk aversion over small- and moderate-stakes gambles relative to large-stakes gambles. Hence, the right kind of heterogeneity in consumption commitments (e.g., committed households have lower risk aversion over large-stakes gambles than other households) could generate a pattern of choices in which committed households choose lower deductibles and lower liability limits than other households, which could explain our results.

---

and DRRA stand for increasing, constant, and decreasing relative risk aversion, respectively.

[15] For instance, Saha's (1993) expo-power utility function can exhibit DARA/IRRA or DARA/DRRA, while the power utility function exhibits DARA/CRRA.

[16] We note that the Rabin critique is not directly related to our main finding of rank instability of risk preferences across small- and large-stakes contexts.

Again, this story, while plausible, is complicated by our sensitivity analysis. One implication of Chetty and Szeidl's (2007) theory is that consumption commitments "have a larger effect on risk aversion when agents are borrowing constrained" (p. 850). It follows that if heterogeneity in consumption commitments were driving our results, we would expect them to be sensitive to differences in access to credit. Our analysis in Section 3.4.2, however, suggests they are not.[17]

Probability distortions headline a third possible story. Suppose that households' subjective beliefs (in a subjective EU model) or decision weights (in a rank-dependent EU model) do not correspond to the objective risks. The right kind of heterogeneity in such beliefs or weights could explain our results. For example, suppose that some households overweight loss probabilities in large-stakes gambles but not small-stakes gambles, while other households overweight loss probabilities in small-stakes gambles but not large-stakes gambles. This could lead the former households to choose higher deductibles and higher liability limits than the latter households. Alternatively, suppose that some households grossly overweight loss probabilities in small-stakes gambles and mildly overweight them in large-stakes gambles, while other households do not overweight loss probabilities in any gambles (cf. Fehr-Duda et al., 2010).[18] If in addition the former households are low risk while the other households are high risk, this could lead the former to choose lower deductibles and lower liability limits than the latter.

The issue with each version of this story is that it requires a peculiar heterogeneity structure. (Indeed, we could level this criticism against the first two stories as well.) We are not aware of any empirical or theoretical support for the kind

[17] Although Chetty and Szeidl (2007) adopt an EU framework, the non-concavities of the utility function insulate their model from the Rabin critique.

[18] Fehr-Duda et al. (2010) present evidence of this kind of pattern, but only for gambles in the gain domain. They find no substantial difference in stake-dependent probability weighting for gambles in the loss domain (like insurance).

of heterogeneity—including, in some versions, the correlation between probability distortions and risk types—that is required by this story.

In future research it would be worthwhile to further probe these and other potential explanations of our results and to explore whether similar results obtain in other comparable datasets.

APPENDIX A

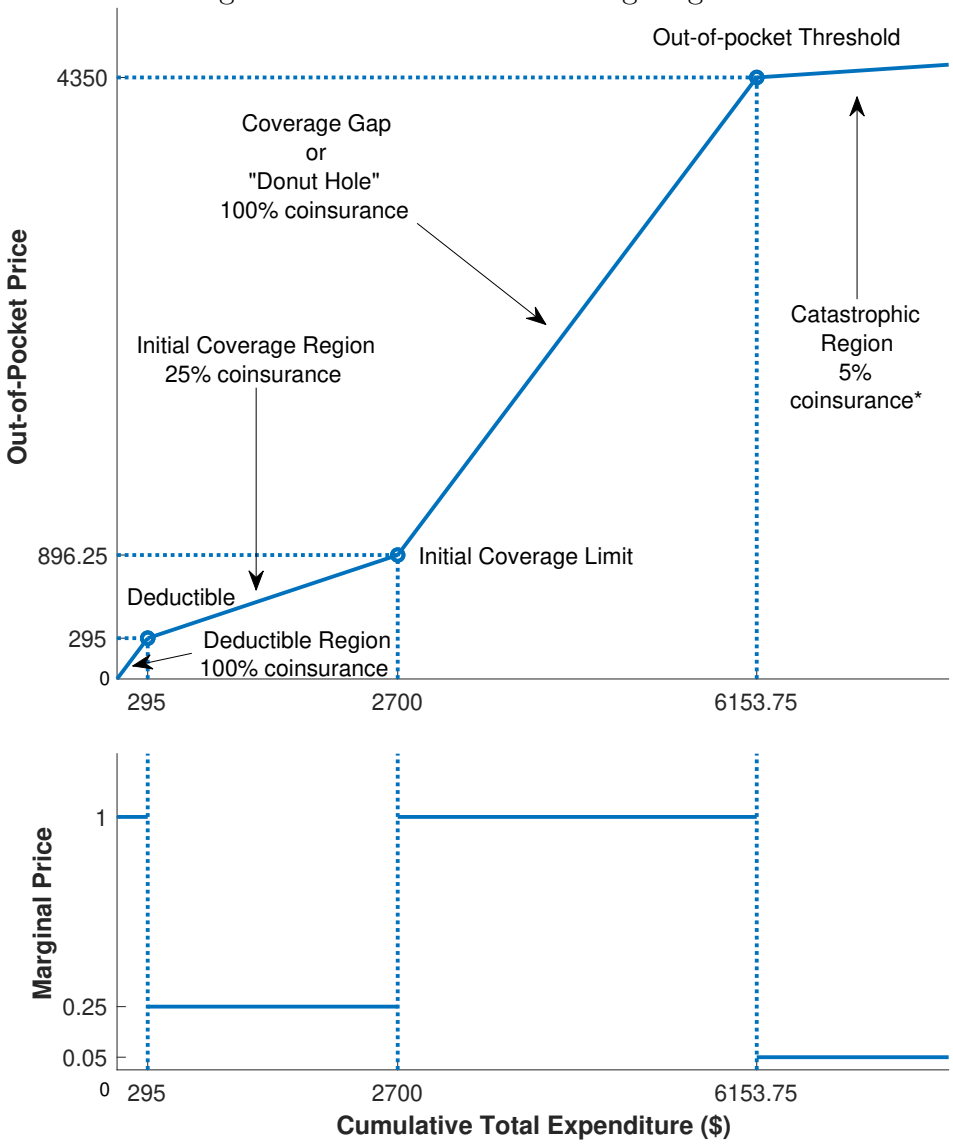
APPENDIX OF CHAPTER 1: DYNAMIC PURCHASING BEHAVIOR IN HEALTHCARE CONSUMPTION

A.1 Data Appendix

Appendix Figure A.1 displays the 2009 Medicare-defined standard plan.

Appendix Table A.1 details the reasons why beneficiaries are excluded from the full sample and also includes the percentage of the 5% sample they encompass. Beneficiaries are limited to those who are 65 or older, are enrolled in Medicare PDPs in 2009-2012, and who originally and are currently enrolled in Medicare through the Old Age and Survivors Insurance (OASI) and not for disability insurance or other qualifiers for Medicare. Similar to Einav et al. (2015), the sample also excludes individuals who are dual eligible for Medicaid financial assistance or receive other types of low-income subsidies (LIS) for premiums or cost-sharing. These individuals are excluded because they face very low cost-sharing and minor differences in their marginal costs. Individuals who only receive premium subsidies are also omitted, because they are more likely lower income and are more likely to be influenced by budget constraints. Joyce et al. (2013) used these LIS individuals as a control group for their analysis of the donut hole, but this paper takes a different approach. Further, my analysis also excludes individuals whose Medicare Part B claims indicate beneficiaries were in long-term care institutions (LTI) such as nursing homes in the prior year. The LTI beneficiaries are excluded on the chance that their Part D prescription purchases are managed by the nursing homes. In addition, the full sample excludes individuals who did not have a Medicare Part D plan for the full year ("Not Same Plan in 12 Months"), because those individuals

Figure A.1: Part D 2009 coverage regions



Note: The figure depicts the nonlinear structure of the standard Medicare Part D benefit contract. Actual plans offered in 2009 were either actuarially equivalent or better. The premium or the amount the patient pays out-of-pocket for the benefit package is not displayed. The Total Expenditure includes the drug expenditure between the patient, insurance company, and Medicare, while the Out-of-Pocket (OOP) Cost only includes the patient's drug expenditure. The 5% coinsurance coverage in the catastrophic region is simplified for the figure. The actual 2009 coverage benefit requires beneficiaries to pay the maximum of either 5% the cost of the drug or $2.40 and $6.00 for a one-month supply of generic and branded drugs respectively. This means that patients may pay either the copay dollar amount or a percentage share of the drug price and the remainder is covered by insurance or the government. The bottom panel displays the marginal costs of the plan.

would have varying time-frames over which their contracts run.

Table A.1: Summary of 5% sample of Medicare data and exclusion reason 2009-2012

	(1)		(2)		(3)		(4)	
	2009		2010		2011		2012	
	mean	sd	mean	sd	mean	sd	mean	sd
Dual eligible	20.01	40	20.09	40	22.94	42	22.71	42
Disabled or end-stage renal disease	16.05	37	17.00	38	17.13	38	17.04	38
Originally not OASI	22.73	42	23.86	43	24.15	43	24.26	43
Currently not OASI	16.53	37	17.89	38	18.26	39	18.25	39
Receive state subsidy	17.52	38	17.92	38	18.15	39	18.16	39
No prescription coverage	40.64	49	40.46	49	39.03	49	36.91	48
Low-Income subsidy	22.32	42	22.62	42	22.72	42	22.55	42
Long term care (or undef)	4.38	20	04.65	21	4.69	21	4.58	21
Cost sharing other (or undef)	43.48	50	43.46	50	41.69	49	39.40	49
Employer subsidy	14.94	36	14.65	35	16.38	37	14.93	36
Undefined creditable coverage	0.01	1	0.01	1	0.05	2	0.06	3
Has creditable coverage	18.48	39	18.81	39	16.29	37	15.44	36
Died in plan year	14.97	36	4.30	20	3.97	20	3.90	19
Not same plan in 12 months	43.76	50	44.46	50	41.96	49	40.03	49
Not PDP	63.89	48	64.38	48	63.62	48	62.97	48
Employer group waiver (or undef)	48.62	50	48.42	50	47.57	50	47.05	50
Full 12 month sample	10.89	31	11.99	32	12.13	33	12.27	33
Main 4 year sample	3.68	19	3.88	19	3.76	19	3.63	19
Observations	2677143		2539492		2619222		2716094	

Note: Percentage of the 5% sample that fit the exclusion restriction and reason. An additional exclusion (not shown) was for individuals whose demographics or plan detail data were undefined or unavailable.

Further, beneficiaries may not have the same plan through the entire year if it is their initial year of enrollment, since patients are eligible to enroll for 7 months around their birthdays (three months before and after and including their birthday month).[1] Beneficiaries who died may also pass away mid-year. Beneficiaries who switch from Medicare Advantage (prescription coverage) to Medicare Part D, may also not be in plans for the full 12 months since the Medicare Advantage disenrollment period occurs within a year January 1-February 14, taking effect the

[1]If enrolled in the initial enrollment period, patients would be enrolled on the first day of their birthday month if enrolled prior. If sign-up occurs in the birthday month or in the three months after, coverage start is delayed 1-3 months after enrollment.

first month after disenrollment (either February or March). Enrollees may also have switched plans mid-year under the special enrollment periods: after moving; losing current Medicaid/Employer/etc coverage; new creditable coverage options, specific plan changes, and other special circumstances.[2]

Appendix Table A.2 includes demographic summary statistics for the entire 5% Medicare sample. These individuals in the entire 5% sample do not necessarily have Medicare Part D or meet the sample criteria as shown in Appendix Table A.1. Appendix Table A.3 includes demographic summary statistics for the "full" sample of Medicare Part D individuals who met the exclusion criteria and were in each plan for the full year. The individuals in the full sample are older than the average Medicare Part D enrollee and are more white. They also have higher rates of chronic diseases. The differences from the 5% sample and the "full" sample reflect the fact that older and sicker people are more likely to select into Medicare Part D. The differences in race between the two samples may also have to do with who qualifies for Medicare (and is thus excluded).

[2]https://www.medicare.gov/sign-up-change-plans/when-can-i-join-a-health-or-drug-plan/special-circumstances/join-plan-special-circumstances.html

Table A.2: 5% Sample of Medicare beneficiary demographics 2009-2012

	(1) 2009		(2) 2010		(3) 2011		(4) 2012	
	mean	sd	mean	sd	mean	sd	mean	sd
Age	72.07	12.49	71.56	12.51	71.43	12.51	71.26	12.35
Female	0.58	0.49	0.56	0.50	0.55	0.50	0.55	0.50
Race: Undefined	0.00	0.05	0.00	0.06	0.00	0.07	0.01	0.08
White	0.83	0.37	0.83	0.38	0.82	0.38	0.82	0.39
Black	0.10	0.30	0.10	0.30	0.10	0.30	0.10	0.31
Other	0.02	0.13	0.02	0.14	0.02	0.14	0.02	0.14
Asian	0.03	0.13	0.02	0.14	0.02	0.14	0.02	0.14
Hispanic	0.02	0.15	0.03	0.16	0.03	0.16	0.03	0.16
N. American Native	0.00	0.07	0.00	0.07	0.00	0.07	0.00	0.07
Observations	2,677,135		2,539,486		2,619,219		2,716,093	

	(1) 2009		(2) 2010		(3) 2011		(4) 2012	
	mean	sd	mean	sd	mean	sd	mean	sd
2011 RxHCC weight	0.38	0.43	0.39	0.44	0.40	0.44	0.40	0.45
2011 RxHCC demo. weight	0.49	0.19	0.49	0.19	0.49	0.19	0.49	0.19
Diabetes	0.21	0.41	0.21	0.41	0.21	0.41	0.21	0.41
Hypertension	0.42	0.49	0.41	0.49	0.41	0.49	0.41	0.49
Cancer	0.06	0.25	0.06	0.24	0.06	0.24	0.06	0.24
High Cholesterol	0.43	0.49	0.43	0.49	0.43	0.49	0.43	0.49
Long Term Care	0.05	0.22	0.05	0.22	0.05	0.22	0.050	0.22
Observations	2,318,357		2,381,965		2,430,325		2,507,147	

Note: The age is the age at the end of the reference year. The total number of observations in the first panel differ from Table A.1 due to missing values in gender and race fields. The number of observations differ between the two panels since the information on risk factors and conditions are generated from the Medicare Part A and B claims from the prior year. Patients who did not have relevant claims to be scored from the prior year did not have values for these conditions.

Table A.3: Full sample of medicare beneficiary demographics 2009-2012

	(1) 2009		(2) 2010		(3) 2011		(4) 2012	
	mean	sd	mean	sd	mean	sd	mean	sd
Age at End of Reference Year	75.79	7.18	75.97	7.32	75.95	7.30	75.81	7.30
Start Medicare Year	1998.3	7.2	1999.0	7.3	2000.1	7.3	2001.2	7.3
Years in Medicare	10.7	7.2	10.9	7.3	10.9	7.3	10.8	7.3
Female	0.65	0.48	0.65	0.48	0.64	0.48	0.64	0.48
Race: White	0.95	0.22	0.95	0.22	0.95	0.22	0.95	0.23
Race: Black	0.03	0.17	0.04	0.16	0.03	0.17	0.03	0.17
Race: Other	0.01	0.10	0.01	0.10	0.01	0.10	0.01	0.11
Race: Asian	0.01	0.08	0.01	0.08	0.01	0.08	0.01	0.08
Race: Hispanic	0.00	0.07	0.00	0.06	0.00	0.06	0.00	0.06
Race: North American Native	0.00	0.04	0.00	0.04	0.00	0.04	0.00	0.04
2011 RxHCC weight	0.45	0.32	0.47	0.33	0.48	0.33	0.48	0.33
2011 RxHCC demo. weight	0.42	0.01	0.42	0.01	0.42	0.01	0.42	0.01
Diabetes	0.23	0.42	0.24	0.42	0.24	0.43	0.25	0.43
Hypertension	0.61	0.49	0.62	0.49	0.62	0.48	0.62	0.48
Cancer	0.09	0.29	0.10	0.29	0.10	0.30	0.10	0.30
High Cholesterol	0.66	0.47	0.67	0.47	0.68	0.47	0.68	0.47
Baseline 4 Year Sample	0.34	0.47	0.32	0.47	0.31	0.46	0.30	0.46
Observations	291,550		304,477		317,670		333,309	

A.1.1 Standard Plans

Appendix Tables A.4, A.5, and A.6 present summary statistics on individuals who selected into plans with the standard Medicare-defined plan (with the standard deductible) in all four years.

Table A.4: Average Coinsurance in Phases in Baseline Sample Standard Plans 2009-2012 (person-week)

	(1) 2009 mean sd	count	(2) 2010 mean sd	count	(3) 2011 mean sd	count	(4) 2012 mean sd	count
Deductible	0.78	152,348	0.82	158,028	0.80	159,524	0.85	174,615
	0.30		0.26		0.28		0.24	
ICR	0.22	253,123	0.20	248,750	0.21	243,824	0.23	231,479
	0.13		0.13		0.12		0.13	
Gap	0.68	34,249	0.54	33,136	0.36	35,389	0.38	33,197
	0.40		0.42		0.24		0.25	
Catastrophic	0.06	4,336	0.05	4,397	0.05	5,428	0.05	4,967
	0.08		0.04		0.03		0.03	
Total	0.45	444,056	0.44	444,311	0.43	444,165	0.48	444,258
	0.36		0.36		0.35		0.36	

Note: Table is generated from the baseline sample individuals with standard deductible, ICL, and OOPT limits. The coinsurance rates are averaged over the amount the patient pays (does not include the drug manufacture discounts in 2011 and 2012) divided by the total expenditure cost in the person-week observation where spending occurs. The count reflects the fact that there are more person-week observations in the ICR region than others. These sums do not reflect the counterfactual coinsurance rates that beneficiaries with low spending would have faced if they had reached higher spending.

Table A.5: Proportion of Beneficiaries in Each Phase at the End of the Year in Baseline Sample Standard Plans 2009-2012

	Standard			
	2009	2010	2011	2012
Deductible	11.26	12.97	13.41	16.55
ICR	61.46	60.68	59.26	58.72
Gap	22.76	22.23	22.28	20.41
Catastrophic	4.52	4.13	5.04	4.33
Observations	9,178	9,178	9,178	9,178

Note: Table is generated from the baseline sample individuals with standard limits. The proportion of beneficiaries that end the year in each phase is averaged over the individual beneficiary.

Table A.6: Average Probability of Weekly Spending in Coverage Regions in Standard Plans 2009-2012 (person-week)

	2009	2010	2011	2012
	mean/sd	mean/sd	mean/sd	mean/sd
Deductible	29.08	29.83	30.16	30.18
	45.41	45.75	45.90	45.90
ICR	40.04	40.94	41.13	42.22
	49.00	49.17	49.21	49.39
Gap	47.85	48.50	48.77	51.60
	49.95	48.98	49.99	49.98
Catastrophic	55.74	56.01	56.85	57.34
	49.67	49.64	49.53	49.46
Total	37.05	37.72	38.01	38.37
	48.29	48.47	48.54	48.63

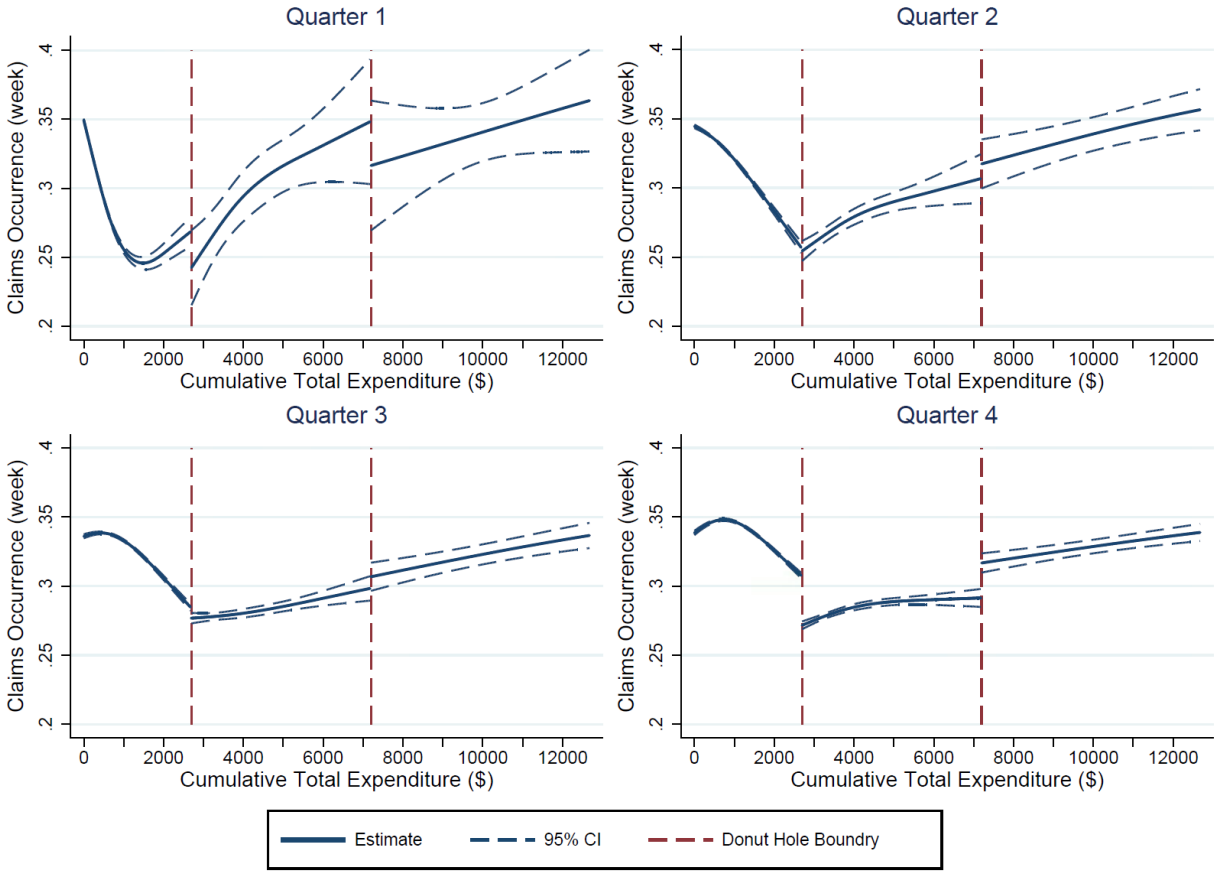
Note: Table is generated from the baseline sample individuals with standard deductible, ICL, and OOPT limits. Table displays the raw probability of spending in a week in each coverage region and year.

A.2 Alternative Model Results

Appendix Figures A.2 and A.3 present the results from the falsification specifications False 1 and 2. The main takeaway from these falsification tests is that even under different specifications for the spline with four and three knots at different locations, the findings from the main specification still hold. It appears that the

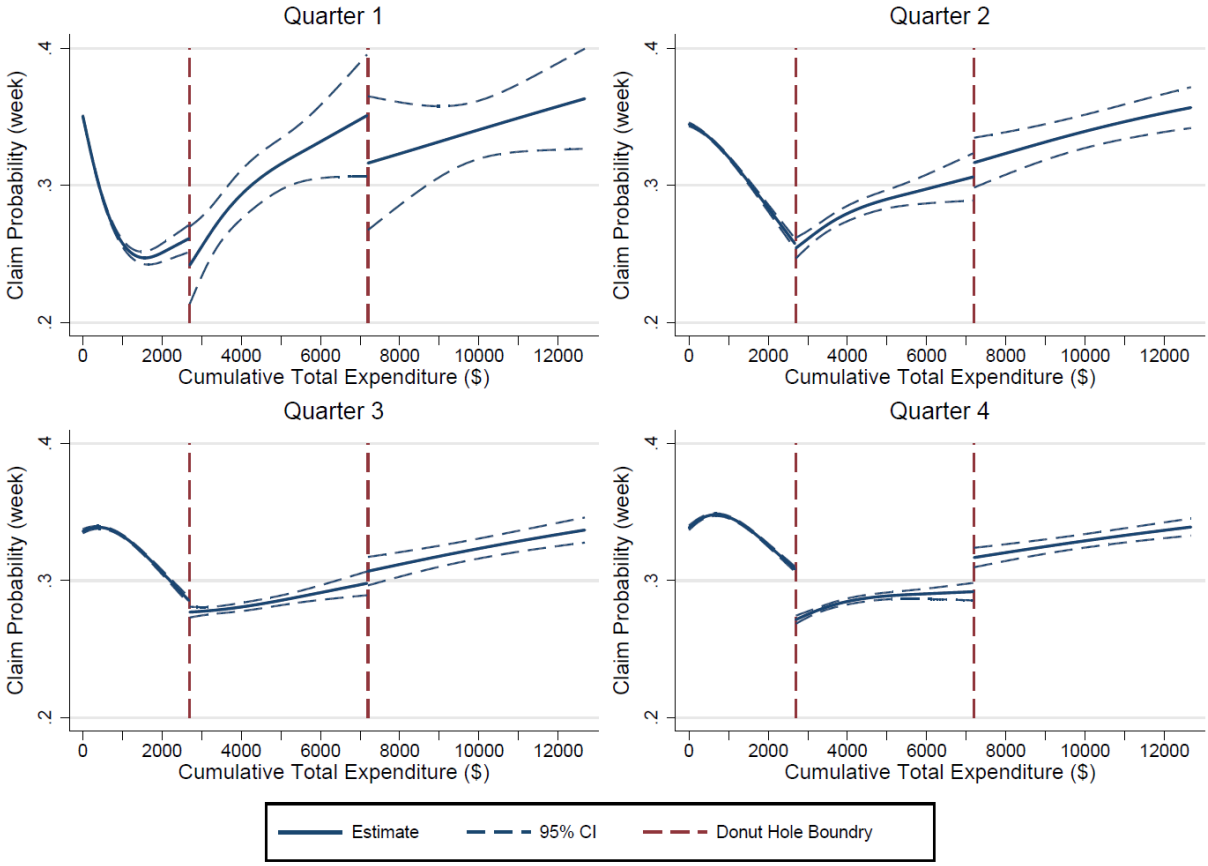
probability of claims occurring decreases far in advance of individuals entering the coverage gap and there are no sharp discontinuities at the kinks (except at the end of the year when they are expected).

Figure A.2: False 1: Probability of Claims Occurring in a Week



Note: Displays the predicted values of the claim occurrence probability $\hat{o}_{iyw}$ in a week from the fixed-effects panel regression of Equation 1.5 under the assumptions of False 1 on the beneficiaries in the No Deductible sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values. The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. See Figure 1.10 for the exact values used to generate the prediction.

Figure A.3: False 2: Probability of Claims Occurring in a Week



Note: Displays the predicted values of the claim occurrence probability $\hat{o}_{iyw}$ in a week from the fixed-effects panel regression of Equation 1.5 under the assumptions of False 2 on the beneficiaries in the No Deductible sample. Each panel represents a quarter of the year where a quarter consists of 13 weeks except for quarter 4, where the last "week" of the year consists of the remainder 8 or 9 days of the year. Images display a 95% confidence interval around the predicted values. The predicted values are generated within each panel by holding all variables constant except for the cumulative total expenditure displayed on the x-axis and the cumulative out-of-pocket expenditure. See Figure 1.10 for the exact values used to generate the prediction.

A.3 Dynamic Panel with Fixed Effects Bias Simulated

Simulated data was created to have a claims occurrence value o_{iyw} and a cumulative total spending value Z_{iyw} with 208 weeks. The data generating process allows for individual heterogeneity in a base probability of spending. For the simulated case, the probability of spending and the amount spent is independent of the Z_{iyw} and region effects R_{iyw}.

The following 3 cases are graphed in Figure A.4. They show the shape the bias would have on the cumulative total spending measure.

T=208

$$o_{iyw} = \alpha_i + R_{iyw} + Z_{iyw} + \beta_r * Z_{iyw} + q_{iyw} + \tau_y + \varepsilon_{it}$$

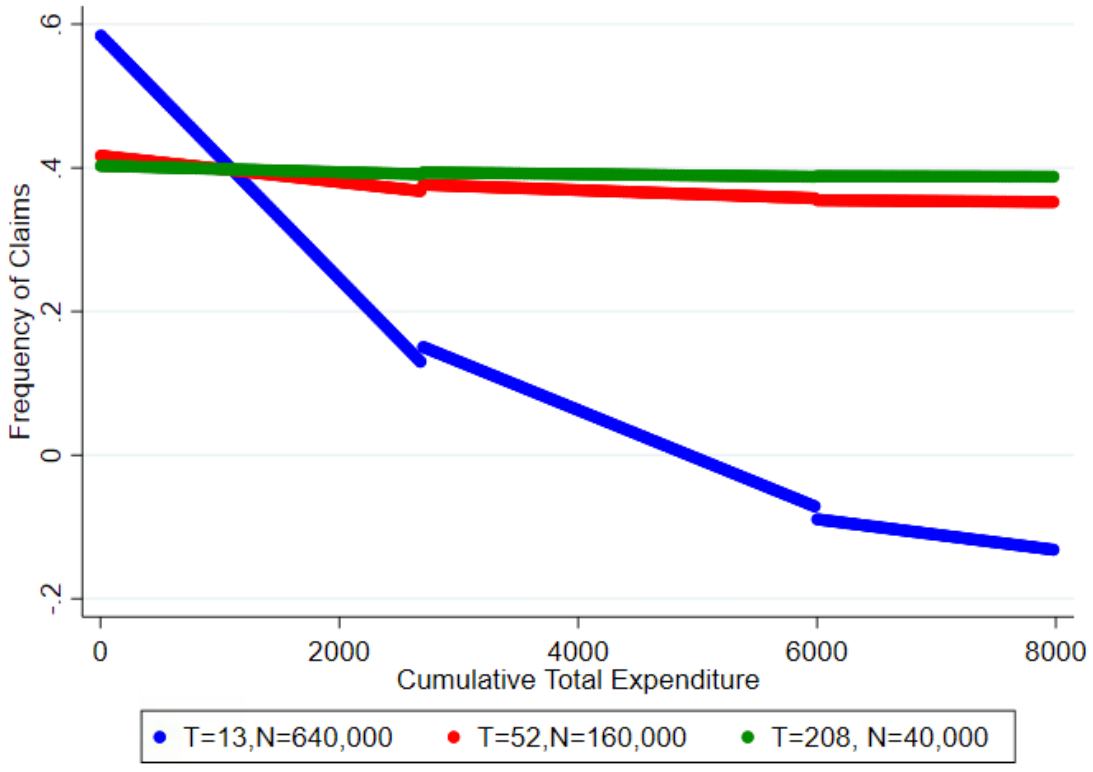
T=52

$$o_{iyw} = \alpha_{iy} + R_{iyw} + Z_{iyw} + \beta_r * Z_{iyw} + q_{iyw} + \varepsilon_{it}$$

T=13

$$o_{iyw} = \alpha_{iyw} + R_{iyw} + Z_{iyw} + \beta_r * Z_{iyw} + \varepsilon_{it}$$

Figure A.4: Estimation of a linear model of simulated data with different time frames



Note: The green- T=13, red - T=52, blue - T=208

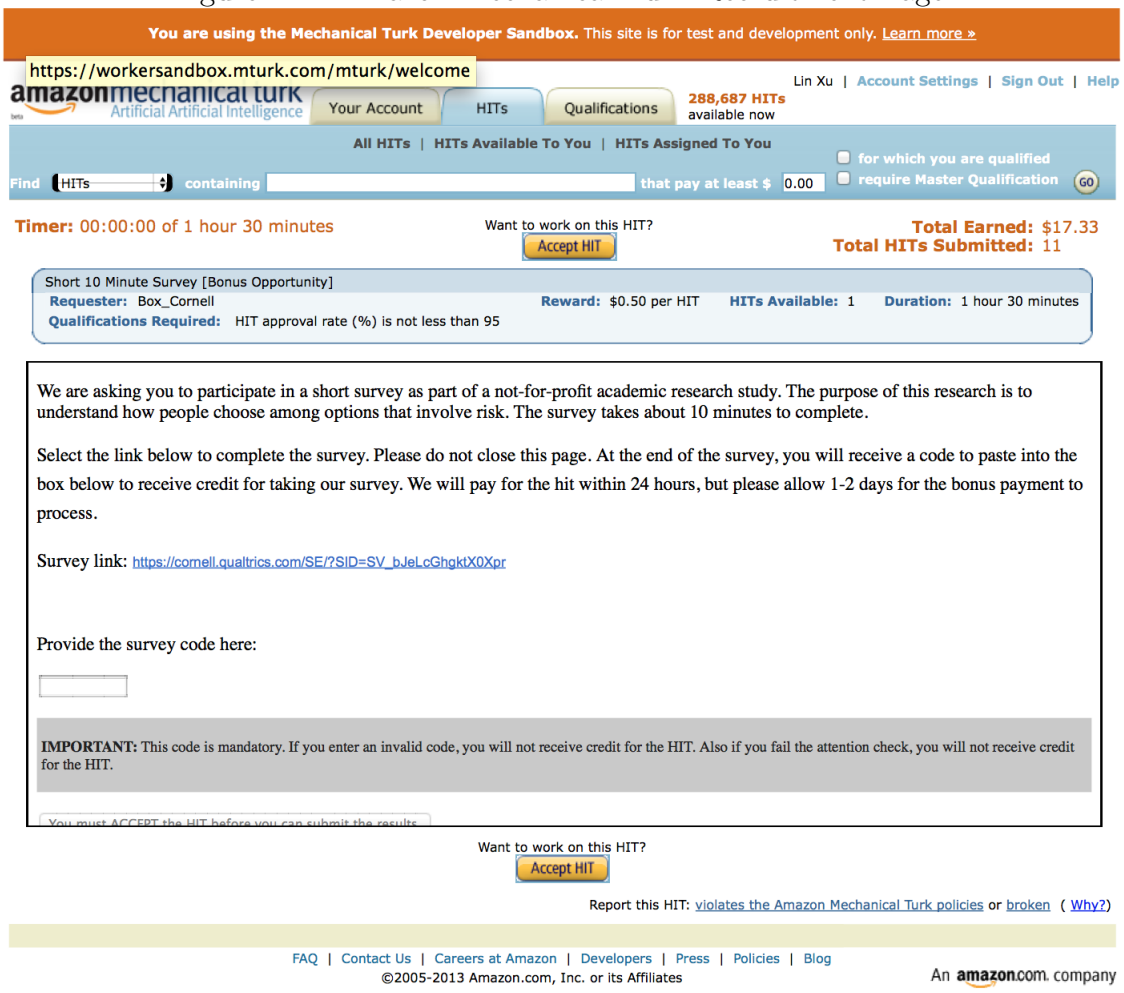
APPENDIX B

APPENDIX OF CHAPTER 2: THE NATURE OF RISK PREFERENCES: EVIDENCE FROM EXPERIMENTAL INSURANCE CHOICES

B.1 Experimental Materials

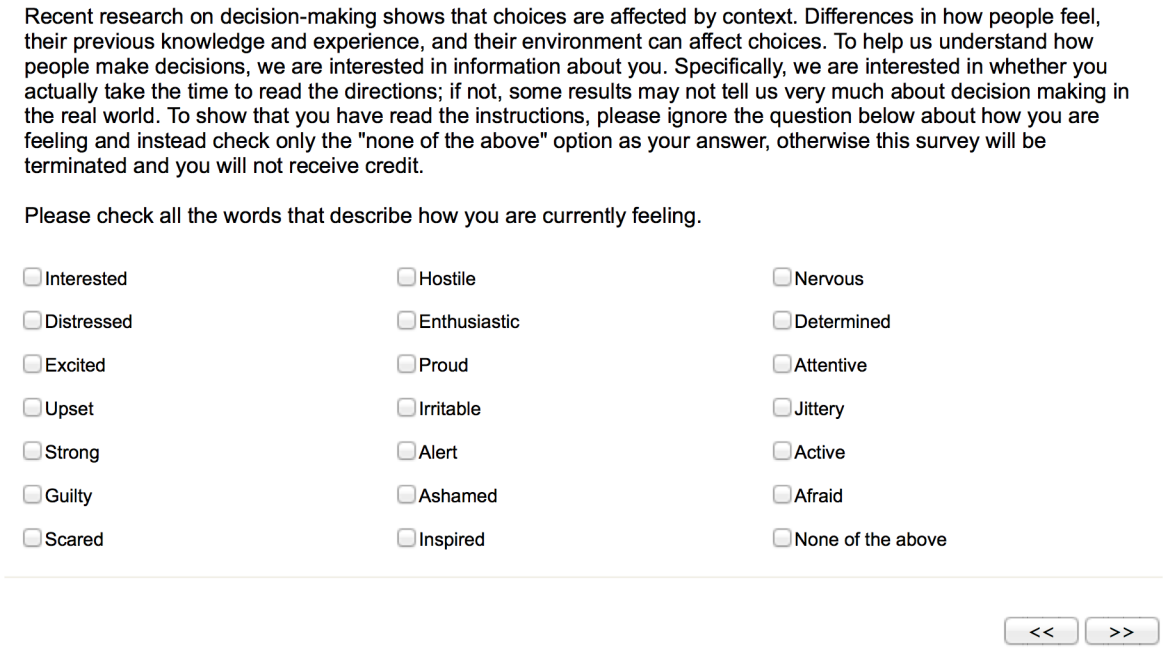
Figure B.1 includes a screenshot of the recruitment page for the Amazon Mechanical Turk workers. Figure B.2 and B.3 include screenshots of the attention check and an example verification check that both of the MTurk and student subjects were required to pass prior to proceeding through the experiment.

Figure B.1: Amazon Mechanical Turk Recruitment Page



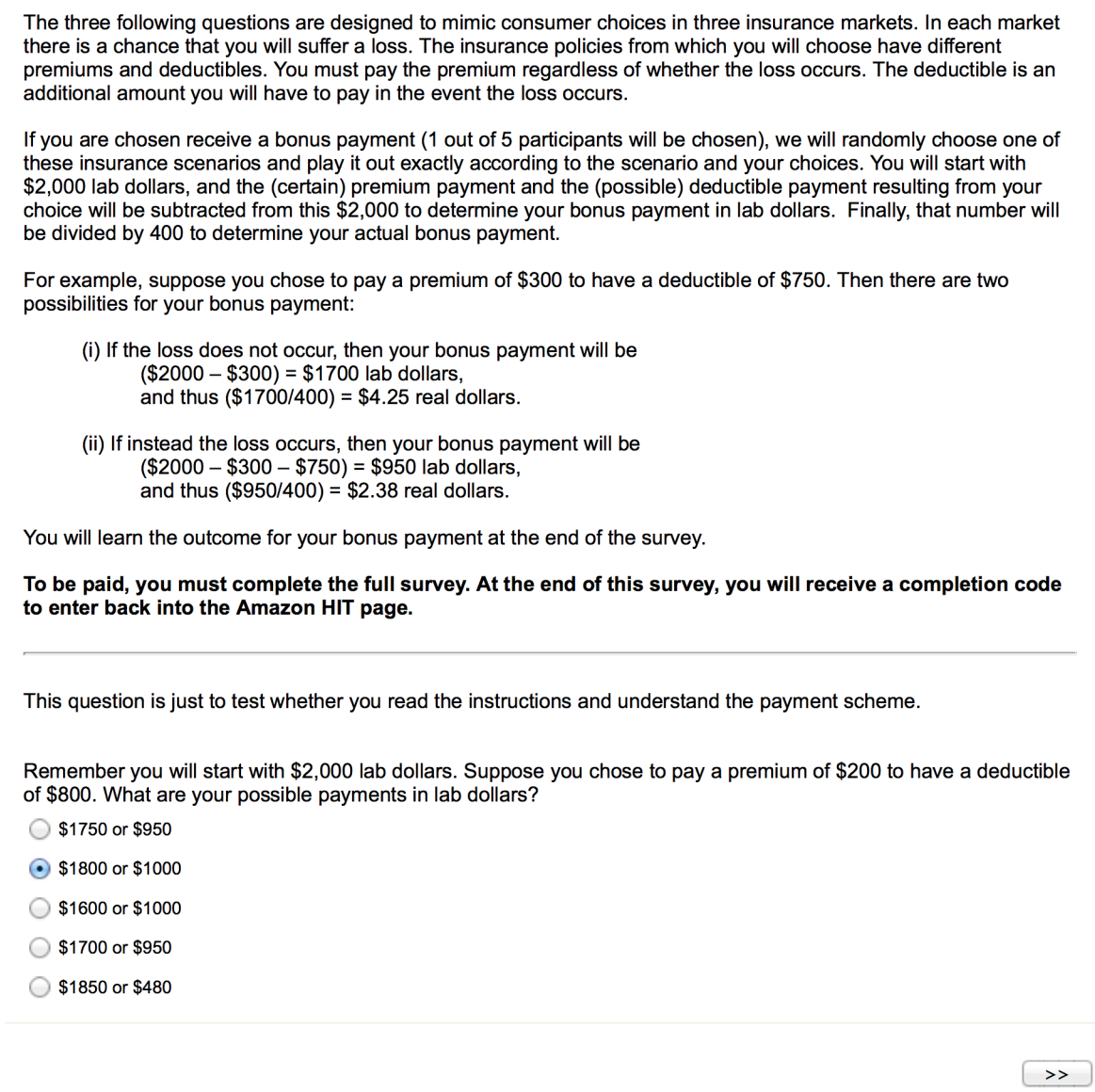
Note: The screenshot is taken of the Mechanical Turk Developer Sandbox platform for testing out survey recruitment prior to implementation. The Sandbox environment is meant to replicate the actual text and functionality that the workers experience.

Figure B.2: Attention Check in Web Experiment



Note: If subjects fail this attention check once, they receive a warning. If they fail it twice, the survey ends and they are not able to proceed. Both MTurk and student subjects received the exact same attention check.

Figure B.3: Verification Question on Payout Scheme in Web Experiment



Note: Image is from the MTurk version of the experiment. The student version divides the payout by 50 and does not include Amazon Hit wording. Wording for other insurance contexts are similar.

B.2 Data summary by subject group

Table B.1 and B.2 include the summary statistics for the deductible choices broken down by the treatment group for the MTurk and student subjects, respectively.

Table B.1: MTurk subjects: Summary of deductible choices between the three treatments

	Abstract			Ambiguity			Insurance		
Deductible	Auto Coll	Auto Comp	Home	Auto Coll	Auto Comp	Home	Auto Coll	Auto Comp	Home
$50		15.7			14.1			8.8	
$100	19.2	9.9	14.7	21.2	17.5	14.1	11.9	9.4	9.1
$200	16.7	14.4		13.1	17.2		11.6	13.5	
$250	25.6	15.4	38.5	35.9	18.1	40.6	27.0	11.3	31.8
$500	18.6	12.8	22.4	19.1	15.0	31.6	29.9	23.0	27.7
$1000	19.9	31.7	24.4	10.6	18.1	13.8	19.5	34.0	31.4
Num obs.		312			320			318	

Note: Values are percent of subjects.

Table B.2: Student subjects: Summary of deductible choices

	Abstract			Ambiguity			Insurance		
Ambiguity Deductible	Auto Coll	Auto Comp	Home	Auto Coll	Auto Comp	Home	Auto Coll	Auto Comp	Home
$50		11.1			9.8			6.8	
$100	15.6	16.0	9.4	15.6	16.5	10.7	13.1	13.6	7.2
$200	13.9	14.3		13.8	22.3		13.6	14.5	
$250	37.7	13.1	53.7	45.1	13.4	52.7	36.7	15.4	47.5
$500	18.0	21.7	19.7	18.3	21.9	26.8	24.4	24.9	25.8
$1000	14.8	23.8	17.2	7.1	16.1	9.8	12.2	24.9	19.5
Num obs.		244			224			221	

Note: Values are percent of subjects.

B.3 Reduced-form analysis of subject groups

Appendix Table B.3 Model I presents the results from Table 2.6, and Model II presents the estimates from an additional specification. Model II is still defined by Equation 2.1, but the X_j independent variables that pertain to insurance context j include those from the original specification (price for the $500 deductible plan,

the average claim rate that the subjects were given, the treatment group, and the subject pool), and also variables controlling for the interaction between the treatment and subject group. Comparing subject groups with just a indicator

Table B.3: Impact of treatment and student group on deductible choices

	I			II		
	Comp	Coll	Home	Comp	Coll	Home
Price of $500 deduc. (per $100)	0.20***	0.22***	0.049***	0.21***	0.22***	0.049***
	(0.057)	(0.035)	(0.015)	(0.057)	(0.035)	(0.015)
Claim rate (per %)	-0.13***	-0.053***	-0.044***	-0.13***	-0.053***	-0.044***
	(0.030)	(0.014)	(0.011)	(0.030)	(0.014)	(0.011)
Ambiguity treatment	-0.37***	-0.32***	-0.36***	-0.46***	-0.42***	-0.44***
	(0.064)	(0.064)	(0.066)	(0.084)	(0.084)	(0.086)
Abstract treatment	-0.17**	-0.17**	-0.24***	-0.22*	-0.25**	-0.29***
	(0.064)	(0.064)	(0.065)	(0.085)	(0.084)	(0.086)
Student	-0.024	-0.060	-0.15**	-0.13	-0.21*	-0.26**
	(0.053)	(0.053)	(0.054)	(0.092)	(0.092)	(0.095)
Ambiguity X student				0.22	0.26*	0.20
				(0.13)	(0.13)	(0.13)
Abstract X student				0.10	0.20	0.13
				(0.13)	(0.13)	(0.13)
κ_{J1}	-1.48***	-1.18***	-1.55***	-1.52***	-1.24***	-1.60***
	(0.092)	(0.11)	(0.13)	(0.098)	(0.11)	(0.13)
κ_{J2}	-0.94***	-0.71***	-0.21	-0.98***	-0.77***	-0.25
	(0.089)	(0.11)	(0.13)	(0.095)	(0.11)	(0.13)
κ_{J3}	-0.49***	0.18	0.55***	-0.53***	0.13	0.51***
	(0.088)	(0.10)	(0.13)	(0.093)	(0.11)	(0.13)
κ_{J4}	-0.11	0.90***		-0.16	0.85***	
	(0.088)	(0.11)		(0.093)	(0.11)	
κ_{J5}	0.43***			0.39***		
	(0.088)			(0.093)		
Observations	1639					

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: The deductible prices, the claim rates, and the estimated coefficients are specific to each insurance context: home, auto collision, or auto comprehensive insurance.

variable for students, student behavior was not significantly different from Amazon Mechanical Turk (MTurk) behavior in our auto insurance contexts, but they were significantly different in the home context. Students were more likely to choose lower deductibles (less risky).

When we analyze the interactions between being a student and being assigned to one of the treatment groups, we find that while on average the direction of the students' and MTurk workers' responses are in the same direction, there is a difference in the magnitudes of their responses. The MTurk workers who are in any treatment are more likely to choose lower deductibles than students who are in the same treatment but only significantly in the collision and home insurance contexts at the 5% and 1% level, respectively. Students who are in the Ambiguity treatment are more likely to choose lower deductibles than students in the insurance treatment. These are one sided significant at less than the 1% level for comprehensive and home, and at the 5% level for collision. (Two sided significant at the 5% level for comp and home, and 10% for collision). Abstract students are more likely to choose lower deductibles than students in the insurance treatment, but these are not significant in any insurance context.

We analyze the impact of student demographics on their choices in Table B.4 Model II. This model is again defined by Equation 2.1 over a sample of only the student subjects. The X_j independent variables include: price for the \$500 deductible plan, the average claim rate that the subjects were given, the treatment group, and subject demographic characteristics. Fewer of the coefficients are significant in part because of the smaller sample size. It is still notable that females were significantly more likely to choose conservative deductibles. Economics majors were significantly more likely to choose less conservative deductibles, which are more actuarially fair according to the given claim rate.[1]

[1]Biology students were the baseline. While there are double majors, there are only six Biology students who were double majors with either Economics or Government.

Table B.4: Impact of treatment and observables on student deductible choice

	I			II		
	Comp	Coll	Home	Comp	Coll	Home
Price of $500 deduc. (per $100)	0.20***	0.22***	0.049***	0.21*	0.26***	0.053*
	(0.057)	(0.035)	(0.015)	(0.095)	(0.058)	(0.023)
Claim rate (per %)	-0.13***	-0.053***	-0.044***	-0.064	-0.079***	-0.063***
	(0.030)	(0.014)	(0.011)	(0.046)	(0.021)	(0.017)
Ambiguity treatment	-0.37***	-0.32***	-0.36***	-0.27**	-0.19	-0.26*
	(0.064)	(0.064)	(0.066)	(0.10)	(0.10)	(0.10)
Abstract treatment	-0.17**	-0.17**	-0.24***	-0.14	-0.083	-0.19
	(0.064)	(0.064)	(0.065)	(0.099)	(0.099)	(0.10)
Student	-0.024	-0.060	-0.15**			
	(0.053)	(0.053)	(0.054)			
Female				-0.37***	-0.29***	-0.36***
				(0.083)	(0.083)	(0.087)
Graduation year				-0.11**	-0.082*	-0.078
				(0.040)	(0.040)	(0.042)
Econ				0.091	0.26**	0.18
				(0.097)	(0.097)	(0.10)
Govt				-0.013	0.14	-0.16
				(0.13)	(0.13)	(0.13)
κ_{J1}	-1.48***	-1.18***	-1.55***	-221.5**	-166.3*	-158.2
	(0.092)	(0.11)	(0.13)	(80.5)	(81.5)	(84.6)
κ_{J2}	-0.94***	-0.71***	-0.21	-220.9**	-165.8*	-156.5
	(0.089)	(0.11)	(0.13)	(80.5)	(81.5)	(84.6)
κ_{J3}	-0.49***	0.18	0.55***	-220.4**	-164.7*	-155.7
	(0.088)	(0.10)	(0.13)	(80.5)	(81.5)	(84.6)
κ_{J4}	-0.11	0.90***		-220.0**	-163.9*	
	(0.088)	(0.11)		(80.5)	(81.5)	
κ_{J5}	0.43***			-219.3**		
	(0.088)			(80.5)		
Observations	1,639			687		

Note: The deductible prices, the claim rates, and the estimated coefficients are specific to each insurance context: home, auto collision, or auto comprehensive insurance. While we observed the choices of 689 students, two were omitted from the Model II results here because of missing data on gender and graduation year, respectively.

B.4 More on Probability Weighting

To our knowledge, there has been an extensive amount of theoretical and experimental work to estimate risk preferences and the probability weighting function. However, ours will be the first experimental piece to estimate

the probability weighting function from subject choices over menus (similar to actual insurance data) rather than elicited certainty equivalents. Tversky and Kahneman (1992) suggest a functional form for probability weighting of $\pi(p) = \frac{p^\gamma}{[p^\gamma+(1-p)^\gamma]^{1/\gamma}}$for some $\gamma \in (0.279, 1)$ while Prelec (1998) suggests $\pi(p) = \exp(-(-\ln(p))^\alpha)$for some $\alpha \in (0, 1)$. Gonzalez and Wu (1999) present a discussion on the shape of the probability weighting function along with a nonparametric estimation algorithm. However, their data is collected from a total of 10 participants who were asked for their certainty equivalents over multiple lotteries. Bruhin et al. (2010) also estimate the parameters of a mixture model of risk preferences based on experimental data, and while their experiments included questions framed in the insurance context, they elicited certainty equivalents. In addition, they assume an inverse-S shaped functional form on the probability weighting function using the two-parameter specification suggested by Goldstein and Einhorn(1987) and Lattimore, Baker, and Witte (1992).

B.5 Model 1: Homogenous r and $\Omega(\mu)$ for Ambiguity and Abstract treatments

Table B.5 and Figure B.4 present the estimates from Model 1a and 1b for the Ambiguity treatment. Similarly, Table B.6 and Figure B.5 present the estimates from Model 1a and 1b for the Abstract treatment.

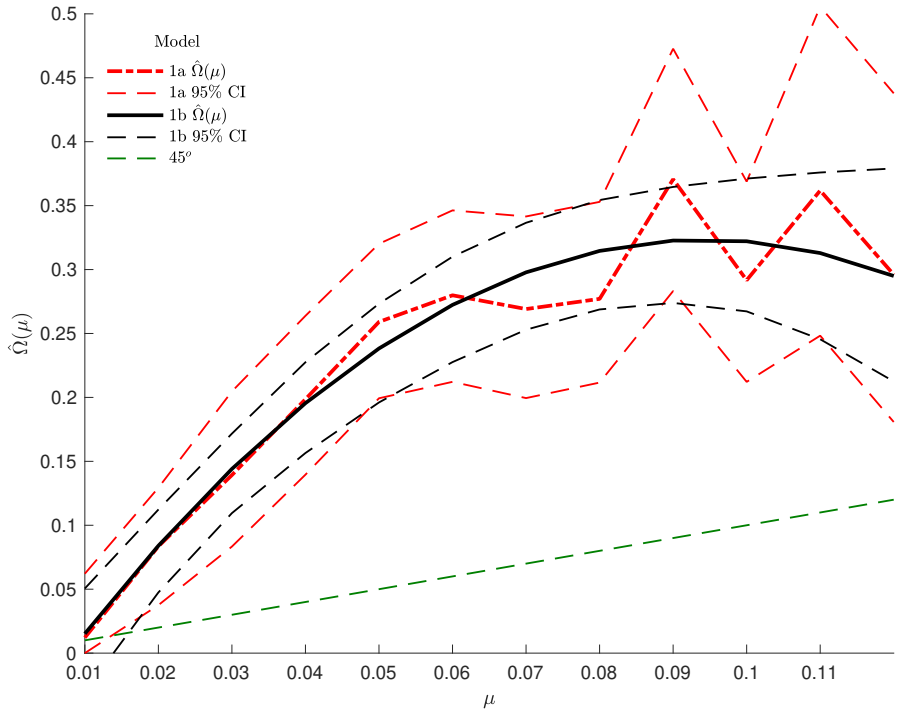
The estimates from the Ambiguity and Abstract treatment are similar to the baseline Insurance treatment in that the estimated risk aversion parameter is small in all cases (less than 0.00024), there are significant probability distortions, and the scale parameter σ is large relative to estimates of risk preferences from other

Table B.5: Model 1a and 1b for Ambiguity

	Model 1a			Model 1b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	89.6	72.8	125.7	90.0	74.1	134.6
r	4.6e-05	7.8e-12	0.00058	3.6e-05	1.2e-10	0.00057

Note: 544 households in the Insurance treatment. See additional notes in Figure B.4 on how estimates are constructed.

Figure B.4: Model 1a and 1b probability weighting functions in Ambiguity treatment



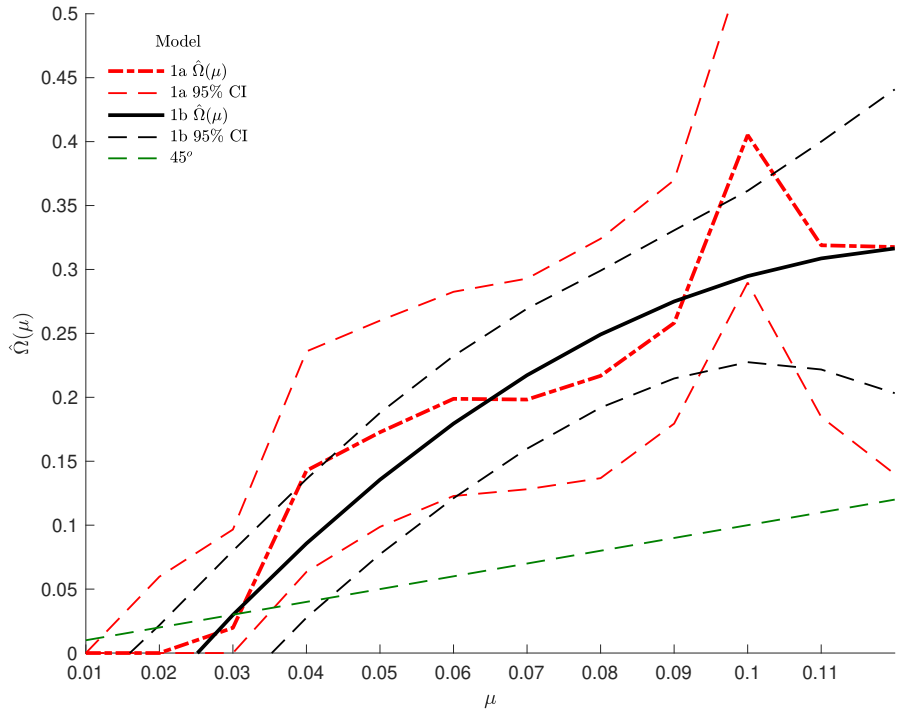
Note: This figure displays the estimated $\Omega(\mu)$ from Model 1a and Model 1b across the 544 subjects in the Ambiguity treatment. The 95% confidence intervals from these estimated values are constructed with 1,000 bootstraps with a subsample of 80 percent of 544 with replacement. The bootstrap was conducted with a subsample due to estimated r, $\omega_{0.01}$ and $\omega_{0.02}$ being close to or at the lower bound of 0.

Table B.6: Model 1a and 1b for Abstract

	Model 1a			Model 1b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	134.3	102.5	166.9	174.2	115.4	504.7
r	6e-14	6e-14	6e-14	0.00024	1.2e-11	0.0028

Note: 556 households in the Insurance treatment. See additional notes in Figure B.5 on how estimates are constructed.

Figure B.5: Model 1a and 1b probability weighting functions in Abstract treatment



Note: This figure displays the estimated $\Omega(\mu)$ from Model 1a and Model 1b across the 566 subjects in the Abstract treatment. The 95% confidence intervals from these estimated values are constructed with 1,000 bootstraps with a subsample of 80 percent of 556 with replacement. The bootstrap was conducted with a subsample due to estimated r, $\omega_{0.01}$ and $\omega_{0.02}$ being close to or at the lower bound of 0.

specifications. The $\hat{\Omega}(\mu)$ estimates are overall slightly higher than those in the Model 1 Insurance treatment. In the Abstract treatment, the estimate $\hat{\omega}_{0.01}$ does still indicate probability underweighting at extremely low probabilities. In the Ambiguity treatment, the estimate $\hat{\omega}_{0.01}$ does not indicate probability underweighting, but the confidence interval around the estimate does not reject underweighting nor overweighting.

As with the Insurance treatment, the estimates of the probability distortion functions using Model 1 a/b are not robust to other specifications.

B.6 Additional tables and graphs for models with heterogenous risk preferences: Model 2, 3, and 4

Table B.7, Figure B.6, and Figure B.8 display the estimates and heterogeneity from Model 2, 3, and 4a/b for the subjects in the Ambiguity treatment. Similarly, Table B.8, Figure B.7, and Figure B.9 display the estimates from Model 2, 3, and 4a/b for the Abstract treatment.

The qualitative results from these models are similar to the results presented for the Insurance treatment. The risk aversion parameter is estimated to be negligible throughout all of the Model 2 and 3 results with the probability distortions functions (and heterogeneity in these functions) driving the apparent differences in choices.

Table B.7: Model 2, 3, and 4 for Ambiguity

	Model 2a			Model 2b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	39.13	32.53	45.53	39.83	33.50	47.34
r_i 10^{th} percentile	6.0e-26			2.2e-22		
$median(r_i)$	5.3e-11	5.3e-11	5.3e-11	1.0e-08	9.8e-09	1.2e-08
r_i 90^{th} percentile	1.1e-10			2.1e-08		
Φ_Ω	2.63	2.23	3.10	2.61	2.24	3.08
$\Phi_{\Omega,r}$	-0.01	-28.02	23.39	-64.95	-74.55	-3.3e-03
Implied corr$(\xi_{r,i}, \xi_{\Omega,i})$	-1.1e-04	-0.38	0.36	-0.99	-0.99	-4.5e-05

	Model 3a			Model 3b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	39.13	32.58	45.53	39.83	33.51	47.34
r	1.8e-12	1.8e-12	1.8e-12	2.4e-11	2.4e-11	2.4e-11
Φ_Ω	2.63	2.23	3.08	2.62	2.24	3.06

	Model 4a			Model 4b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	39.13	33.50	44.94	39.83	34.39	46.17
Φ_Ω	2.63	2.26	3.04	2.62	2.25	3.03

Note: 544 subjects in the Ambiguity treatment. See additional notes in Figure B.6 on how estimates are constructed.

Table B.8: Model 2, 3, and 4 for Abstract

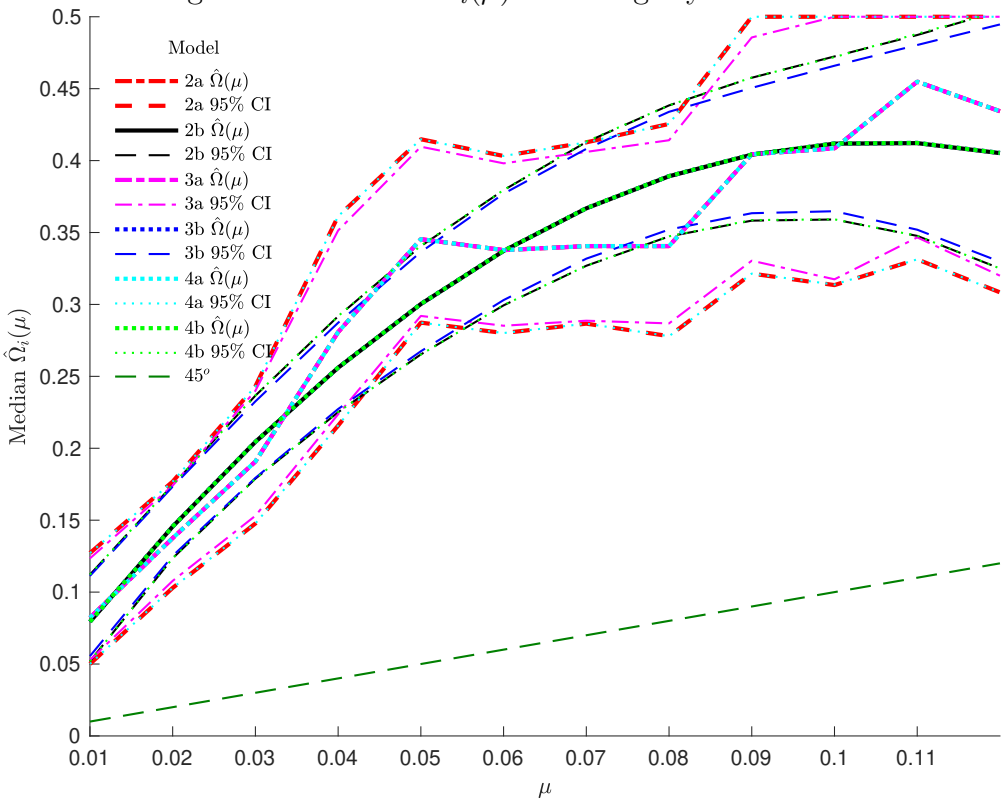
	Model 2a			Model 2b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	44.29	36.69	53.12	45.18	37.87	54.95
r_i 10^{th} percentile	2.8e-15			7.4e-23		
$median(r_i)$	3.8e-12	3.8e-12	3.8e-12	4.6e-15	4.6e-15	4.6e-15
r_i 90^{th} percentile	7.6e-12			9.2e-15		
Φ_Ω	4.93	3.86	7.30	4.99	3.90	16.16
$\Phi_{\Omega,r}$	27.17	18.87	42.28	64.96	0.00	97.34
Implied corr$(\xi_{r,i}, \xi_{\Omega,i})$	0.89	0.71	0.99	0.89	0.00	0.99

	Model 3a			Model 3b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	45.34	36.68	56.54	46.47	38.08	58.48
r	9.0e-09	8.1e-09	9.0e-09	2.7e-10	2.6e-10	2.7e-10
Φ_Ω	5.06	3.74	57.66	5.20	3.78	94.00

	Model 4a			Model 4b		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
σ	45.33	37.41	56.54	46.47	38.76	57.82
Φ_Ω	5.06	3.82	55.74	5.20	3.87	78.35

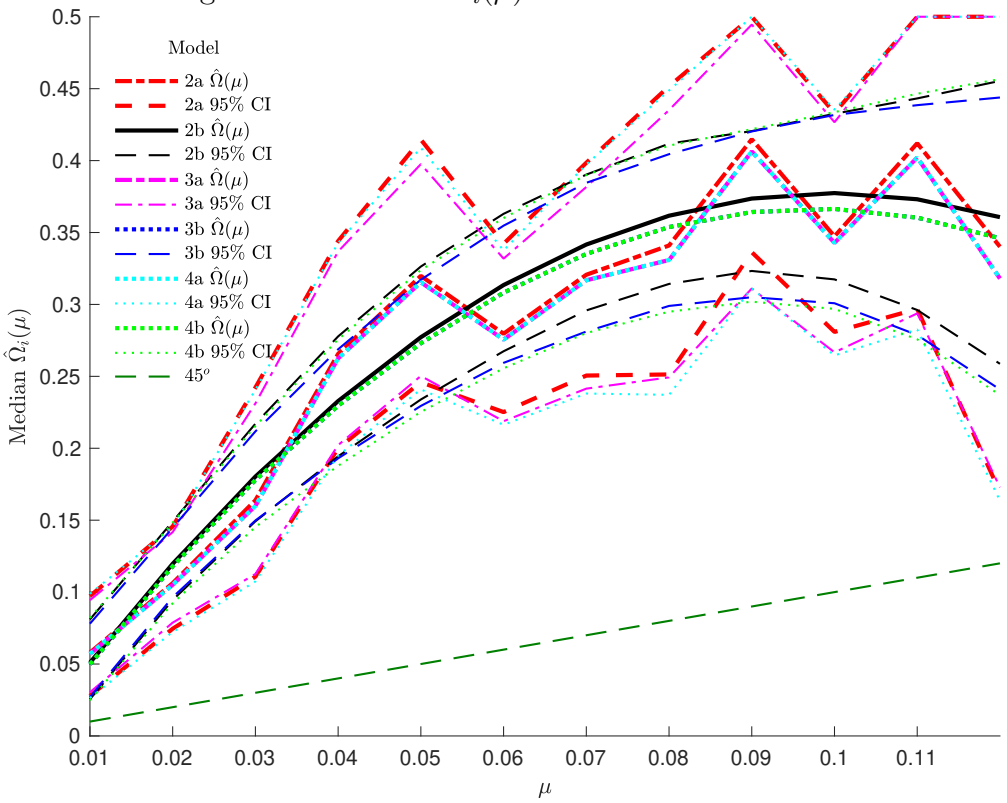
Note: 556 subjects in the Abstract treatment. See additional notes in Figure B.7 on how estimates are constructed.

Figure B.6: Median $\Omega_i(\mu)$ in Ambiguity treatment



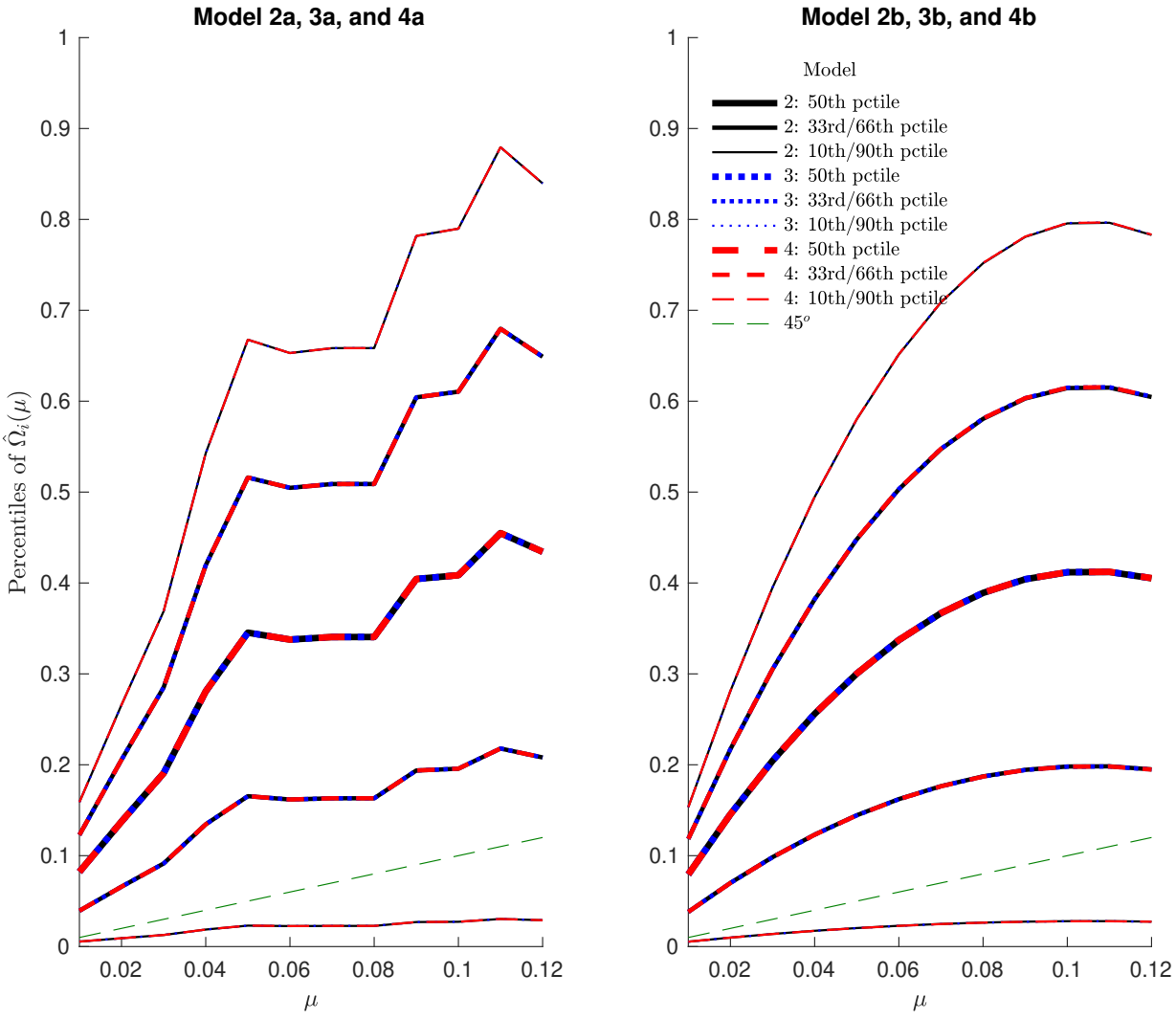
Note: This figure displays the estimated median $\Omega_i(\mu)$ from Model 2 and 3 across the 544 subjects in the Ambiguity treatment. Because the probability weighting function and r has a lower bound at 0, the 95% confidence intervals from the estimated values of Model 2a, 2b, 3a, and 3b are constructed with 1,000 bootstraps with random subsamples that are approximately 80% the size of the Ambiguity treatment subjects. Subsampling is used in these cases because the estimate of r approaches a lower bound in these models. The 95% confidence intervals for Model 4a and 4b are constructed from 1,000 bootstraps of N=544 drawn from the original sample with replacement.

Figure B.7: Median $\Omega_i(\mu)$ in Abstract treatment



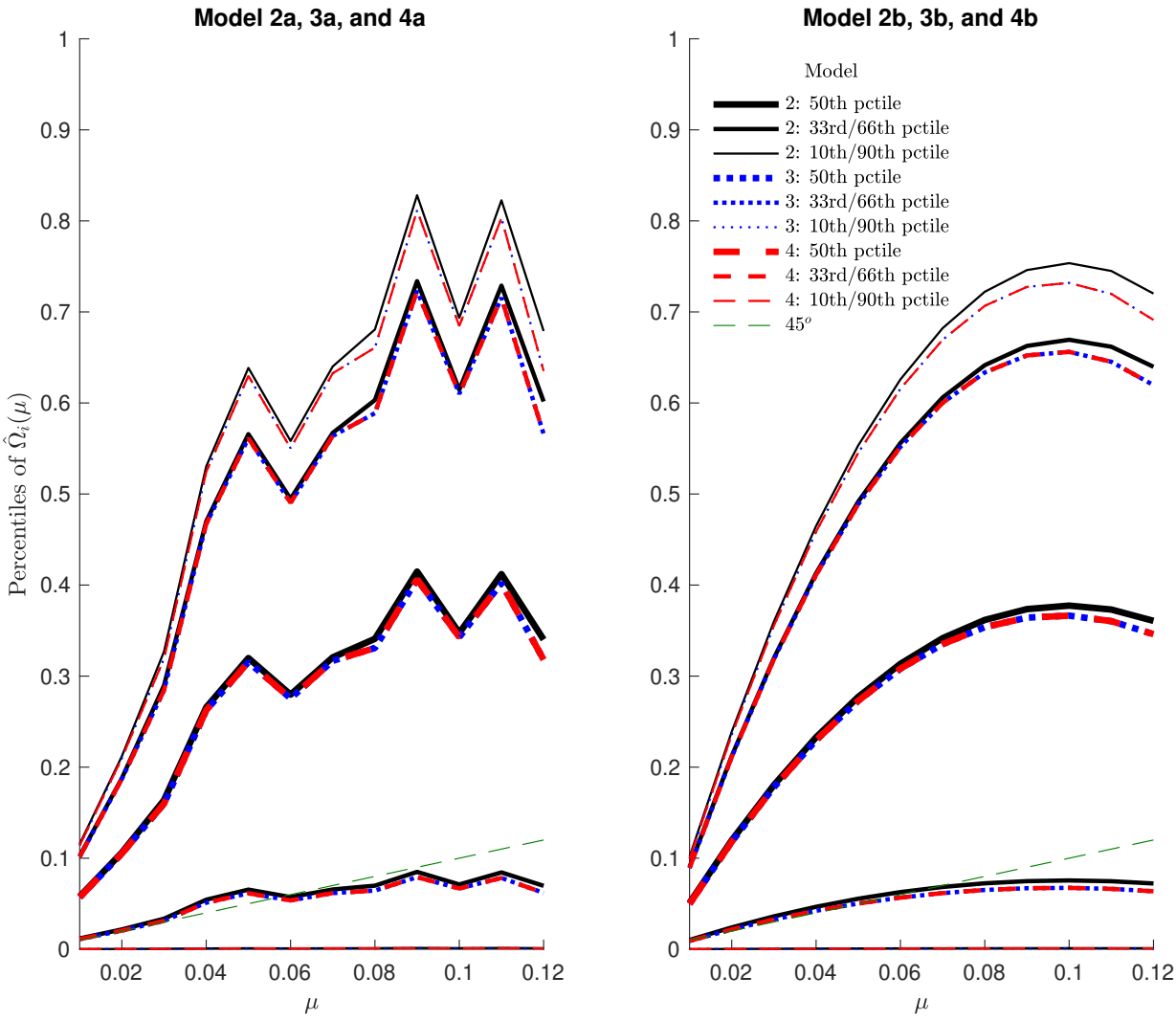
Note: This figure displays the estimated median $\Omega_i(\mu)$ from Model 2 and 3 across the 556 subjects in the Abstract treatment. Because the probability weighting function and r has a lower bound at 0, the 95% confidence intervals from the estimated values of Model 2a, 2b, 3a, and 3b are constructed with 1,000 bootstraps with random subsamples that are approximately 80% the size of the Abstract treatment subjects. Subsampling is used in these cases because the estimate of r approaches a lower bound in these models. The 95% confidence intervals for Model 4a and 4b are constructed from 1,000 bootstraps of N=556 drawn from the original sample with replacement.

Figure B.8: Heterogeneity of $\Omega_i(\mu)$ in Ambiguity treatment



Note: Two panels that show the quantiles of the estimated $\Omega_i(\mu)$ given the estimated noise $\Phi_{\Omega,i}$ using subjects from the Ambiguity treatment. The quantiles of $\Omega(\mu)$ from Model 2a, 3a, and 4a are displayed on the left, while the quantiles from model 2b, 3b, and 4b are displayed in the right panel. See additional notes in Figure B.6 on how estimates are constructed.

Figure B.9: Heterogeneity of $\Omega_i(\mu)$ in Abstract treatment



Note: Two panels that show the quantiles of the estimated $\Omega_i(\mu)$ given the estimated noise $\Phi_{\Omega,i}$ using subjects from the Abstract treatment. The quantiles of $\Omega(\mu)$ from Model 2a, 3a, and 4a, are displayed on the left, while the quantiles from Model 2b, 3b, and 4b are displayed in the right panel. See additional notes in Figure B.7 on how estimates are constructed.

APPENDIX C

APPENDIX OF CHAPTER 3: DIFFERENT CONTEXTS, DIFFERENT RISK PREFERENCES?

C.1 Risk Controls

The risk controls are expected annual claims under each coverage based on separate Poisson-gamma Bayesian credibility models. More specifically, we assume that household i's claims under coverage j in year t follow a Poisson distribution with arrival rate λ_{ijt}. We treat λ_{ijt} as a latent random variable and assume that $\ln \lambda_{ijt} = \mathbf{z}'_{ijt}\alpha_j + \epsilon_{ij}$,where $\mathbf{z}_{ijt}$ is a vector of observables, α_j is a vector of coefficients, ϵ_{ij} is an iid error term, and $\exp(\epsilon_{ij})$ follows a gamma distribution with unit mean and variance ϕ_j. Utilizing our full dataset, we perform separate Poisson panel regressions with random effects to obtain maximum likelihood estimates of α_j and ϕ_j for each coverage j. For each household i in the baseline sample, we then calculate the expected number of claims $\widehat{\lambda}_{ij}$ for each coverage j, conditional on the household's ex ante characteristics $\mathbf{z}_{ij}$ and ex post claims experience γ_{ij}, as follows: $\widehat{\lambda}_{ij} = \exp(\mathbf{z}'_{ij}\widehat{\alpha}_j)E(\exp(\epsilon_{ij})|\gamma_{ij})$, where $E(\exp(\epsilon_{ij})|\gamma_{ij})$ is calculated assuming $\exp(\epsilon_{ij})$ follows a gamma distribution with unit mean and variance $\widehat{\phi}_j$. Observe that by construction $\widehat{\lambda}_{ij}$ takes into account both the systematic and idiosyncratic components of a households' risk type.[1]

[1]We refer to the above-described model as a Bayesian credibity model because $\widehat{\lambda}_{ij}$ corresponds to the Bayesian credibility premium in the actuarial literature (Denuit et al., 2007, ch. 3).

C.2 Classification of the Contexts in Einav et al. (2012)

In order to compare our results with those of Einav et al. (2012), we classify each of their contexts according to the magnitude of the stakes involved. Moreover, we apply the same criteria to classify their contexts that we use to classify our contexts. Recall that in our small-stakes contexts the values of the options and the inter-option increments range in the hundreds and thousands dollars, whereas in our large-stakes contexts the value of the options range in the hundreds of thousands and millions of dollars with inter-option increments that range in the hundreds of thousands dollars.

Based on their description of the coverage options in each context (Einav et al., 2012, pp. 2612-2616),[2] we conclude that none of their contexts involve large-stakes choices. Three of their six contexts—health, drug, and dental insurance—involve small-stakes choices. In health insurance, employees effectively choose among deductible options that range from zero to $3,000 (with a mean inter-option increment of $750) for in-network care and from $500 to $6,000 (with a mean inter-option increment of $1,375) for out-of-network care. In drug insurance, employees choose among brand drug cost-sharing percentages that range from 30 percent to 50 percent for retail purchases and from 20 percent to 40 percent for mail-order purchases. The mean of the resulting annual drug claims is approximately $1,500 and the 95th percentile is approximately $5,500. In dental insurance, employees effectively choose between a maximum annual benefit of $1,000 or $2,000.

In two of the three remaining contexts—short-term disability insurance and 401(k) investments—the stakes range in the thousands and tens of thousands of dollars but not the hundreds of thousands of dollars, and so we classify them

[2] See also pp. 4-5 in their Online Appendix.

as moderate-stakes contexts. In short-term disability insurance, which replaces lost wages due to disability for up to six months, employees choose among wage-replacement rates that range from 60 percent to 100 percent.[3] The mean annual wage of the employees in their baseline sample is approximately $58,000 and the 95th percentile is approximately $114,000. At the mean claim duration, which Einav et al. (2012) report is approximately two months, this suggests that the value of the benefit ranges approximately from $5,800 to $9,700 for the average employee and does not exceed $19,000 for 95 percent of employees. Even at the maximum claim duration, the value of the annual benefit ranges approximately from $17,000 to $29,000 for the average employee and does not exceed $57,000 for 95 percent of employees. In 401(k) investments, contributing employees choose how to allocate their contributions among 13 different funds whose prospective monthly returns range from -11.69 percent to 16.79 percent.[4] The mean annual contribution is approximately $4,600 and the maximum allowable is $18,000,[5] with Alcoa matching contributions up to six percent. This suggests that the stakes range approximately from $-\$2,200$ to $8,300 for the average contributor and from $-\$8,500$ to $32,400 for all contributors.

We also classify the remaining context—long-term disability insurance—as involving moderate-stakes choices, though the classification is less straightforward than in the other contexts. Alcoa's long-term disability plan replaces lost wages due to disability for durations longer than six-months, subject to a six-month

[3] In their Appendix Table A1, Einav et al. (2012) note that "sometimes" the wage-replacement rates in short-term disability insurance range instead from 40 percent to 80 percent.

[4] Einav et al. (2012) abstract from the employees decisions as to whether and how much to contribute, but rather focus on how contributing employees choose to allocate their contributions across the funds. The range of monthly returns is taken from Appendix Table A2 in Einav et al. (2012), which reports summary statistics of the funds' monthly returns from August 2005 to December 2007.

[5] Einav et al. (2012) state that the choices were made in 2004. We assume they reflect benefit elections for 2005. In 2005, the annual contribution limit was $14,000 for employees under age 50 and $18,000 for older employees.

elimination period.[6] Employees choose among three wage-replacement rates: 50 percent, 60 percent, or 70 percent. At the mean claim duration, which Einav et al. (2012) report is approximately one year,[7] this suggests that the value of the benefit ranges approximately from $29,000 to $41,000 for the average employee and does not exceed $80,000 for 95 percent of employees. At the maximum claim duration, which we assume could be as long as 45 years,[8] the present value of the benefit could range into the hundred of thousands of dollars; but even in this extreme case the present value of the inter-option increments would range in the tens of thousands of dollars.[9] All things considered, we conclude the stakes are best classified as moderate.

[6]The elimination period is the period of time between the onset of disability and the time at which the employee is eligible to receive benefits.

[7]Einav et al. (2012) note that their claims data are truncated at about two years, which suggests the mean claim duration may be longer than one year. In a recent study of employer-provided long-term disability insurance, Autor et al. (2014) report a mean claim duration of 1.55 years and a median of one year. Their sample consists of approximately 8 million quarterly observations from nearly 10,000 unique employers, and their claims data span eight years.

[8]Einav et al. (2012) do not report the maximum claim duration (or the 95th percentile) in their data, nor do they report the maximum benefit period under Alcoa's long-term disability plan. The maxmimum benefit period under many long-term disability plans is 2, 5, or 10 years, but under the most generous plans it runs until the employee's social securty full retirement age, which is 67 for employee's born in 1960 or later. Assuming that Alcoa's plan has the most generous maximum benefit period and that its youngest eligible employee is 22 years old, we arrive at the assumption that the maximum claim duration could be as long as 45 years.

[9]We are assuming annual discount rates well in excess of 10 percent, which is consistent with the preponderance of the empirical evidence on time preferences (Frederick et al., 2002, pp. 377-380). For instance, Warner and Pleeter (2001) estimate the personal discount rates of approximately 66,000 U.S. millitary personnel who were offered separation benefits that consisted of a choice between a lump sum or an annuity, where the break-even discount rate was at least 17.5 percent. They find that "over half of the officers and over 90 percent of enlisted personnel chose the lump-sum payment, implying that the vast majority of personnel had discount rates of at least 18 percent" (p. 33). Based on regression analysis, they report mean discount rates of between 10 percent and 19 percent for officers and between 35 percent and 54 percent for enlisted personnel, depending on the model specification (p. 48, tbl. 6). As Frederick et al. (2002, p. 385) note, this field study "is particularly compelling in terms of credibility of reward delivery, magnitude of stakes, and number of subjects."

BIBLIOGRAPHY

(2016). The Medicare Part D Prescription Drug Benefit. Technical report, The Henry J. Kaiser Family Foundation.

Abaluck, J. and Gruber, J. (2016). Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance. *American Economic Review*, 106(8):2145–2184.

Abaluck, J., Gruber, J., and Swanson, A. (2018). Prescription drug use under Medicare Part D: A linear model of nonlinear budget sets. *Journal of Public Economics*, 164:106–138.

Anderson, L. R. and Mellor, J. M. (2009). Are risk preferences stable? comparing an experimental measure with a validated survey-Based measure. *Journal of Risk and Uncertainty*, 39(2):137–160.

Anderson, T. W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297.

Aron-Dine, A., Einav, L., Finkelstein, A., and Cullen, M. (2015). Moral Hazard in Health Insurance: Do Dynamic Incentives Matter? *Review of Economics and Statistics*, 97(4):725–741.

Arrow, K. J. (1971). *Essays on the Theory of Risk Bearing*. Markham Publishing, Chicago.

Autor, D., Duggan, M., and Gruber, J. (2014). Moral hazard and claims deterrence in private disability insurance. *American Economic Journal: Applied Economics*, 6(4):110–141.

Baicker, K., Mullainathan, S., and Schwartzstein, J. (2015). Behavioral Hazard in Health Insurance. *The Quarterly Journal of Economics*, 130(4):1623–1667.

Baker, D. (2006). The Origins of the Doughnut Hole: Excess Profits on Prescription Drugs. Technical Report 2006-19, Center for Economic and Policy Research (CEPR).

Baker, S. R. (2017). Debt and the response to household income shocks: Validation and application of linked financial account data. *Journal of Political Economy*, forthcoming.

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

Barseghyan, L., Molinari, F., O'Donoghue, T., and Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, 103(6):2499–2529.

Barseghyan, L., Molinari, F., O'Donoghue, T., and Teitelbaum, J. C. (2018). Estimating Risk Preferences in the Field. *Journal of Economic Literature*, 56(2):501–564.

Barseghyan, L., Molinari, F., and Teitelbaum, J. C. (2016). Inference under stability of risk preferences. *Quantitative Economics*, 7(2):367–409.

Barseghyan, L., Prince, J., and Teitelbaum, J. C. (2011). Are risk preferences stable across contexts? evidence from insurance data. *American Economic Review*, 101(2):591–631.

Berinsky, A. J., Huber, G. A., and Lenz, G. S. (2012). Evaluating online labor

markets for experimental research: Amazon.com's mechanical turk. *Political Analysis*, 20(3):351–368.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American Journal of Agricultural Economics*, 62(3):395–407.

Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78(4):1375–1412.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5.

Chetty, R. and Szeidl, A. (2007). Consumption commitments and risk preferences. *Quarterly Journal of Economics*, 122(2):831–877.

Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007). Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, 97(5):1921–1938.

Cohen, A. and Einav, L. (2007). Estimating risk preferences from deductible choice. *American Economic Review*, 97(3):745–788.

Collier, B. L., Schwartz, D., Kunreuther, H. C., and Michel-Kerjan, E. O. (2017). Risk preferences in small and large stakes: Evidence from insurance contract decisions. National Bureau of Economic Research Working Paper 23579.

Dalton, C., Gowrisankaran, G., and Town, R. (2018). Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D. *Unpublished*.

Denuit, M., Maréchal, X., Pitrebois, S., and Walhin, J.-F. c. (2007). *Actuarial*

Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems. John Wiley, Chichester.

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

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

Einav, L., Finkelstein, A., and Polyakova, M. (2018). Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D. 10(2):August 2018, forthcoming.

Einav, L., Finkelstein, A., and Schrimpf, P. (2015). The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D. *The Quarterly Journal of Economics*, 130(2):841–899.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669.

Etner, J., Jeleva, M., and Tallon, J.-M. (2012). Decision theory under ambiguity. *Journal of Economic Surveys*, 26(2):234–270.

Fehr-Duda, H., Bruhin, A., Epper, T., and Schubert, R. (2010). Rationality on the rise: Why relative risk aversion increases with stake size. *Journal of Risk and Uncertainty*, 40(2):147–180.

Fox, C. R. and Tversky, A. (1998). A Belief-Based Account of Decision under Uncertainty. *Management Science*, 44(7):879–895.

Frederick, S., Lowenstein, G., and O'Donohue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401.

Friedman, M. and Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, 56(4):279–304.

Galizzi, M. M., Machado, S. R., and Miniaci, R. (2016a). Temporal stability, cross-Validity, and external validity of risk preferences measures: Experimental evidence from a uk representative sample.

Galizzi, M. M., Miraldo, M., and Stavropoulou, C. (2016b). In sickness but not in wealth: Field evidence on patients' risk preferences in financial and health domains. *Medical Decision Making*, 36(4):503–517.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153.

Gonzalez, R. and Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38(1):129–166.

Gul, F. (1991). A theory of disappointment aversion. *Econometrica*, 59(3):667–686.

Handel, B., Hendel, I., and Whinston, M. D. (2015). Equilibria in health exchanges: Adverse selection versus reclassification risk. *Econometrica*, 83(4):1261–1313.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–2682.

Harrell, F. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer Series in Statistics. Springer-Verlag, New York.

Hausman, D. M. (2012). *Preference, Value, Choice, and Welfare*. Cambridge University Press, Cambridge.

Ho, K., Hogan, J., and Morton, F. S. (2017). The impact of consumer inattention on insurer pricing in the Medicare Part D program. *The RAND Journal of Economics*, 48(4):877–905.

Hoadley, J., Summer, L., Hargrave, E., and Cubanski, J. (2011). Understanding The Effects of The Medicare Part D Coverage Gap in 2008 and 2009. Technical Report 8221, The Henry J. Kaiser Family Foundation.

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

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). Estimating Vector Autoregressions with Panel Data. *Econometrica*, 56(6):1371–1395.

Ioannou, C. A. and Sadeh, J. (2016). Time preferences and risk aversion: Tests on domain differences. *Journal of Risk and Uncertainty*, 53(1):29–54.

Jaffe, S. and Malani, A. (2017). The welfare implications of health insurance. Human Capital and Economic Opportunity Working Group Working Paper 2017-045.

Joyce, G. F., Zissimopoulos, J., and Goldman, D. P. (2013). Digesting the doughnut hole. *Journal of Health Economics*, 32(6):1345–1355.

Kachelmeier, S. J. and Shehata, M. (1992). Examining risk preferences under high monetary incentives: Experimental evidence from the people's republic of china. *American Economic Review*, 82(5):1120–1141.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Kőszegi, B. and Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, 97(4):1047–1073.

Kowalski, A. (2016). Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care. *Journal of Business & Economic Statistics*, 34(1):107–117.

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy*, 60(2):151–158.

Mason, W. and Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, 44(1):1–23.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics*, pages 105–142. Academic Press, New York.

McFadden, D. (1981). Econometric models of probabilistic choice. In Manski, C. F. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*, pages 198–272. The MIT Press, Cambridge, Mass.

MedPAC (2016). Prescription Drugs. In *Health Care Spending and the Medicare Program*, pages 170–171.

Menezes, C. F. and Hanson, D. L. (1970). On the theory of risk aversion. *International Economic Review*, 11(3):481–487.

Morris, D. S., Schwarcz, D., and Teitelbaum, J. C. (2017). Do credit-based insurance scores proxy for income in predicting auto claim risk? *Journal of Empirical Legal Studies*, 14(2):397–423.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417–1426.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. SSRN Scholarly Paper ID 1626226, Social Science Research Network, Rochester, NY.

Polinski, J. M., Bhandari, A., Saya, U. Y., Schneeweiss, S., and Shrank, W. H. (2010). Medicare beneficiaries' knowledge of and choices regarding Part D, 2005 – present. *Journal of the American Geriatrics Society*, 58(5):950–966.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122–136.

Prelec, D. (1998). The probability weighting function. *Econometrica*, 66(3):497.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343.

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

Read, D., Loewenstein, G., and Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, 19(1-3):171–197.

Saha, A. (1993). Expo-power utility: A 'flexible' form for absolute and relative risk aversion. *American Journal of Agricultural Economics*, 75(4):905–913.

Smith, S. (2013). 4 ways to ensure valid responses for your online survey. http://www.qualtrics.com/blog/online-survey-valid-responses/.

Snowberg, E. and Wolfers, J. (2010). Explaining the Favorite—Long shot bias: Is it risk-love or misperceptions? *Journal of Political Economy*, 118(4):723–746.

Sydnor, J. (2010). (Over)insuring modest risks. *American Economic Journal: Applied Economics*, 2(4):177–199.

Tversky, A. and Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, 59(4):S251–S278.

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

Warner, J. T. and Pleeter, S. (2001). The personal discount rate: Evidence from military downsizing programs. *American Economic Review*, 91(1):33–53.

Wolf, C. and Pohlman, L. (1983). The recovery of risk preferences from actual choices. *Econometrica*, 51(3):843–850.

Zeckhauser, R. and Keeler, E. (1970). Another type of risk aversion. *Econometrica*, 38(5):661–665.

Zhang, Y., Donohue, J. M., Newhouse, J. P., and Lave, J. R. (2009). The Effects Of The Coverage Gap On Drug Spending: A Closer Look At Medicare Part D. *Health Affairs*, 28(2):w317–w325.