Decision-Making in Health Insurance Markets

by

Tomas P. Sanguinetti

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2020 by the
Graduate Supervisory Committee:

Nicolai Kuminoff, Chair
Edward Schlee
Jonathan Ketcham
Daniel Silverman

ARIZONA STATE UNIVERSITY

May 2020

ABSTRACT

Prior research on consumer behavior in health insurance markets has primarily focused on *individual* decision making while relying on strong *parametric* assumptions about preferences. The aim of this dissertation is to improve the traditional approach in both dimensions. First, I consider the importance of joint decision-making in individual insurance markets by studying how married couples coordinate their choices in these markets. Second, I investigate the robustness of prior studies by developing a non-parametric method to assess decision-making in health insurance markets.

To study how married couples make choices in individual insurance markets I estimate a stochastic choice model of household demand that takes into account spouses' risk aversion, spouses' expenditure risk, risk sharing, and switching costs. I use the model estimates to study how coordination within couples and interaction between couples and singles affects the way that markets adjust to policies designed to nudge consumers toward choosing higher value plans, particularly with respect to adverse selection.

Finally, to assess consumer decision-making beyond standard parametric assumptions about preferences, I use second–order stochastic dominance rankings. Moreover, I show how to extend this method to construct bounds on the welfare implications of choosing dominated plans.

DEDICATION

To My Wife

ACKNOWLEDGMENTS

I am grateful to my advisors Nick Kuminoff, Eddie Schlee, Jonathan Ketcham, and Dan Silverman for their guidance and support. I am profoundly grateful to them for all their patience and all the time they have dedicated to me.

I also want to thank the entire faculty and staff at ASU Department of Economics. I want to thank professors of the department who were important during my graduate studies, including Gustavo Ventura, Domenico Ferraro, Alvin Murphy, Kelly Bishop, Natalia Kovrijnykh, Esteban Aucejo, Hector Chade, and Manjira Datta.

Many of my classmates and colleagues provided valuable comments and support during these years, including Sophie Mathes, Jacob French, Jakob Dowling, Jorge Mesias, Santiago Garcia, Nazim Tamkoc, Sayahnika Basu, Juhee Bae, Nirman Saha, and Mehrdad Esfahani.

Finally, this dissertation would have not been possible without the support of my family. In particular, I want to give special thanks to my wife Laura, for all her support during these difficult years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Consumer behavior often departs from predictions made by standard models in economics, (Dellavigna, 2009; Bernheim *et al.*, 2019). Researchers have called attention to this in insurance markets, Kunreuther *et al.* (2013), energy markets, Allcott and Taubinsky (2015), and markets for employer-sponsored savings plans, Madrian and Shea (2001). Within the literature on insurance markets, several studies focus on health insurance in particular, and conclude that many consumers make sub-optimal choices (Abaluck and Gruber, 2011; Handel, 2013; Sinaiko and Hirth, 2011; Bhargava *et al.*, 2017; Liu and Sydnor, 2018). However, these conclusions are based on strong *parametric* assumptions about the shapes of consumer preferences and also assume that consumers make choices *independently*.

In chapter 2 of this essay, I investigate how married couples coordinate their choices in individual insurance markets. I do this by estimating a stochastic choice model of household demand that takes into account risk aversion, expenditure risk, risk sharing, and switching costs. I use the estimates to analyze how coordination within couples and interactions between couples and singles affects the way that markets adjust to policies designed to nudge consumers toward choosing higher value plans, particularly with respect to adverse selection. The data reveal several striking facts about insurance choice. In particular, I find that 78% of couples decide to "pool" by buying the same plan. This figure remains constant even for couples with extremely different health risks. My estimates imply that the average couple is willing to pay $1,584 per year to avoid searching for separate plans, which is approximately three times the annual average plan premium. In contrast, a version of the model that ignores marital status and therefore, pooling incentives, would predict that only 4% of couples pool. I find that nudging households to choose the plans that maximize their expected utility would yield larger welfare gains for couples than singles.

In chapter 3 I return to modeling individual behavior but relax the parametric assumptions on preferences. Instead, I use a non-parametric framework to evaluate consumer choices. Specifically, I use stochastic dominance rankings to assess the quality of consumer decision-making under uncertainty when consumers choose prescription drug insurance plans in Medicare Part D. I use this framework to develop measures of decision-making quality at both the extensive margin (e.g. choosing an object off the efficient frontier) and the intensive margin (e.g. distance from the frontier). I also compare my method with mean-variance frontier analysis which is often used in insurance markets. My results can be summarized as follow. Mean-variance and second-order stochastic dominance have similar implications for average consumer behavior: 70% of consumers select dominated plans (20% if I allow for unrestricted heterogeneity in preferences for unobserved plan quality by focusing exclusively on plans sold by the same insurer). However, the two approaches to evaluating consumer decision making differ in the set of plans that they label as dominated, especially when I control for non-financial attributes. The welfare ambiguity of choosing dominated plans is large. The average welfare loss captured by the upper bound is eight times higher than the lower bound, \$1,307 versus \$170 respectively. This implies that the average welfare loss is never smaller than 34% of the annual premium. Finally, while some consumers suffer welfare losses higher than \$2,000 per year, the probability of being in this group is not systematically correlated with observable characteristics such as income, education, and gender.

Chapter 2

HOW DO COUPLES CHOOSE INDIVIDUAL INSURANCE PLANS

Couples often participate in *individual* health insurance markets, many of which are federally regulated. Examples include Medicare Advantage, Medicare Part D and Medigap. In these markets, married couples have incentives to coordinate their plan choices since they share a budget constraint, risks, and information obtained from their search processes. Virtually nothing is known about this form of coordination and its implications for equity and efficiency of markets. Understanding how couples coordinate their enrollment decisions is potentially important for evaluating consumer welfare from health insurance and for assessing policies targeting market design. For example, it could improve our ability to predict insurance enrollment decisions if spouses' choices are highly correlated. Equally important, if couples are less risk-averse than their individual members because they are able to share risks, then their collective enrollment decisions will differ from the plans they would have chosen separately. Marital status may also contain policy relevant information about risk beyond what can be learned from existing medical conditions and demographics.

This research is the first economic study to investigate how couples make enrollment choices in individual insurance markets. I leverage administrative records to determine the marital status of a large panel sample of Medicare Part D enrollees. I first distinguish households that are comprised of married couples from households that consist of singles who are widowed or divorced. Then, for each type of household, I estimate a stochastic choice model of insurance demand that incorporates risk aversion, expenditure risk, and inertia. Finally, I combine my estimates for household-level insurance demand with a parsimonious model of plan pricing to study the distributional welfare consequences of policies designed to nudge individuals or households toward choosing certain types of plans. This exercise allows me to investigate how coordinated decision making by couples modifies the implications of nudging for adverse selection. I also use the model to investigate how standard regulations in health insurance markets affect the ways in which singles and couples would sort themselves across the market in response to nudges, and what this sorting behavior implies for consumer welfare.

I start by using information on a random sample of approximately 2 million beneficiaries' residential locations, last names and basis for social security eligibility to identify different household types. These data allow me to identify approximately 75,000 couples making repeated insurance plan choices over the first five years of the Medicare Part D program, 2006-2010. These data enable me to provide the first direct evidence on how couples make their insurance plan enrollment decisions, and how their behavior compares with singles who are widowed or divorced.

The linked household-level data reveal several striking facts about couples' insurance plan choices. First, nearly 80% of couples buy the same plan. Second, this statistic is virtually invariant to the difference between spouses' health risks. Third, inertia affects couples and singles similarly, implying that couples do not fully exploit economies of scale in information. Indeed, approximately 85% of married couples reenroll in their default plan combinations each year. Fourth, while I find some evidence of positive assortative mating in prescription drug risk, the magnitude is small compared to evidence on assortative mating in other contexts such as education, (Fernández *et al.*, 2005). Fifth, couples account for a substantial share of the market (54%) and, on average, have substantially lower costs for insurance companies ($871) compared to widows, who account for 30% of the market and have average costs of $1,116.[1] Moreover, this difference is exacerbated by the way the federal government adjusts subsidies to insurance companies based on the risk of its

[1]Costs here represent total prescription drug expenditures minus the out of pocket expenditure of each household. This difference represents the cost that insurance companies incur.

consumer pool, which inflates the cost differential between couples and widows by 35%.

I model household behavior with a stochastic choice model. Households are assumed to choose plans based on a deterministic core theory, expected utility, and a random error, (Hey and Orme, 1994; Harless *et al.*, 1994; von Gaudecker *et al.*, 2011). The scale of the logistic error will be household-type-specific to capture heterogeneity in households' decision processes. I follow standard practice in the health insurance literature by adding "inertia" parameters describing the disutility of switching plans to help explain the low rates of consumer switching. Similarly, I add "pooling" parameters describing the disutility for couples to choose separate plans to help explain the low rates at which couples choose different plans. Each household type will choose plans that maximize the certainty equivalent corresponding to expected utility, plus the combined effects of inertia, pooling, and a random shock. This representation allows me to represent plan choices as lotteries with inertia and pooling defined in certainty equivalent terms, similar to Handel (2013).

Estimation proceeds in two stages. First I estimate a distribution of individual-specific parameters describing the degree of constant absolute risk aversion (CARA) using data from the Health and Retirement Study (HRS). Then I use observable measures of individual demographics and prescription drug spending to project this distribution onto the Medicare Part D population. The risk aversion parameters are identified by individuals' responses to a set of questions in the HRS that were designed to elicit risk aversion by asking each individual to choose among hypothetical monetary gambles. I find that risk aversion tends to be higher among older individuals, those with higher prescription drug expenditures, and females. Then I use the risk aversion measures predicted for each individual in each household to construct a CARA specification for household utility. This implies that widows are the most risk averse household types, and married couples the least.

Next I estimate household utility parameters for each household type. A model in which spouses maximize expected utility (EU) jointly predicts 35% of couples will choose the same plan. In contrast, a version of the model that ignores marital status and therefore, pooling incentives, would predict that only 4% of couples pool. To reproduce the fact that 78% of couples pool I augment the model with a pooling parameter that captures the residual value of this behavior. The monetary value of pooling ($1,584) is approximately half the implied monetary estimates of status quo inertia for couples, ($3,152). This is striking, given that both behaviors likely reflect some of the same mechanisms. Status quo inertia for widows is $1,975 compared to $1,472 for divorced women. Inertia is higher among widows because they are the most risk averse agents in the market and stand to gain the most by switching. The fact that they switch at the same rate as less risk averse household types is rationalized by higher inertia. I demonstrate that the difference in inertia between couples and singles is consistent with the hypothesis that one spouse selects both plans.

I use my estimates to simulate a counterfactual policy experiment in which a regulator nudges consumers to conform with expected utility. The experiment is replicated in environments with and without premiums subsidies and risk adjustment payments. Importantly, I recognize that the policy may affect couples and singles differently. In the actual Part D setting, both household types are made better off by the policy. In the preferred specification, welfare increases up to a 43% for couples and 5% for singles. Premiums decrease on average, revealing that most plans get advantageously selected. To better understand the mechanisms driving these results and explore model predictions in less regulated environments I replicate the same policy after eliminating premium subsidies. This alteration reduces the size of couples' welfare gains from the nudge to 28%, whereas singles are made worse off by the policy. Their welfare decreases by 5% on average. The nudge tends to induce (relatively low cost) couples to reduce their expenditures on plan premiums by moving to plans with smaller shares of widows. The subsequent increase in premiums makes (relatively high cost) singles who remain in those plans worse off. When I further eliminate risk adjustment I find that nudging enrollees increases couples' welfare and decreases singles' welfare moderately. The decrease among singles' is smaller relative to the environment with risk adjustment because risk adjustment makes singles more costly relative to couples. When couples sort into plans with smaller shares of widows, the variation in premiums is smaller relative to the environment with risk adjusted payments. Overall, these policy experiments reveal that premium subsidies are the key institutional feature that make the policy welfare-enhancing for most enrollees. This is important because it suggests that nudges are more likely to be welfare reducing in individual insurance markets that are not as heavily subsidized as Part D.

Finally, I investigate the importance of accounting for collective decision making in policy evaluations by repeating the policy analysis with and without accounting for the way that couples interact in their decision-making. A model where spouses are choosing individual insurance plans independently, predicts an increase in premiums and welfare gains only for couples. Moreover, a model where spouses choose plans independently can never reproduce the high pooling rates of spouses, with only 6% of couples choosing the same plan. This figure increases up to 90% after the intervention in a model where spouses choose plans jointly. Interestingly, this result impacts the reduction in default rates after the policy. In a model where spouses choose plan separately, the decrease in default rates is of 30 percentage points, three times higher than in a model where spouses choose plans jointly.

This paper contributes to several pieces of literature. First, it adds to the literature on behavioral health economics, Chandra *et al.* (2019), by providing the first evidence on how couples choose health insurance plans. Similar to conventional models of individual decision making in insurance markets, I estimate a stochastic choice model that incorporates inertia, (Abaluck and Gruber, 2011; Handel, 2013; Polyakova, 2016; Ketcham *et al.*, 2016). I extend this literature by providing the first evidence on the prevalence of within-household pooling and show that, like inertia, it has first-order implications for how couples choose individual insurance plans. Further, I demonstrate that pooling can change conclusions about the effects of policies that are intended to help consumers make more informed choices and have implications for adverse selection, building on prior work by the interaction between nudging and adverse selection by (Handel, 2013; Polyakova, 2016; Handel *et al.*, 2019).

My findings also contribute to the empirical literature of Medicare Part D, (Abaluck and Gruber, 2011; Kling *et al.*, 2012; Ketcham *et al.*, 2012; Ericson, 2014; Ketcham *et al.*, 2015; Abaluck and Gruber, 2016; Ketcham *et al.*, 2016; Polyakova, 2016; Ketcham *et al.*, 2019). However, my work differs from these prior studies in two important ways. First, I provide the first analysis of how household type affects the demand for prescription drug insurance plans and how different types of households interact in this market. Second, I adopt an expected utility framework and estimate a distribution of risk aversion measures in the Part D population in a novel way. I exploit the similarities between the HRS and the Medicare Part D population to project a distribution of absolute risk aversion parameters that were elicited using hypothetical gamble questions.

Finally, this paper also contributes to the broader literature on how household structure affects household decision making, (Fonseca *et al.*, 2012; Addoum, 2017). The main contribution relative to this literature is the unique financial setting where my spouses are making decisions: individual insurance markets.

The rest of the paper proceeds as follows. Section 2.1 briefly review relevant literature on inertia and adverse selection in Medicare Part D. Section 2.2 describes the data. Section 2.3 shows descriptive evidence of household inertia, plan pooling, assortative mating, and costs. Section 2.4 introduces the stochastic choice model of household type. Section 2.5 shows the estimates of risk aversion and model parameters. Finally, section 2.6 describes the different pricing models and counterfactual policies. Section 2.7 concludes.

2.1 Medicare Part D, Inertia and Adverse Selection

Medicare Part D was established in 2006 and was the largest expansion of the Medicare program since its inception. A novel feature of Medicare Part D, relative to traditional Medicare (Part A and Part B), was the creation of markets in which private insurance companies can sell standalone prescription drug insurance plans (PDP) to Medicare enrollees at prices that are subsidized by the federal government.[2] In 2016, about 40 million Medicare beneficiaries choose to enroll in plans that offered prescription drug coverage (70% of the Medicare population) and had average spending of 2,130 dollars per enrollee, (Hoadley et al 2016). From those 40 million, 60% were enrolled in a Part D prescription drug plan (PDP) while the rest where enrolled in Medicare Advantage.

The U.S. Centers for Medicare and Medicaid Services divides the United States into 34 geographic regions, each of which offers a distinct menu of plans. Insurance companies can offer multiple plans in a single region; they can offer different plans in different regions; and they can change the attributes of a given plan in a given region (e.g. premiums, co-payment rates) from year to year. Thus each region-year comprises a distinct market in the sense that all Medicare beneficiaries within the market choose among the same menu of plans. The default for new Medicare beneficiaries is to be uninsured. They must enter the market and actively choose a plan to become insured. Their choice becomes their automatic default plan for the following year. They will be re-enrolled in the same plan unless they actively switch plans, opt out of the market during the annual open enrollment period or if their plan exits the market the following year. Importantly, Medicare Part D, like Medicare Adavantage and Medigap, is a market for individual health insurance. When married seniors buy plans, they have to buy individual plans for each spouse. No family

[2]The other option for Medicare enrollees to get coverage for prescription drug expenses is to purchase Medicare Advantage drug plans.

plans or premium discounts for families are offered.

CMS regulates the PDP markets in several ways. First, people who enroll after age 65 are required to pay a penalty that increases their monthly premiums. Second, premiums are subsidized by the federal government and risk-adjusted in order to prevent adverse selection and "cream skimming". Third, firms that want to participate in the market must adhere to a regulated bidding process. Each year, firms submit bids that reflect the cost to supply the basic benefits to a person of average health.[3] The difference between the plan's bid and the government subsidy determines the plan premium that enrollees must pay.[4] Once a plan submits its bid for the upcoming year, it must accept all enrollees at the predetermined premium.[5] Finally, the payment that each firm receives for insuring an individual is equal to the bid, risk-adjusted by the individual's health condition. Thus, subsidies and risk-adjustment have key implications for costs, premiums and insurance payments in the Part D markets.

Although these policies induced most Medicare enrollees to participate in the market, they did not prevent some generous plans from suffering death spirals, (Heiss *et al.*, 2009). A plan suffers from an adverse selection death spiral when the plan's market share and premium start experiencing a rapid decrease and increase respectively. Indeed, most plans that were offering generous coverage in 2006 were no longer available in 2009, (Polyakova, 2016), suggesting a considerable degree of adverse selection. Polyakova (2016) confirms this by constructing a non-parametric test in the spirit of Chiappori and Salanie (2000). She finds that most generous plans attracted individuals with higher annual expenditures. The definition of adverse selection used in these studies is similar to the one I will employ. A plan is defined to be adversely selected relative to a baseline scenario if the expected costs

[3]The basic benefit parameters are set by CMS each year. They consist of three numbers: an annual deductible, an initial coverage limit and an out of pocket catastrophic threshold.

[4]Enrollees who qualify for the Low Income Subsidy (LIS) pay less than this resulted premium.

[5]See Stocking et al (2014) for more details about the bidding process.

of the plan is higher relative to the baseline scenario. This difference in costs will depend on the pool of consumers who choose the plan.

Another striking feature of the PDP market is the high rates of inertia. Kling *et al.* (2012), Polyakova (2016), Ketcham *et al.* (2016), Ho *et al.* (2017) document that enrollees rarely switch plans, with 90% of them passively reenrolling in their default option when available. This fact sparked considerable research on trying to understand the consequences of this behavior for market outcomes and welfare (e.g. Ericson, 2014; Ketcham *et al.*, 2016; Polyakova, 2016; Ho *et al.*, 2017). Some of these studies estimate the amount of money that enrollees would have to be paid ex-ante to switch to another alternative. For example, Polyakova (2016) estimates a monetary value of status quo inertia around 1,159 dollars and Ketcham *et al.* (2016) estimate a range between 809 and 3,660 dollars depending on enrollees' characteristics. These estimates are generally interpreted as reflecting a combination of search and switching costs, and inattention.

Evidence on the quantitative importance of inertia motivated the study of policies that nudge consumers toward different choices. For example, Polyakova (2016) explores the welfare consequences of nudging Part D enrollees away from their default plans. Her results suggest moderate adverse selection and higher welfare gains from better plan-person matches. Ketcham *et al.* (2019) compare the distributional welfare consequences of three specific policies that nudge consumers: restricting menus, smart defaults and providing personal information about potential savings from switching plans to Part D enrollees. They conclude that although none of these policies are Pareto efficient, personalized information benefits most enrollees.

I advance this literature in several ways. First, I study how couples choose plans and how their choices affect the trade-off between nudging and adverse selection. Understanding how couples make choices could improve our ability to predict insurance enrollment decisions and evaluate distributional welfare implications of policies. Second, I use expected utility to consider the role of risk aversion, how it is distributed across different type of households, and how it is correlated with consumer risk. These features are crucial to predict how polices that aim to nudge consumers toward enrolling in certain types of plans (e.g. lower cost, greater risk protection) will affect adverse selection, (de Meza and Webb, 2001; Cutler *et al.*, 2008). The expected utility framework also has the advantage of making sharp predictions for how counterfactual policies will affect adverse selection through resorting. Third, I study how nudging policies interact in markets with standard regulations like premium subsidies and risk-adjustment payments. With a stylized model of plan pricing I am able to study how risk adjustment payments and subsidies affect enrolees' sorting patterns.

2.2 Data

2.2.1 Medicare Part D

I begin with a random 20% sample of all Medicare beneficiaries age 65 and above who participated in Part D market between 2006 and 2010. This sample comprises more than two million individuals. I also observe all of the financial characteristics of plans including plan premiums, deductibles and coinsurance rates, as well as non-financial characteristics including brand names and CMS star-ratings.[6] Finally, I observe the quantities of each specific drug each person purchased each year under their chosen plan.

Table 1 shows summary statistics for the evolution of the choice set over the first five years of the program. The number of plans and brands change each year. This change in plan menus will be crucial for the identification of status quo inertia. The other striking feature about the market is the substantial variation in annual premiums.

[6]The Centers for Medicare Medicaid Services (CMS) created a Five Star Quality Rating System that rates Part D plans. Ratings are between 1 and 5, 5 being the highest, for health plan quality based on measurements of customer satisfaction and quality of care the plan delivers.

Table 2.1: Medicare Part D 2006-2010

	2006	2007	2008	2009	2010
Average # Plans	44	56	55	50	47
Average # Brands	20	25	23	24	21
Mean Premiums ($)	450	440	478	547	566
sd Premiums ($)	160	185	238	245	235

Notes: Table 2.1 shows summary statistics of the Medicare Part D market for the first five years of the program. The average number of plans, brands and premiums are calculated across regions.

2.2.2 CMS Administrative Records

I match the Part D data to administrative records containing rich information on chronic medical conditions, demographics, annual residential location at the level of a zip-9 code, dates of death, and last names. The CMS records also include a beneficiary identification code (BIC) that specifies the basis of the individual's eligibility for cash payment programs, mainly Social Security. When the individual qualifies under another person's account, e.g. as a spouse, the code identifies the type of relationship between the individual and the primary beneficiary. In particular, widows and divorced people, are entitled to claim their ex-spouses social security benefits.[7] I will use this variable, BIC, to identify singles in the CMS records.

2.2.3 MCBS

I merge the Part D data with survey responses for all individuals who participated in the Medicare Current Beneficiary Survey (MCBS) between 2005 and 2011. I use the MCBS

[7] In general, widows can claim their deceased spouse's benefits when they are age 60 or older and they don't remarry before age 60. Divorced people can also claim benefits based on their ex-spouses work. Generally, they can do it if the following conditions are met; they reach age 62, the marriage lasted 10 years or longer, they are still unmarried, and the benefits they are entitled through their ex-spouses work are higher than the benefit they are entitled through their own work.

to compare my match rates of couples and singles with the true rates in the Medicare population.[8] From here I can obtain the marital status of approximately 3,500 seniors.

2.2.4 HRS

In order to estimate the degree of risk aversion for each individual in the Part D sample, I use a set of questions that were specially designed to elicit risk aversion parameters on the Health and Retirement Study (HRS). The HRS is a longitudinal panel study that surveys a representative sample of approximately 20,000 seniors in the United States. Importantly, the survey contains rich information on demographic characteristics and prescription drug expenditures. While I am unable to match individuals across the HRS and Part D samples, I leverage the fact that they describe the same population. I first estimate a distribution of absolute risk aversion measures in the HRS as a function of individual demographics and prescription drug spending. Then I project this distribution onto the Part D sample. This procedure for generating imputed variables in CMS-samples using HRS data is similar to the approach used in Fang *et al.* (2008).

2.2.5 Risk Scores

I use data on each enrollee's chronic medical conditions to calculate risk scores using the RxHcc Risk Adjustment Model developed by CMS. This model was created to adjust CMS's subsidies to insurance companies offering Part D plans. The scores are non-negative numbers normalized to be one for the average risk score in the Medicare population. Individuals with higher scores have higher expenditure risk.[9]

[8]The MCBS operates as a rotating panel survey with rich information on the demographics of people in Medicare.

[9]See Robst *et al.* (2007) for more details about the model.

2.2.6 Identifying Couples and Singles

To the best of my knowledge, this is the first paper to identify couples in CMS administrative data. I define a "couple" as a pair of beneficiaries, one female and one male, who have the same last name and who share the same residential ZIP+9 code during the same year. The rationale for the matching algorithm is simple. First, zip-9 codes are close to street addresses in terms of spatial precision; each code corresponds to a single mail delivery point such as a unique address, one floor of an apartment building, or one side of a street on a city block. Equally important, only 17% of women who married in the 1970s kept their maiden names, (Cain and Derek, 2015). This was a spike relative to prior and future years given the rise of the feminist movement. Moreover, according to the US Census Bureau, the median age of first marriages in 1970 was 20.6 years old for women. A woman who was 65 years old in my study period, 2006-2010, was at least 30 years old in 1970 implying that most of them got married before 1970. Thus, the majority of women in my sample, if married, took their husband's last name.

I rely on several variables to identify singles separately from individuals who are married to someone who was not present in my 20% random sample of beneficiaries. First, to identify widows and divorced women I use the BIC variable described earlier. The BIC is not ideal to identify single men since most men, whether single or married, claim their own social security benefits. Most of my sample of widowers are identified using the death dates of wives of men that are determined to be married according to my algorithm. Finally, I augment the sample of divorced men using the MCBS sample.

Table 2.2 assesses the performance of my matching algorithm by comparing the total number of couples that I identify in the Medicare Part D sample each year with the total number of expected matches given the demographic information in MCBS. According to MCBS, 54% of people in Medicare Part D are married. Using this fact and the 20% random sample of Medicare Part D enrollees, I use a simple back-of-the-envelope calculation to estimate how many couples I should expect to observe in my sample. Therefore, the statistical prediction for the number of matches for each year is $N_y 0.54 * 0.2 * 0.54$ where N_y is the size of the random sample of individuals in Part D each year.

Table 2.2: Couples in Medicare Part D Sample: MCBS vs Algorithm

	2006	2007	2008	2009	2010
Back-of-the-envelope	74,564	77,919	78,292	80,242	80,005
Number of matched couples	73,500	76,696	78,867	82,466	82,880

Notes: Table 2.2 compares the number of matches I should expect according to MCBS ("Back-of-the-envelope") and the final number of couples in my sample following the algorithm.

The algorithm comes remarkably close to matching the statistical prediction. This makes sense given that for this cohort the majority of wives took their husband's last name.

Finally, table 2.3 summarizes the demographic variables for each type of household. The shares on marital status reveal an interesting feature about this market; while 54% are married couples, roughly 30% are widows. This means that the preferences and behavior of these two types of households are likely to drive market outcomes.

Table 2.3: Summary Statistics: Demographics in 2006

	singles	wife	husband
age (mean)	80	73	76
white (%)	95	94	94
total CC (mean)	8	7	7
risk score (mean)	1.04	1.01	0.98
male (%)	4		
divorced (%)	7		
widow/er (%)	92		
observations by year	148,245	68,549	68,549

Notes: Table 2.3 shows summary statistics of main demographics variables of Medicare Part D enrollees. The first data column shows demographic variables, marital status and health conditions of single individuals. "total CC" shows the mean number of chronic conditions for each type of household. The the last two columns describe demographics variables and health conditions of members of married couples.

The table shows that singles have worse health on average. This is not surprising since they are on average older than spouses. The demographic characteristics of enrollees will be important to understand not only their costs but also their preferences, e.g. risk aversion. Importantly, insurance companies are not allowed to price age or any other characteristic of enrollees in this market.

2.3 Descriptive Evidence

2.3.1 Inertia and Pooling

Inertia is a common feature of consumer choice in many markets, (Samuelson and Zeckhauser, 1988). In Medicare Part D, Ketcham *et al.* (2016), show that 90% of individuals choose their default plans each year. Table 2.4 shows the share of households who choose their default plan by household type from 2007 to 2010. Consistent with prior literature, the fraction of households who stick with their previous choice is on average 90%.

Divorced men are the most reluctant to switch, with 97% of them choosing the default plan. For couples, the figure is slightly smaller: 85%. In other words, 85 percent of couples decide to keep the same plan combination as last year. I also calculate the share of couples who enroll in the husband's default plan (but not the wife's) or who enroll in the wife's default plan (but not the husband's). Both figures are 3%.

Table 2.4: Inertia in Medicare Part D (2007-2010)

Inertia	Share (%)
singles choosing default plan	92
divorced women	92
divorced men	97
widow	91
widower	92
couples choosing default plan combination	85
only wife choosing default plan	3
only husband choosing default plan	3

Notes: The first six rows of the table table report the fraction of enrollees by type of household who re-enroll in their default plans. The last two rows show the fractions of couples where only one member chooses the incumbent plan.

The table reveals that households types are fairly homogeneous in terms of inertia. This is striking since they are likely heterogeneous in terms of costs and preferences. The fact that heterogeneous households switch at the same rate underscores the pervasive nature of this behavior.

Table 2.5 shows the share of couples who enrolled in the same plan. Overall, 78% of couples buy the same plan. In principle, this could be explained by partners aging into Medicare in different years and the younger spouse choosing the older spouse's default plan. However, the second row shows the same pattern with 76% of couples pooling in 2006, the first year of the program in which couples entered the market and purchased their initial plans simultaneously. Strikingly, this figure hardly changes when I focus on couples with different health needs. The last four rows of the table divides couples based on the similarity of the spouses' risk scores. "Same Risk" describes couples in the same quartile of the distribution of risk scores. "Adjacent risk quartiles" are couples where spouses are in adjacent quartiles. "Nonadjacent risk quartiles I" are couples where the spouses are either in the first and third quartiles or in the second and fourth quartiles. "Nonadjacent risk quartiles II" are couples where one spouse is in the first quartile and the other is in the fourth quartile. Moving down the last four rows shows that the rate of pooling hardly changes as spouses increasingly differ in terms of their prescription drug risk. The fact that more than three quarters of spouses with substantially different prescription drug risks decide to buy the same plan suggests that this behavior is related to causes beyond health needs and costs.

Table 2.5: Couples' Tendency to Choose the Same Plan

Pooling	Share (%)
couples pooling	78
couples pooling in 2006	76
same risk	80
adjacent risk quartiles	77
noadjacent risk quartiles I	76
noadjacent risk quartiles II	75

Notes: The table shows the share of couples who decide to buy the same plan. "Same Risk" corresponds to couples in the same quartiles of the distribution of RxHCC risk scores. "Adjacent risk quartiles" are couples where spouses are in different adjacent quartiles. "Nonadjacent risk quartiles I" are couples where one spouse is in the first, or second quartile and the other spouse is in the third and forth quartile respectively. "Nonadjacent risk quartiles II" are couples where one spouse in the first quartile and the other is in the forth quartile.

Why do so many couples buy the same plan? A vast literature studies why individuals tend to choose their status quo options. Starting with Samuelson and Zeckhauser (1988), the literature has tended to divide mechanisms into two main categories: rational decision-making and cognitive misperceptions. Examples of the former are costly information acquisition or uncertainty about plan features. Both examples imply that plans must be discovered, leading to search rules and cutoff strategies. An example of the second category is loss-aversion, with the incumbent plan being the reference point such that losses from switching will be weighted more than gains from the same action.

The same two sets of mechanisms can explain why couples tend to buy the same plan. There are several reasons why it may be optimal for spouses to choose the same plan. For example, if couples have the same preferences and risk or they are specially matched, e.g. highly risk averse wives married to high cost husbands, then within-household plan "pooling" may emerge endogenously. Bargaining within the household and risk-sharing can also explain why couples buy the same plan. For example, if one spouse is more risk averse than the other, he could agree to buy a less generous plan if the less risk averse spouse is willing to bear most of the risk. In this scenario, the less risk averse spouse sacrifices more private consumption in the bad state and consumes more in the good state. Division of tasks within the household may also explain this behavior. If the couple splits duties to save time and effort, the spouse who is in charge of choosing insurance plans may find it optimal to choose a single plan for both spouses if searching is costly for the same reasons that lead to inertia. In terms of cognitive misperceptions, a possible explanation is the convergence of beliefs of each spouse about their own risks. This may occur, for example, if a specific chronic condition afflicting one member has salience effects over the other spouse, (Fadlon and Nielsen, 2019). I do not attempt to identify the relative importance of the various mechanisms that cause pooling. Instead, I measure how large the combined effect of these mechanisms must be to rationalize couples' observed choices. Identifying model parameters designed to capture the tendency to pool requires characterizing the distribution of risk aversion among single and couples, ex ante differences in expected costs, and rules for risk sharing within households.

Finally, I want to preview the importance of this new finding for evaluating prospective policy reforms. In the policy section, I will investigate the market consequences of a policy that helps consumer make better informed choices. This counterfactual policy, is motivated by the fact that 90% of consumers reenroll in their default plans each year and leave money on the table by doing so, (Handel, 2013; Polyakova, 2016; Ketcham *et al.*, 2019). What is not clear, is whether a policy that improves information will reduce the high degree of pooling among couples. Will spouses each buy different plans, will they jointly switch to a different plan but still choose the same one or will they stay in their default plans? The three actions may have very different implications for market functioning and welfare. How couples behave in this situation will depend on the relative importance of the different mechanisms that underlie pooling. If couples are choosing the same plan because search costs preclude them comparing more plans, then a policy intended to target inertia may also impact pooling. On the other hand, if couples are choosing the same plan because it is optimal given their combined preferences and expenditure risk that they face, then they may still decide to continue to pool.

2.3.2 Household Costs and Residual Costs

Table 2.5 suggests that assortative mating on prescription drug use does not explain why couples decide to buy the same plan. Nevertheless, understanding the degree of assortative mating may be important for policy. In the education literature for example, assortative mating on educational attainment is of interest because it has implications for household income inequality, (Fernández *et al.*, 2005; Eika *et al.*, 2019). Similarly, assortative mating in health may contribute to household health inequality, (Fleurbaey and Schokkaert, 2011).

Further, predicting market outcomes and welfare consequences of policies that nudge consumers toward different choices requires knowing the costs of couples and how they are related to costs of other type of households.

To better understand the degree of assortative mating in prescription drug expenditure, I compare three different measures in 2006. First, I assign each spouse to their corresponding quartile of the distribution of risk scores at the beginning of the year. Then I compare the Pearson correlation coefficient of these variables with the correlation coefficient estimated for randomly assigned couples from the same geographic area. Panel A of Table 2.6 shows the correlation of risk scores for actual couples, the correlation of risk scores for random couples from the same state and the correlation of risk scores for random couples from the same zip-5 code. Panel B of the table shows the same statistics but calculated based on quartiles of total cost under each plan in 2006. The two measures differ in that the second measure is affected by the choice of plan while the first is not.

Table 2.6: Assortative Mating - Prescription Drug Expenditure

Actual couples	Random couples	
	same ZIP5	same state
	A. Risk Score Correlation	
0.147	0.061	0.011
	B. Prescription Drug Expenditures Correlation	
0.261	0.104	0.018

Notes: Panel A shows the Pearson correlation coefficient of risk scores of actual couples, the correlation of risk scores of randomly matched couples from the same state and the correlation of risk scores of randomly matched couples from the same zip-5 at the beginning of 2006. Panel B of the table shows the same statistics but calculated with quartiles of total costs under each plan in 2006.

The correlation coefficient for actual couples is 2.5 times larger than the correlation coefficient for randomly matched couples from the same zip-5, and 13 times larger than the correlation of random matches from the same state. The correlation among actual couples is small relative to the measures of assortative mating estimated in the education literature. In Fernández *et al.* (2005) for example, the Pearson correlation coefficient describing assortative mating in education in different countries ranges from 0.32 to 0.76. Further, compared to the education setting my estimates are more likely to be increased by changes in behavior after marriage that would increase the estimated degree of assortative mating in health.

Predicting how the market will evolve after a specific policy also requires understanding how the costs of married couples compare to the costs for other types of households. As shown in Table 2.3, widows constitute the majority of single households. According to MCBS, they represent 30% of the market. Table 2.7 compares the distribution of costs of wives, husbands, couples and widows. Costs are measured as total prescription drug expenditures minus the out of pocket expenditure of each household. This difference represents the cost that insurance companies incur.

Table 2.7: Moments of the Cost Distribution for Spouses, Couples, and Widows

Costs	wife	husband	couple	widow
10 percentile	0	0	20	2
25 percentile	56	55	195	207
median	448	497	650	813
75 percentile	1,271	1,359	1,165	1,583
90 percentile	1,845	1,890	1,702	2,020
mean	868	873	871	1,116

Notes: Table 2.7 compares the distribution of costs of wives, husbands, couples and widows. Costs are measured as total prescription drug expenditures minus the out of pocket expenditure of each household. This difference represents the cost that insurance companies incur.

The table shows that the average cost of wives and husbands is smaller than the cost of widows. If the majority of couples buy the same plan, then we can think that the evolution of the market will be driven by the behavior of these two types of households, with married couples being the "good type" and widows being the "bad type" in the sense of Rothschild and Stiglitz (1976). Importantly, since the average cost of widows exceeds the average costs of each spouse, the presence of widows in the market makes the degree of assortative mating less important for predicting market outcomes when couples pool. In general, the difference in the cost of widows relative to the cost of couples will be positive, regardless of their decision to pool. Widows are on average $245 more costly than married couples.

As noted earlier, the process for risk adjusting payments to insurance companies is an important feature of the Part D markets. The costs shown in Table 2.7, measure the total costs but not the residual costs that matter for insurance companies' profits, (Layton, 2017). Without risk adjustment, premiums will likely reflect the average cost of the enrollee pool. Layton (2017) shows that in the presence of risk adjustment, premiums will reflect average residual costs, which are defined to be costs that are not predicted by the risk adjustment model. More formally, under risk adjustment the premium of plan j will be:

$$\text{Premium}_j = E(\text{cost}_j) - E(\text{RA}_j) + E(\text{cost}) \tag{2.1}$$

Here $E(\text{cost}_j)$ reflects the average cost of the enrollees who selected plan j, $E(\text{RA}_j)$ the average risk adjustment payments of the pool and $E(\text{cost})$ reflects the average cost of the market. Equation 2.1 implies that the difference in premiums will reflect the difference in average residual costs, $E(\text{cost}_j) - E(\text{RA}_j)$. Prior studies have shown that the risk-adjustment model that CMS uses for Part D (RxHCC Model) tends to overpredict costs for beneficiaries with low actual costs, and underpredict costs for beneficiaries with high actual costs, (Hsu *et al.*, 2009). These systematic prediction errors will likely distort the differences in residual costs relative to total costs. Table 2.8 illustrates this point by showing the average costs and residual costs of couples and widows, and the differences between them. With imperfect risk adjustment payments, the difference in residual costs is 35% higher than the difference in total costs.

Table 2.8: Residual Costs for Couples vs Widows

	Total Costs			Residual Cost		
	couple	widow	difference	couple	widow	difference
mean	871	1,116	245	121	453	332

Notes: Table 2.8 shows the average costs and residual costs of couples and widows and the differences between them. Residual costs are the costs that are not predicted by the risk adjustment model.

This difference in total costs and residual costs is important because it may exacerbate adverse selection if couples and widows select different plans. In section 6, I explore the consequences of this difference in the context of a policy that nudges consumers toward different choices. Importantly, costs are just one factor in the calculations. The preferences of these two type of households also matter for predicting the evolution of the market and how it responds to different regulations. More specifically, the distribution of risk aversion across households is crucial for understanding how households will react.

2.4 Empirical Framework

2.4.1 Model

The Choice Set

Consumers are modeled as choosing lotteries of prescription drug expenditure, where a lottery is defined by a distribution of prescription drug expenditures under all possible health states of the world and a set of probabilities for realizing those states. To construct these lotteries I rely on several variables contained in the CMS Administrative records, including the diagnosis dates of more than 30 chronic conditions from CMS Chronic Condition Data Warehouse. Specifically, I use knowledge of the health conditions and demographics of each enrollee, to calculate their individual risk scores using the RxHcc Risk Adjustment Model developed by CMS.[10] I define a "cell" as a set of individuals with the same risk score in year T-1 who live in the same CMS region. Ex-ante distributions of out-of-pocket (*oop*) expenditures of each plan and type in year T are generated with the realized *oop* costs in year T of all beneficiaries that belong to the same cell. This means that the *oop* expenditure of each beneficiary is a possible state of the world for all beneficiaries that belong to the same cell. This actuarial method of creating oop distributions has been used extensively, (Pauly and Zeng, 2004; Abaluck and Gruber, 2011; Handel, 2013; Ketcham *et al.*, 2016).

In order to construct the ex-ante distributions of *oop* expenditures of every plan-person match, I make the standard assumption of no moral hazard. I use a cost calculator developed by Ketcham *et al.* (2015) to construct counterfactual out-of-pocket expenditures for the bundle of drugs that each beneficiary purchased under all of the plans in the beneficiary's choice set. This is essentially the same as the approaches used in Abaluck and Gruber (2011), Abaluck and Gruber (2016), Ketcham *et al.* (2016), and Ketcham *et al.* (2019). The no moral hazard assumption in these studies is justified by the small drug-specific price elasticities estimated in the literature and the high persistence of drug use, both of which are indicators of moderate moral hazard, (Abaluck *et al.*, 2018). As long as the presence of moral hazard is mild the estimated distributions will approximate the true distributions.

The oop expenditure of each beneficiary can then be used to construct the empirical CDF for plan j and type t as follows:

$$\hat{F_{jt}}(x) = \frac{1}{n_t}\sum_{i=1}^{n_t} 1(X_{ij} \leq x) \tag{2.2}$$

[10]The two demographic variables that enter in the RxHcc model together with the chronic conditions, are age and gender.

In the equation n_t is the number of observations of type t, i.e. the number of people that belongs to that cell, x is a non-negative number that belongs to the support of the distribution of out-of-pocket expenditure and $1(X_i \leq x)$ is an indicator function for whether realization *i* in plan j is less than x.[11] I assume that the distributions of out-of-pocket expenditure implied by each plan and type belong to the bounded and common support $[a, b]$.[12] If a plan has a smaller realized support I define the density function of that distribution to be zero outside this range.

To construct the distributions of out of pocket expenditures for couples I use copula methods.[13] Intuitively, a copula expresses a joint distribution as a function of marginal distributions and a correlation parameter. I use Gaussian copulas to generate the bivariate distribution of oop for the couple as a function of the marginal distribution of each spouse. The Gaussian copula form is $C(u_1, u_2, \rho) = \Phi_b(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)$, where Φ_b is the CDF of the standard bivariate normal distribution and Φ^{-1} is the inverse CDF of the standard normal distribution. The copula is a function of a correlation parameter ρ and $u_i = F_i(x)$, the range of the marginal CDF distribution function of each spouse. Appendix A provides a detailed description of the steps used to construct the distribution of couples' oop expenditures and estimate ρ.

Under this formulation, the only parameter to be estimated is ρ.[14] ρ captures the joint dependency of both distributions $F_i(x)$ which could, in principle, be very different from

[11] Like Handel (2013) I require the minimum cell size to be 75 individuals.

[12] Here a will be the smallest realization of out-of-pocket expenditure among all plans and types while b will be the maximum. I will discretize the support in r pieces as is often done when working with empirical CDFs.

[13] See Trivedi and Zimmer (2006) for an introduction to copula methods in economics.

[14] We should not interpret this parameter as a reflection of assortative mating. The later would be captured by how similar are the $F_i(x)$ of each couple.

each other. An intuitive example of what ρ measures is the "broken heart syndrome" in life insurance, in which the death of a spouse reduces the survival probability of the other spouse, (Denuit *et al.*, 2005). For tractability, I assume that ρ is constant across diseases. There is medical evidence that people with partners that have ischaemic heart disease, diabetes, or experienced a stroke have no increased risk of contracting the disease themselves. While this is not true for asthma, depression, and hypertension, these diseases are likely to be less expensive to treat (Hippisley-Cox *et al.*, 2002). I estimated $\hat{\rho}$ to be equal to 0.3 (p=0.0000). This means that couples' risk exhibits a positive dependence, consistent with the intuition for broken heart syndrome.

I follow standard practice in assuming that the relevant distributions from which households are choosing are the ones estimated by the researcher. This assumption could be violated for two reasons. The first is private information. Specifically, the estimated F_{ij} and the true distributions could differ if enrollees possess information that is not available to the researcher. This is relatively minor concern for my analysis because I observe detailed information on chronic condition diagnoses and other demographic characteristics that together with the RxHCC software allows me to calculate risk types for each enrollee. Further, I focus exclusively on a very specific type of medical risk: prescription drug expenditures. Estimating ex-ante distributions of plans that cover many types of medical services like surgery, hospitalizations and doctor visits would require more information. A second reason why F_{ij} could be substantially different from the distribution determining consumer choice is because of systematic differences in subjective beliefs. Without a complementary survey eliciting subjective beliefs of prescription drug expenditures, it is impossible to assess the validity of this assumption. Thus, I maintain the assumption that subjective beliefs match objective beliefs.

Preferences

Financial markets exist because people have different tastes for risk. Thus, to understand how demand for insurance will change in response to a prospective policy it is essential to know the degree of risk aversion, how it is distributed across households and the extent to which it is correlated with household risk.

How individuals feel about taking risks is also necessary to understand how groups, like couples, make joint decisions in risky environments. Within the tradition of methodological individualism it is individuals and not groups, who are presumed to have preferences.[15] How groups make choices depends on the preferences of their members and how their members interact with each other. The interaction could be by a voting rule, within household bargaining or any other mechanism.

To understand how different types of households decide among insurance plans and how policies will affect their decisions, I represent household behavior with a stochastic choice model. In other words, consumers' choices will be consistent with a deterministic core theory, specifically expected utility, plus a shock, (e.g. Harless *et al.*, 1994; Hey and Orme, 1994; von Gaudecker *et al.*, 2011). The shock is meant to represent errors in the decision process of households when comparing two plans.

For the deterministic component of choice I use exponential utility which implies constant absolute risk aversion (CARA). This parametric form has many advantages. First, it allows the household to be modeled as a representative agent where the household's risk aversion parameter is the harmonic mean of each spouses risk aversion parameter divided by two. This representation is observationally equivalent to a collective model of the household, (Chiappori, 1992). It is also equivalent to group preferences under uncertainty

[15] See Chiappori and Mazzocco (2017) for a discussion of methodological individualism and its implications in the unitary approach of household preferences.

as in Harsanyi (1955) with constant Pareto weights across the study period.[16] Importantly, if couples share risk efficiently, their choices will conform with this utility specification, (Bone, 1998).

More specifically, if both members of the couple have CARA preferences with risk aversion parameters σ_w and σ_h, then the couple can be represented with a utility function of the following form:

$$u_c(x) = -e^{-\sigma_i^* x}, \text{ where } \sigma_i^* = \frac{1}{\frac{1}{\sigma_w} + \frac{1}{\sigma_h}} \tag{2.3}$$

This representation captures the fact that group risk aversion is derived from individual risk aversion. Another advantage of this particular parametric form for utility is that it will allow me to compare my results directly with previous studies that used CARA specifications to depict household-level choices among employer sponsored insurance plans that offer coverage for employees, spouses, and their children, (Handel, 2013; Handel and Kolstad, 2015).

If consumers adhere completely to expected utility, then when they first enter the market they will choose the plan that maximizes the certainty equivalent (CE).[17] However, the random element added to their utility function implies that households may choose plans of lower value. For example, when comparing two plans j and j', household i will select plan j whenever:[18]

$$\text{CE}(L_{jhi}, \sigma_i) - \text{CE}(L_{j'hi}, \sigma_i) + \lambda_h \epsilon_{jj'hi} \geq 0 \tag{2.4}$$

[16]In general, additional information on private consumption of each member of the couple and changes in distribution factors are needed to identify Pareto weights, (Chiappori and Mazzocco, 2017).

[17]With exponential utility, $\text{CE}(L, \sigma) = \frac{1}{\sigma}\ln\Big(\text{E}(e^{\sigma x})\Big)$.

[18]In equation 2.4 the plan premium is included in the certainty equivalent.

$\epsilon_{jj'hi}$ are independent household plan logistic shocks, i indexes the household (which may be comprised of an individual or a couple), h indexes the household type, e.g. couples or singles. For a couple, L_{jhi} represents the lottery associated with plan combination j, given the couple's joint distribution of potential health shocks.

Insurance plan enrollment is repeated each year. As seen above, during open enrollment most households default into their incumbent plans and most couples enroll in the same plan. Equation 2.5 shows that households will choose the default plan combination j over j' if:

$$\text{CE}(L_{jhi}, \sigma_i) + K_h 1_{d=j} + 1_{j=p}\Omega - \text{CE}(L_{j'hi}, \sigma_i) - 1_{j'=p}\Omega + \lambda_h \epsilon_{jj'hi} \geq 0 \tag{2.5}$$

$1_{d=j}$ is an indicator if the plan or plan combination is the household's default plan. Similarly, $1_{j=p}$ is an indicator for whether couples choose the same plan. This formulation highlights an advantage of using certainty equivalents as a cardinalization of preferences. It allows us to measure the values that particular household types attribute to status quo plans and plan-pooling in monetary terms using the parameters K_h and Ω.[19] In particular, the way I am measuring status quo inertia is identical to Handel (2013) and Handel and Kolstad (2015).[20]

The interpretation of λ_h, the scale of the logistic shock, is important and related to the interpretation of $\epsilon_{jj'hi}$. In random utility models, $\epsilon_{jj'hi}$, is typically used to represent la-

[19] Equation 2.5 can alternatively be represented similarly to a random utility model: $\beta_{0h}\text{CE}(L_{ji}, \sigma_i) + \beta_{1h}1_{d=j} + \beta_{2h}1_{j=p} - \beta_{0h}\text{CE}(L_{j'i}, \sigma_i) - \beta_{2h}1_{j'=p} + \epsilon_{hi}$ with $\frac{\beta_{1h}}{\beta_{0h}} = K_h$, $\frac{\beta_{2h}}{\beta_{0h}} = \Omega$ and $\beta_{0h} = \frac{1}{\lambda_h}$.

[20] They define inertia in terms of a "bidding price". However, given that they also use a CARA specification and assume independence between the distribution of out of pocket expenditures and inertia, both measures coincide, (Pratt, 1964).

tent attributes that provide utility to households. Under this interpretation λ_h determines the relative importance of expected utility for household decision-making compared to all other non-modeled attributes. If, however, we treat EU as a normative decision-theory in the sense that households should behave as EU maximizers then $\epsilon_{jj'hi}$ are interpreted as a "mistake" and λ_h as a measure of conformity with expected utility. In my main specifications $\epsilon_{jj'hi}$ is used to represent errors in the decision process. This interpretation is meant to facilitate the counterfactual scenarios I explore later in the paper, in which a regulator who cares paternalistically for consumers welfare will attempt to nudge consumers to conform with expected utility by introducing a generic policy that reduce the magnitude of λ_h.

2.4.2 Identification

Risk aversion

I first estimate a distribution of individual-specific parameters describing absolute risk aversion as a function of demographics and prescription drug spending using data from the Health and Retirement Study (HRS). Then I use observable demographics to project this distribution onto the Medicare Part D population. I use a set of questions in the HRS that were designed to elicit risk aversion by asking each individual to choose among hypothetical monetary gambles.

The HRS data allow me to overcome the problem explained in Apesteguia and Ballester (2018) that risk aversion cannot be identified in many discrete choice settings. The problem is that CARA and CRRA preferences embedded in stochastic choice models can generate the same choice probabilities with different values of risk aversion. Moreover, estimating the risk aversion level of each household member is necessary for predicting how choices would change if individuals were to choose in isolation. Thus, I assume that the hypothetical gambles that were used to elicit risk aversion parameters for a random sample of individuals in HRS capture spouses' levels of risk aversion in scenarios where they can't share risks. This assumption is consistent with the design of the survey questions, which are described in more detail below.

Using stated-preferences methods to elicit risk aversion parameters, such as hypothetical gambles, instead of revealed-preferences methods has well-known trade-offs, (Diamond and Hausman, 1994; Beshears *et al.*, 2008; Mata *et al.*, 2018). While stated preferences methods are usually better in controlling for possible cofounders, revealed preference methods are often thought to perform better in real-world scenarios that are difficult to represent on a survey or in the laboratory. Given these tradeoffs, many studies have investigated the validity of HRS risk measures and have found a strong relationship between these measures and individuals' financial decisions, (Mazzocco, 2004; Kimball *et al.*, 2008). Finally, my approach to transferring risk aversion parameters from the HRS to CMS assumes that the levels of risk aversion for individuals are constant across domains. So that risk aversion parameters elicited with monetary gambles can be used to assess risky choices of health insurance plans. While there is some debate over this assumption (Barseghyan *et al.*, 2011), Einav *et al.* (2012) find that individuals' willingness to take risk relative to their peers remains stable across domains.[21]

I elicit risk aversion parameters using the following questions that were asked in the 2004 HRS wave, two years prior to the introduction of Medicare Part D:

Suppose you have an additional USD 10,000 saved for the future. You can choose to invest this money one of two ways. One is to invest in a government bond that will be worth

[21] This result is important for the present paper because I will be comparing status quo inertia for different types of households: widows, widowers, divorced women, divorced men, and couples. Importantly, the monetary estimates will depend on the different levels of risk aversion of each type of household. As long as risk preferences, relative to these demographic groups, are stable across domains; the relative size of status quo inertia will be stable as well.

USD 10,000 in two years for sure. The other way is to invest in a mutual fund that may increase or may decrease in value in the next two years. On average the mutual fund will be worth 20,000 in two years, but has a 50-50 chance of being worth USD 5,000 and a 50-50 chance of being worth USD 35,000. Would you invest your money in the government bond that guarantees you USD 10,000 or in the mutual fund I have just described?

Individuals who choose the riskier option, were then asked:

Suppose instead that the average return on the mutual fund is lower. On average the mutual fund will be worth USD 15,000 in two years, but has a 50-50 chance of being worth USD 5,000 and a 50-50 chance of being worth USD 25,000. Would you invest your money in the government bond that guarantees you USD 10,000 or in the mutual fund I have just described?

If the individual opted for the risk-free option in response to the first question, he would then be asked:

Suppose instead that the average return on the mutual fund is higher. On average the mutual fund will be worth USD 25,000 in two years, but has a 50-50 chance of being worth USD 5,000 and a 50-50 chance of being worth USD 45,000. Would you invest your money in the government bond that guarantees you USD 10,000 or in the mutual fund I have just described?

This procedure identifies lower and upper bounds on the absolute risk aversion parameter. Following the approach described in Barsky *et al.* (1997), Kimball *et al.* (2008).[22]

After using respondents' answers to assign them to mutually exclusive categories, I use their resulting bin assignments to estimate a continuous distribution of risk aversion. I assume that the distribution of risk aversion is log-normal: $\log\sigma \equiv x \sim N(\mu, \phi)$, with

[22]They use relative gambles that were designed to elicit relative risk aversion coefficients for CRRA preferences. The only difference is that I use a different set of questions.

the mean $\mu = \mu_0 + \gamma_1 X + \gamma_2 M$ being a function of demographics, X, and different bins of prescription drug expenditure, M. Gender and age will be included in X and the prior year's total expenditure on prescription drugs will be included in M.[23] The probability of being in category j is then:

$$P(c = j) = P(\log\sigma_{lj} < x < \log\sigma_{uj}) \tag{2.6}$$

$$P(c = j) = \Phi((\log\sigma_{uj} - \mu)/\phi) - \Phi((\log\sigma_{lj} - \mu)/\phi), \tag{2.7}$$

where Φ is the cumulative normal distribution function and σ_{lj} and σ_{uj} denote lower and upper bound of absolute risk aversion of category j. I estimate $\mu_0, \gamma_1, \gamma_2$ and ϕ via maximum likelihood.

Inertia and Plan Pooling

The identification of status quo inertia relies on two main sources of variation in the data. Two sets of enrollees serve as control groups for people who have a status quo plan in their menus. The first group is new enrollees. As noted by Samuelson and Zeckhauser (1988), the active choices of new enrollees capture what the choices of old (and similar) enrollees would have been absent the status quo plan. I have detailed data on chronic conditions for old enrollees and for most of the new enrollees as well. This is one of the advantages of having data in the early years of this market. However, it is important to distinguish

[23] I group people in four different bins of prescription drug expenditure: people whose annual spending was below \$50, people whose spending was between \$50 and \$500, between \$500 and \$2,500 and people whose annual spending was above \$2,500.

enrollees that are new to Part D and enrollees that are new to the entire Medicare system (Part A and B). Because chronic conditions diagnoses are collected for all enrollees who are already in Part A and Part B, I can construct ex-ante distributions of oop for enrollees who are new to Part D but not to the rest of Medicare. This group is formed by enrollees who enrolled late in the market, after turning 65 years old. The second control group is composed of enrollees who are forced to choose actively because their incumbent plan was discontinued. These two groups constitute the "active" choosers who lack a default plan. The second feature of the data that allows me to identify status quo inertia is the continuing change in plans' menus that happen each year. This was depicted in Table 1. New plans enter the market each year and some old plans exit.

Finally, "willingness to pool" is identified by the active choice of new couples, e.g. the choice in 2006, and couples who switch plans. This is, each year new couples express their preferences for choosing plan combinations with the same pair of plans or choosing plan combinations with different plans. The strength of their preferences for pooling beyond expected utility will be captured by Ω_h.

2.5 Results

2.5.1 Risk Aversion

Table 2.9 reports maximum likelihood estimates for heterogeneity in individual risk aversion from equation 2.7. A casual interpretation of the estimated coefficients is unnecessary because the purpose of this exercise is ultimately to project the demographic variation in individual risk onto the medicare population. Nevertheless, the coefficients on demographic variables are broadly consistent with causal estimates from previous literature. Women appear to be more risk averse than men, a finding that has been documented several times in different environments, (Borghans *et al.*, 2009). There is less agreement in the literature on how age affects risk aversion. Cohen and Einav (2007) document a U-shaped relation between age and risk aversion across the life cycle while Dohmen *et al.* (2011) find a positive slope. Both findings are consistent with the positive slope that I estimate for the final years of the life cycle.

Interpreting the coefficients on medical expenditures and comparing them to prior studies is more complicated. First, my estimates could reflect some reverse causality. That is, people who are less risk averse may take less precaution in their daily life choices, like eating healthy food and exercising, which could result in them requiring more medical services. Second, these variables could reflect the medical risk that each bin is exposed to, and impact risk aversion through this background risk channel.[24] It could also represent health shocks that the individual suffered in the previous year, and impact risk aversion through this channel. The empirical literature established that both channels affect risk aversion, (Courbage *et al.*, 2018; Decker and Schmitz, 2016).

In any case, the positive correlation between prescription drug expenditure and risk aversion implies that more costly enrollees have a higher willingness to pay for insurance, because of higher monetary expenses and because they are more risk averse. This positive correlation between risk and risk aversion is also present in the auto insurance setting of Cohen and Einav (2007), whereas Finkelstein and McGarry (2006) document a negative correlation in markets for long term care insurance.

[24]Recall that the hypothetical gambles that each individual responds to in the HRS are not specifically about medical expenditures. So from this perspective, prescription drug expenditures are a background risk, (Gollier *et al.*, 2001).

Table 2.9: Estimates for Demographic Heterogeneity in Risk Aversion

Parameter	Estimates
constant	-15.79***
	(1.12)
male	-2.02***
	(0.25)
age	0.05***
	(0.01)
prescription drug expenditure I: (USD 50-500)	1.01*
	(0.58)
prescription drug expenditure II: (USD 500-2500)	2.55**
	(0.59)
prescription drug expenditure III: (> USD 2500)	2.58***
	(0.59)
standard deviation	5.27***
	(0.10)
Observations	451

Notes: Table 2.9 shows maximum likelihood estimates (standard errors) of parameters describing heterogeneity in absolute risk aversion based on responses to survey questions in the HRS. The estimates represent the influence of demographic characteristics on the mean and median absolute risk aversion of the HRS population. "Prescription drug expenditure categories I, II, and II" corresponds to different bins of prescription drug expenditures in the previous year.

Finally, I project the median absolute risk aversion parameter for each demographic group and medical expenditure bin onto the Medicare sample.[25] The mean absolute risk aversion of the Medicare population is .0000369 and the median is .0000131.[26] For an economic interpretation of these estimates, Figure 1 shows the distribution of implied risk premia when individuals face a hypothetical gamble in which they can win or lose $900 with the same probability. This hypothetical gamble is scaled to approximately capture the risk an average individual is facing in Medicare Part D, where $900 is the standard deviation of out of pocket cost in the first five years of the market. The figure shows the distribution of

[25]Given my assumption of a log-normal distribution, the median is a better representation of central value tendency than the mean.

[26]This estimates are similar to previous studies, (Cohen and Einav, 2007).

risk premiums as a fraction of \$900. For a risk-neutral individual, this number is zero and for someone who is extremely risk averse it is 1. Although most individuals have moderate values for the risk premium, the distribution is right-skewed suggesting a high degree of heterogeneity.

Figure 2.1: Distribution of Risk Aversion Among the Part D Population

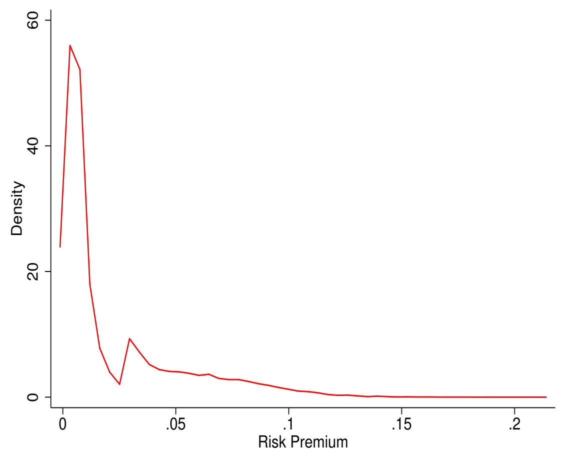


Figure 2.1: Figure 2.1 shows the projected distribution of risk premiums in the Medicare Part D population. The risk premium is expressed as a fraction of the \$900 gamble. For a risk-neutral individual this number is zero and for someone who is extremely risk averse it is 1.

The estimated distribution of risk aversion parameters allows me to test the hypothesis that couples tend to pool because people who are highly risk averse tend to be married to partners who are relatively sick, inducing both partners to optimally choose high coverage plans. I test this hypothesis by calculating the following correlation:

$$\text{corr}(\sigma_w - \sigma_h, cost_w - cost_h) \tag{2.8}$$

A negative correlation means that spouses who are more risk averse are in general married to spouses with higher costs, and vice versa. The estimated correlation coefficient is 0.085, allowing me to reject the hypothesis.

2.5.2 *Inertia and Pooling*

The following table shows the estimates for the parameter describing the relative importance of random factors driving choices (λ_h) together with inertia and the value of plan pooling in certainty equivalent terms. I report estimates for the four types of households that represent the largest fractions of consumers in the market; married couples (54%), widows (30%), widowers (6%) and divorced women (5%).

Table 2.10: Type-specific Estimates for Status Quo Inertia and the "Willingness to Pool"

	Couples		Widows		Divorced Women		Widower	
	Estimates	$P > \|Z\|$	Estimates	$P > \|Z\|$	Estimates	$P > \|Z\|$	Estimates	$P > \|Z\|$
λ	372	0.000	306	0.000	230	0.000	272	0.000
K (inertia)	3,152	0.000	1,975	0.000	1,472	0.000	1,754	0.000
$\Omega(pooling)$	1,584	0.000						
couple last year			1,946	0.000			1,871	0.000
Observations	5,078		14,216		1,134		3,828	

Notes: The first row of the table shows the coefficient on the certainty equivalent, or equivalently the inverse of the variance of the shock. The second row of the table shows the size of the status quo bias. Ω repersents the estimated "willingness to pool". The third row shows the estimated "willingness to pool" for couples. In the fourth row, "couple last year" measures status quo inertia for widowed people who were married in the previous year.

The first row of the table shows the scale of the logistic error for each type of household. The higher the λ_h the less the household type conforms with expected utility maximization, given my assumption about the parametric form of utility. An example can help to illustrate the economic implication of these estimates. Imagine there are two plans or plan combinations: A and B, with CE(A, h) $-$ CE$(B, h) = 500$. The estimates of λ_h imply that 22% of married couples will select the lower value plan compared to 16% of widows, 14% of widowers and 10% of divorced women.

The second row of the table shows the money metric estimates for status quo inertia. The magnitude of the estimates for single households is similar to previous studies by Handel (2013), Polyakova (2016), and Ketcham *et al.* (2019). The estimate for married couples is \$3,152, two times the estimate for divorced women. Note that on one hand, couples face a harder problem, in the sense that they have to choose from a menu with more options. If each spouse chooses among 50 plans, couples have to decide among 50 by 50 plan combinations. On the other hand, couples can exploit information economies of scale, (Wilson, 1975) or help each other in the search process. Interestingly, my estimates of inertia for couples are similar to Handel's (2013) largest estimates for families in markets for employer-sponsored health insurance plans ($3,006). This is somewhat surprising, because married couples in Part D have to choose two plans among 50^2 options, whereas families in his setting choose a single plan among 5 options.

Interestingly, the difference between widows and divorced women is quite large. The difference is approximately equal to the average (subsidized) annual premiums in Part D, $500. Recall that these measures are expressed in terms of certainty equivalents so they should be interpreted from an ex-ante perspective. For example, suppose that a widow has to choose between a gamble and a riskless position. If she chooses the gamble she can lose 5,000 dollars with probability p and zero with complementary probability. In the riskless position, she loses zero dollars with certainty. The estimated status quo inertia, $1,975, implies that a widow with average risk aversion will be indifferent between taking the gamble and maintaining a riskless position when $p = 0.35$.[27] This means that plans have to get significantly worse in terms of coverage in order to induce widows to switch to another alternative. The reason why status quo bias is higher among the set of widows is simple. Widows are the most risk averse agents in the market (because they tend to be older than divorced women and have higher medical spending), so they are the ones who stand to gain the most by switching. The fact that they switch at the same rate as less risk averse agents, 10%, can only be rationalized with higher status quo inertia.

The coefficient on "couple last year" measures status quo inertia for widowed people who were married in the previous year to test whether death of a spouse may reduce inertia. The estimates are not different from enrollees who were already widowed last year,

[27] The average absolute risk aversion for widows is 0.00007.

suggesting that death of a spouse does not reduce inertia, at least in the short term.

The third row of the table reports a money metric for couples' implied willingness to pool, Ω. Interestingly, Ω is approximately half the size of K, (\$1,584). The revealed preference logic of the maximum likelihood estimator requires Ω to be smaller than K to rationalize the fact that most couples decide to buy the same plan.

Returning to the coefficients on "couple last year", notice that the estimates conflate the effects of two changes: a year-to-year change in plan menu, and, a change in the preference function of households who were previously choosing plans jointly with their spouse. When compared with the choice of an active widower, both effects must be taken into account. I can not disentangle these two effects because there is no region where plans menus were unchanged from one year to the next between 2006 and 2010. The estimates for men and women are similar to enrollees who were already widowed last year.

Comparing these measures for inertia between couples and singles is consistent with the hypothesis that one member of the couple is in charge of selecting both plans. Under this hypothesis, the estimates for couples should be double the estimates for singles. Another observationally equivalent hypothesis is that both spouses choose their own plans with complete autonomy. However, this second hypothesis seems less likely to drive behavior because it does not seem capable of explaining the high rate of pooling.

2.6 Policy Analysis

The policy counterfactuals envision a regulator who will nudge consumers to conform with expected utility. The counterfactual scenarios simulate how premiums will respond to changes in the way that consumers sort themselves across plans as a consequence of the policy. Section 6.1 describes the institutional details of how premiums and plan payments are set in Part D. Section 6.2 defines "nudge" in the context of my stochastic choice model. Section 6.3 describes the policy counterfactuals. Finally, section 6.4 summarizes results.

2.6.1 Insurance Payments and Premium Subsidies

As noted earlier, plan payments in Medicare Part D are risk adjusted. This means that plan providers are compensated with payments that vary with the chronic condition of their pool. For example, the risk adjustment payment that a plan provider receives for insuring individual i is:

$$\sum_{cc} W_{cc} D_{icc} \,, \tag{2.9}$$

where W_{cc} is the risk adjustment payment for chronic condition cc and D_{icc} is a dummy variable equal to one if individual i has chronic condition cc. In the same fashion, the plan receives a demographic risk adjustment component depending on the demographic characteristics d of individual i, $\sum_d W_{cc} D_{id}$.[28] W_{cc} and W_d are measured in dollars and the risk score described in previous sections results from the following formula:

$$\text{risk score}_i = \frac{\sum_x W_x D_{ix}}{\sum_x W_x D_{\bar{i}x}} \,, \tag{2.10}$$

where $\bar{i}$ represents the average enrollee and x includes the chronic conditions and demographic risk adjusters. It is clear from equation 2.10 that enrollees sicker than average will have a risk score greater than one, while enrollees who are healthier than average will have a risk score less than one.

I assume that insurance companies are risk neutral and that the market is competitive. Therefore, each plan must make zero profits in equilibrium. In competitive markets with no risk adjustment plan bids will reflect average costs of each plan. In this scenario, differences in plan bids reflect differences in average costs across plans:

[28] See Carey (2017) for a more comprehensive treatment of risk adjustment payments in Medicare Part D.

$$bid_j = E(\text{cost}_j) \tag{2.11}$$

In contrast, with partial risk adjustment plan bids are defined by the following equation:

$$bid_j = E(\text{cost}_j) - E(\text{RA}_j) + E(\text{cost}) \tag{2.12}$$

Now differences in plan bids reflect differences in average residual costs, (Layton, 2017). The residual cost of individual i is the difference between total cost and his risk adjusted payments; in other words, costs that can not be predicted by the model. Note that if risk adjustment was perfect, then insurance companies would be bidding the average cost of the market and there would be no difference in plan bids.

Following the actuarial literature I assume that firms form expectations of future costs (residual costs) with the average cost (residual cost) of enrollees that are currently under the plan.[29] This stylized model of plan pricing attempts to capture the key features of how insurance companies set their bids in this market. In the bidding process, insurance companies have to send a bid that represents the estimated cost for providing the basic benefit. Bids for the upcoming year are submitted in the current year. Given the timing of this process, the bid will likely carry information on the current pool of enrollees in each plan.

Plan premiums in Part D are defined by the following equation:

$$\text{premium}_j = \text{bid}_j - \theta\text{NAB} \tag{2.13}$$

[29]This backward looking depiction of firm behavior is motivated by the empirical presence of adverse selection "death spirals" in insurance markets, (Cutler and Reber, 1998; Handel, 2013).

This means that enrolleess pay the difference between the plan bid and the national average bid (NAB) multiplied by a factor of θ.[30] This last parameter represents the share of premiums that are subsidized by the government. The counterfactual scenario without plan subsidies corresponds to $\theta = 0$:

$$\text{premium}_j = \text{bid}_j \tag{2.14}$$

In summary, policymakers can adjust the level of government involvement in the market through the risk adjustment formula and through the subsidy level. Table 2.11 summarizes alternative market environments with and without each of these features.

Table 2.11: Policy Environments

	Subsidy		No Subsidy	
	Payments	Premiums	Payments	Premiums
Risk Adjustment	(Eq. 2.12	Eq. 2.13)	(Eq. 2.12	Eq. 2.14)
No Risk Adjustment	(Eq. 2.11	Eq. 2.13)	(Eq. 2.11	Eq. 2.14)

Notes: Table 2.11 depicts four types of environments depending on the presence or not of risk adjustment payments and premium subsidies. Plan premiums and plan payments can be represented by any combination of equations 2.11-2.14. The current environment of Medicare Part D is represented by the combination of equation 2.12 and 2.13.

I analyze how these four environments affect the relative costs of couples and singles when the government nudges them to adjust their choices. The exercise is meant to provide insight on how couples' behavior would be likely to affect adverse selection in markets with different regulations, including but not limited to the current environment of Part D.

[30] θ is on average 75% in the first five years of the program.

2.6.2 Nudges

The policy counterfactuals envision a regulator who will nudge consumers to conform with expected utility, his preferred normative theory. The regulator, faces a trade-off. The nudge will create an incentive for consumers to choose higher value plans. At the same time, because risk is not fully priced, consumer sorting may increase adverse selection, (Akerlof, 1970; Rothschild and Stiglitz, 1976). This trade-off is studied in the health insurance context by Handel (2013) and by Polyakova (2016). My analysis extends this literature by investigating how this trade-off is modified by consumer demographics, specifically, how the interactions within couples and the interaction between couples and widows.

Unlike prior studies, I assume that the policy affects household choices by reducing λ_h, the scale of the logistic shock, by some percentage κ. Modelling the nudge as a reduction in the scale parameter has two advantages. First, because the main concern of the policy is if it will exacerbate adverse selection. By reducing the scale parameter, consumers preferences will conform with expected utility. The two main seminal papers on adverse selection (Akerlof, 1970; Rothschild and Stiglitz, 1976) were written under expected utility. It has sharp predictions for this phenomenon. My policy then will be less conservative in potentially exacerbating adverse selection compared with previous research designs (Polyakova, 2016; Ketcham *et al.*, 2019). Second, like previous studies that incorporate inertia, I will not identify the underlying mechanisms behind it. Unlike these studies I also estimate a pooling parameter, that in principle, could be affected by the policy. By modelling the nudge as a reduction in λ_h, I address the problem of taking a stand on if the pooling parameter will also be affected. Under my research design, couples may still decide to default or to pool after the nudge if they maximize expected utility by doing so. I also recognize that the policy may affect different household types differently. For example, prior studies have proposed nudging seniors to use the Medicare Part D Plan Finder tool, (Kling *et al.*, 2012).

Plan Finder is meant to be a friendly platform on Medicare.gov that allows seniors to compare plans in Part D or Medicare Advantage. Married couples and widows may react quite differently to such a policy. Widows tend to belong to older cohorts who are less likely to use a computer and the internet as shown by the following table.

Table 2.12: Question on Internet Use MCBS

Do you personally ever use the Internet to get information of any kind?	yes	no
Spouses	46	54
Widows	21	77

Notes: Table 2.12 is based on a question asked on the Medicare Current Beneficiary Survey between years 2006-2010. The table shows the average share of seniors by answer across this sample period.

Table 2.12 is derived from Medicare Current Beneficiary Survey and suggests that widows will be less likely to use the internet to compare plans. This example is simply meant to motivate the exercise that I am interested in exploring: the welfare consequences of nudging heterogeneous consumers such as couples and widows who are likely to respond differently to the policy.[31] Modelling the nudge as a reduction in the scale parameter and allowing the effect to be heterogeneous across households with different costs makes adverse selection a concern. The results will be indicative of how successful the policy may be in the worst case scenario for adverse selection.

I simulate market adjustment in scenarios where couples and widows are affected differently by the policy with a wedge $w = \kappa_c - \kappa_w$, where κ_c and κ_w are the fractions by

[31] Although in the example I am suggesting that widows could be less affected by the policy, we could imagine policies where the opposite happens. For example, widows will be likely more affected than couples with a policy that automatically enrolls seniors in a default plan. In general, more information will be needed to assess the plans that best meets the needs of a couple than for one single person. I show the results of a policy with this feature in Appendix B.

which λ_c and λ_w are reduced as a result of the policy. Since the level of w would be policy-specific, I bound welfare for different levels of w. Throughout the simulations, I set κ_w to 0.9, a 10% reduction, and report the welfare implications and outcomes for differently sized wedges w.

The error terms that define deviations from expected utility in the stochastic choice model are assumed to be irrelevant for calculating consumer welfare. Of course, the errors may also reflect sources of unobserved heterogeneity that affect the willingness to pay for insurance such as the quality of customer service or pharmacy networks.[32] Thus, my simulation exercises assume that the policy targets the component of the error that is not explained by unobserved heterogeneity.

2.6.3 Policy Counterfactual

I simulate the effects of nudges in four counterfactual environments.

Counterfactual I: This scenario considers policy that reduces λ_s, the scale parameter of singles, by 10% and at the same time reduces the scale parameter of couples by different magnitudes. The policy is modeled as starting in 2007 and continuing through 2010. Importantly, I conduct this experiment in a market without premium subsidies and risk adjustment payments. Therefore, bids and premiums will be determined by equations 2.11 and 2.14.

Counterfactual II: This scenario is the same as counterfactual I but adds risk adjustment payments. This experiment is meant to reveal how risk adjustment payments influence the sorting patterns of couples and singles in response to a nudge. Thus, bids will be determined by equation 2.12 and premiums by equation 2.14.

Counterfactual III: This scenario is the same as counterfactual II but adds subsidies to

[32]In one extreme, one could interpret ϵ as completely driven by unobserved heterogeneity, (Bundorf *et al.*, 2012) and in the other extreme we can interpret ϵ as completely driven by errors, (Abaluck and Gruber, 2011).

mirror the real Part D environment. The subsidies will obviously limit adverse selection and improve consumer welfare. The question is by how much. Here bids and premiums will be determined by equations 2.12 and 2.13.

Taken together, the first three counterfactuals will allow me to determine how each feature of this market affects outcomes when considering policies that nudge consumers toward different choices. This knowledge is important for assessing the implications of similar policies that could be introduced in markets that do not share the same regulations as Part D, such as the Medicare Advantage markets where plan premiums are not as heavily subsidized.

Counterfactual IV: The final counterfactual is a thought experiment in which I compare the evolution of the market with two models. The first model corresponds to my estimation results where spouses choose plans as a group and share risks. The second model treats each spouse separately. When they decide in isolation, each spouse will make a choice with their own risk aversion parameter σ_i and when they decide as a couple they will still do it with σ_i^*. The comparison of both models will shed light on how couples' behavior affects outcomes in individual insurance markets.

Simulation mechanics: Each simulation consists of solving for market outcomes with the estimated model. I simulate the baseline and each counterfactual 500 times and report average outcomes. The policy consists of reducing λ_h from 2007 onward. In each replication a baseline scenario with an initial allocation of consumers is compared to the policy scenario. I am interested in isolating the effect of different households responding differently to the designed policy. Welfare will be measured with the certainty equivalent money metrics. Therefore, the comparison of scenario A with scenario B for household i will be done with the money equivalent $me_i = \text{CE}_i(A) - \text{CE}_i(B)$.[33]

[33]See Pope and Chavas (1985) for a comparison of this metric with other welfare measures under uncertainty.

2.6.4 Results

Tables 2.13, 2.14 and 2.16 summarize results from the first three counterfactuals. In each table, the first two rows show the average change in the welfare of couples and singles in years 2008-2010. The third row shows the average increase in premiums in the last three years of the policy relative to the baseline scenario. The next row shows the variation of plans that suffer adverse selection death spirals. I define a death spiral as a situation where the plan share decreases every year and premiums increase every year. This patterns implies that enrollees who exit the plan are relatively healthier than enrollees who stay. The variation in premiums and death spirals provide two quantitative measures for the degree of adverse selection. Finally, the fifth and sixth row show the share of couples who decide to pool after the policy ("share pooling") and the default rates in the first year of the intervention. This last outcome will be useful to compare my results with Kling *et al.* (2012) who find that providing information to Part D enrollees about plans cost increases switching rates by 11 percentage points.

Table 2.13 summarizes counterfactual I. Each column reports results for a different wedge: $w = 0$ means that the policy reduces λ_c by 10%, so there is no difference with respect to singles. $w = 3$ is the situation where the policy reduces λ_c by 30%, three times higher than singles, and so on. The welfare and premium changes are always calculated relative to the baseline scenario where enrollees are not nudged. Couples' welfare increases after nudging enrollees. The welfare of singles decreases slightly. The nudge tends to induce couples who decide to switch, to move to plans with smaller shares of widows. Since couples' risk premium is small relative to widows, they are not willing to pay plan premiums that do not reflect their own costs. At the same time, this sorting behavior makes widows pay higher premiums in subsequent years relative to the baseline scenario. After the policy, 90% of couples still decide to pool on average. So most couples switch, but still decide to buy the same plan. Under expected utility, choosing the same plan is optimal for most couples. They, however, were not enrolled in the best pooling plan originally. The average increase in premiums and death spirals relative to the baseline scenario reveals that most plans get adversely selected after the policy.

Note that both types of heterogeneity are needed for the results. If couples were similar to widows in terms of cost, the fact that they have different preferences will not be sufficient to generate the results because their movement to other plans would not lead to large changes in premiums. At the same time, if couples differed in terms of costs, but had the same preferences with widows, then they would be likely to buy the same plans, subsidizing widows and preventing plans with higher shares of widows from suffering death spirals. In summary, couples' increase in welfare comes at the cost of a slightly decrease in welfare for widows.

Table 2.13: Counterfactual I. No Risk Adjustment - No Subsidies

	w=0	w=3	w=5	w=7	w=9
welfare change couples (%) 2008-2010	0%	3%	12%	29%	34%
welfare change singles (%) 2008-2010	-3%	-2%	-2%	-2%	-3%
premium variation 2008-2010	14%	12%	11%	13%	14%
(%) Increase Death Spirals	0%	11%	11%	11%	22%
Share Pooling - Policy	84%	90%	92%	94%	95%
Share Defaulting - 2007	91%	89%	84%	71%	57%

Notes: Table 2.13 shows the results from counterfactual I. The environment consists of a market without risk adjustment and without premium subsidies. κ_w is set to 0.9 in 2007 onward, a 10% decrease in λ_w. Each columns shows the results of the policy for different wedges. $w = 0$ corresponds to the scenario where couples are equally affected by the policy. $w = 3$ is the situation where the policy reduces λ_c by 30%, three times higher than singles, and so on. The first rows of the table report the welfare change for couples and singles after the policy. The premiums variation shows the average increase in premiums relative to the baseline scenario in the last three years of the policy. The next row shows the variation of plans that suffer adverse selection death spirals. The fifth and sixth row show the share of couples who decide to pool after the policy ("share pooling") and the default rates in the first year of the intervention.

Table 2.14: Counterfactual II. Risk Adjustment - No Subsidies

	w=0	w=3	w=5	w=7	w=9
welfare change couples (%) 2008-2010	0%	9%	28%	54%	58%
welfare change singles (%) 2008-2010	-5%	-7%	-4%	-4%	-4%
premium variation 2008-2010	14%	13%	13%	13%	13%
(%) Increase Death Spirals	13%	13%	25%	38%	38%
Share Pooling - Policy	90%	91%	92%	94%	95%
Share Defaulting - 2007	90%	88%	80%	68%	57%

Notes: Table 2.14 shows the results from counterfactual II. The environment consists of a market with risk adjustment and without premium subsidies. κ_w is set to 0.9 in 2007 onward, a 10% decrease in λ_w. Each columns shows the results of the policy for different wedges. $w = 0$ corresponds to the scenario where couples are equally affected by the policy. $w = 3$ is the situation where the policy reduces λ_c by 30%, three times higher than singles, and so on. The first rows of the table report the welfare change for couples and singles after the policy. The premiums variation shows the average increase in premiums relative to the baseline scenario in the last three years of the policy. The next row shows the variation of plans that suffer adverse selection death spirals. The fifth and sixth row show the share of couples who decide to pool after the policy ("share pooling") and the default rates in the first year of the intervention.

Table 2.14 summarizes how adding risk adjustment payments to the environment changes outcomes. Recall from section 3 that risk adjustment makes widows a riskier proposition for insurance companies relative to couples. Therefore, when couples move away from plans with high shares of widows, premiums adjust more relative to an environment without risk adjustment. This implies that couples' welfare will increase more relative to the environment without risk adjustment and widows will be worse off, and Table 2.14 confirms these predictions. Imperfect risk adjustment, increases the effects of the nudge.

These results beg the question of whether there is information content in the marital status of enrollees that could usefully be taken into account in the risk-adjustment model. Table 2.15 explores this idea by comparing the difference in costs and residual costs of widows and divorced women relative to couples. The last row shows the increase in the difference between average cost and residual costs relative to married couples. When we compare couples with divorced women, there is little difference in average costs. On average, divorced women are 68 dollars more expensive than couples. The difference in residual costs however increases by almost 150%. Interestingly, the risk adjusted model generates less distortion between the costs of widows and divorced women. The difference in residual costs in roughy 15% less than the difference in average cost.[34] The table suggests that there are factors that make couples' true costs more similar than singles' that the risk adjustment model is not able to capture. This could include behavioral factors as suggested by Einav *et al.* (2016) or factors that can influence health beyond chronic conditions. Thus, the inclusion of marital status in the risk score could reduce the gap between residual costs among households and mitigate adverse selection.

Table 2.15: Costs and Residual Costs Relative to Couples

	Widow	Difference with couple	Divorced Women	Difference with couple
Average Cost	1,116	260	923	68
Average Residual Cost	453	332	289	168
Variation		28%		148%

Notes: Table 2.15 shows the average cost and residual cost for widows and divorced women. It also shows for each of these two variables the difference with married couples. The last row shows the increase in the difference between average cost and residual costs relative to married couples.

Table 2.16 shows the results from implementing the nudge in the actual Part D environment, with risk adjustment payments and premium subsidies. The subsidy is the primary factor that limits adverse selection after we nudge enrollees. Both household types are made better off by the policy. Premiums decrease on average and the increase of plans that suffer from death spirals is lower relative to previous scenarios. This exercise shows that the presence of the subsidy is the main reason why nudging can be successful from the policymaker's perspective. This is important because in markets that are not as heavily subsidized as Part D, the effects of the policy will be likely to work in the opposite direction as shown in the first two counterfactuals. It is also notable that the policy is successful in achieving its objectives in the presence of multiple mechanisms that feed adverse selec-

[34]This calculation follows from the difference in average costs of widows and divorced women (1,116-923=193), with the difference in residual costs of these households (453-289=164).

tion. First, positive correlation between the risk aversion parameter and costs will tend to increase adverse selection. Second, allowing household types to react differently to the policy, can also increase adverse selection because premium differentials will be determined by how couples adjust relative to widows. Third, I exclude other aspects of preferences that could moderate sorting according to risk such as preferences over brand dummies or other non-financial attributes of plans. Remarkably, the policy still succeeds in this environment.

Table 2.16: Counterfactual III. Medicare Part D

	w=0	w=3	w=5	w=7	w=9
welfare change couples (%) 2008-2010	10%	18%	43%	74%	82%
welfare change singles (%) 2008-2010	4%	4%	5%	7%	11%
premium variation 2008-2010	-39%	-26%	-28%	-25%	-20%
(%) Increase Death Spirals	0%	0%	17%	33%	33%
Share Pooling - Policy	90%	91%	92%	94%	95%
Share Defaulting - 2007	90%	87%	80%	68%	57%

Notes: Table 2.16 shows the results from counterfactual III. The environment consists of a market with risk adjustment and with premium subsidies. κ_w is set to 0.9 in 2007 onward, a 10% decrease in λ_w. Each columns shows the results of the policy for different wedges. $w = 0$ corresponds to the scenario where couples are equally affected by the policy. $w = 3$ is the situation where the policy reduces λ_c by 30%, three times higher than singles, and so on. The first rows of the table report the welfare change for couples and singles after the policy. The premiums variation shows the average increase in premiums relative to the baseline scenario in the last three years of the policy. The next row shows the variation of plans that suffer adverse selection death spirals. The fifth and sixth row show the share of couples who decide to pool after the policy ("share pooling") and the default rates in the first year of the intervention.

In summary, combining my results with finding from prior studies suggest that a policy that helps consumers make more informed choices has potential to be successful in the current Medicare Part D environment even after accounting for adverse selection. The fact that the subsidy is able to overcome the positive correlation between prescription drug expenditure and risk aversion together with the distortion in costs generated by the risk adjustment model speaks about the important role of premiums subsidies in the functioning of this market.

Finally, Table 2.17 summarizes results where I compare the effects of the policy with and without accounting for the way that couples interact in their decision-making. I fix the wedge equal to 5 in order to simplify the exposition. Under this wedge, the policy counterfactual is able to reproduce the quantitative findings of Kling *et al.* (2012) for switching rates. The first column shows the results from a model where spouses choose plans independently. The second columns shows the results of the main model, where spouses choose plans jointly.

The two models make very different predictions. A model where spouses are choosing individual insurance plans independently, predicts an increase in premiums and welfare gains only for couples. Moreover, a model where spouses choose plans independently can never reproduce the high pooling rates of spouses, with only 4% of couples choosing the same plan. This figure increases up to 6% after the intervention whereas in a model where spouses choose plans jointly it increases up to 90%. Interestingly, this result impacts the reduction in default rates after the policy. In a model where spouses choose plan separately, the decrease in default rates is of 30 percentage points, three times higher than in a model where spouses choose plans jointly.

Table 2.17: Counterfactual IV. Two Models

	Part D	
	Isolation	Couple
welfare change couples (%) 2008-2010	12%	43%
welfare change singles (%) 2008-2010	0%	5%
premium variation 2008-2010	24%	-28%
(%) Increase in Death Spirals	0%	17%
Share Pooling - Policy	6%	92%
Share Defaulting - 2007	59%	80%

Notes: Table 2.17 shows the results from counterfactual IV. The wedge is fixed equal to 5. Under this wedge, the policy counterfactual is able to reproduce the quantitative findings of Kling *et al.* (2012) for switching rates. The first column shows the results from a model where spouses choose plans independently. The second columns shows the results of the main model, where spouses choose plans jointly.

This result shows that joint decision-making is important for predicting how markets respond to a nudge, and the welfare implications involved differ from a model where joint decision-making is not taken into account. Overall, the four counterfactuals examined in this section reveal how different features of this market interact with policies that nudge consumers to shape market outcomes and their welfare implications. The heterogeneity in costs and preferences of the main two consumer groups exacerbates adverse selection after the policy. Risk adjustment increases the differences in costs of couples and widows. This increases the distributional welfare consequences of the policy and it exacerbates adverse selection. The federal subsidy is the key feature of the market that allows the nudge to achieve the policymaker's objective. It increases the likelihood that couples and widows choose the same plans and therefore mitigates adverse selection. An implication of these findings is that markets that are not as heavily subsidized as Part D, are likely to have greater disparities between winners and losers from policies that help consumers make more informed choices.

2.7 Conclusion

In recent years much attention has been devoted to understanding the equity and efficiency of federally regulated health insurance markets in the US. From understanding the choices of consumers to the incentives of firms, the emergence of these markets sparked many studies that advanced knowledge of the economics of regulated health insurance markets and their policy implications. However, the literature to date has focused on modeling individual choices. The fact that most federally regulated health insurance markets (Part D, Medigap, Advantage) only sell individual plans does not imply that consumers will choose plans as individuals. A large proportion of consumers in these markets are married and married couples have strong incentives to coordinate on their enrollment choices.

This paper is the first economic study to analyze the behavior of married couples in individual health insurance markets. It sheds light on how their behavior and interactions with other types of households affects market functioning. I documented that, strikingly, more than two thirds of couples decide to buy the same plan, that their degree of positive assortative mating in expenditure risk is small compared to other contexts such as education, and that their degree of inertia is similar to singles. However, I also found that couples differ from singles in terms of risk aversion and costs. Spouses are in general less costly to insurance companies compared to other types of households (widowed, divorced, etc) and less risk averse. These differences are crucial to understanding the consequences of policies that nudge consumers toward different choices. The fact that couples behave less risk adversely than single households because they are able to share risks makes it harder to contain adverse selection when consumers are nudged toward enrolling in higher-value plans.

I also characterized how standard regulatory features of insurance markets like risk adjustment payments and premium subsidies modify the ways in which different types of households sort themselves across the market in response to a nudge. In Medicare Part D, the imperfect risk adjustment model makes single households more costly relative to couples, exacerbating adverse selection. The inclusion of marital status as a risk adjustment component would likely moderate some of the distortions. Finally, my results suggest that premium subsidies are essential for nudges to be broadly welfare improving in the Part D context.

Chapter 3

STOCHASTIC DOMINANCE TESTS OF HEALTH INSURANCE ENROLLMENT DECISIONS

Consumer behavior often departs from predictions made by simple models of consumer choice under full information, (Dellavigna, 2009; Bernheim *et al.*, 2019). These departures are often found to be quantitative important when consumers choose among insurance plans, (Kunreuther *et al.*, 2013; Richter *et al.*, 2019), products that differ in energy efficiency, Allcott and Taubinsky (2015), and retirement savings plans, Madrian and Shea (2001). Loss aversion, procrastination, inertia, time inconsistency, peer-effects, and search costs are among the mechanisms that have been suggested to better predict how consumers make choices in these and other markets. This study focuses on health insurance markets, where it is commonly found that substantial fractions of consumers choose plans that are somehow "dominated" by other viable plans, (Abaluck and Gruber, 2011; Handel, 2013; Sinaiko and Hirth, 2011; Bhargava *et al.*, 2017; Liu and Sydnor, 2018). For example, Sinaiko and Hirth (2011), Handel (2013), and Bhargava *et al.* (2017) found that employees often choose employer-sponsored health insurance plans that are financially dominated, meaning that for any possible level of total health expenditures (annually) the consumer would have lower out-of-pocket expenditures under a different plan.

However, there is very little robust evidence on the welfare implications of choosing dominated plans. Prior evidence on welfare is based almost exclusively on a narrow set of parametric specifications for utility that assume risk neutrality or constant absolute risk aversion, (Abaluck and Gruber, 2011; Handel, 2013; Handel and Kolstad, 2015; Handel *et al.*, 2019; Bhargava *et al.*, 2017). This paper proposes a novel measure of whether health insurance plans are dominated that is robust to a large set of normative theories and can be used to calculate bounds on the welfare loss from choosing dominated plans in the spirit of Bernheim and Rangel (2009) and Bernheim *et al.* (2015). Our approach is based on stochastic dominance rankings. First, we show how researchers can measure second order stochastic dominance (SOSD) across a range of contexts where they have access to insurance claims data. A distribution is dominated in SOSD by another distribution if every risk averse consumer prefers the latter over the former. SOSD rankings allows us to assess consumer choices in insurance markets under mild assumptions. Moreover, it enables researchers to evaluate consumer choices in a broad class of insurance markets where plans need not be state–by–state dominated.[1] Second, we augment the binary SOSD measure with a continuous measure of how far a plan is from the efficient frontier. This allows us to construct non-parametric bounds on the welfare consequences of choosing dominated plans.

Prior research on evaluating consumer decision making quality in ways that are robust to parametric specifications of preferences has been primarily limited to experimental settings, e.g. Choi *et al.* (2014), Apesteguia and Ballester (2015), and Heufer (2014), and Kourouxous and Bauer (2019). These studies detect systematic violations of rational choice when agents are repeatedly choosing small stakes "lotteries" or hypothetical gambles. These research designs are well suited to distinguish decision-making ability from unobserved heterogeneity in preferences but it is unclear whether their findings are informative about how consumers behave when facing high stakes financial decisions.

Our research design takes the stochastic dominance ranking measures for evaluating individual rationality from the experimental literature and adapts them to health insurance markets.

We use this method to ask two questions. First, how robust are measures of plan dom-

[1]Prior studies have measured dominance in highly specialized settings where insurance plans were made available to employees by a single employer (Handel (2013), Bhargava *et al.* (2017)).

ination when we only assume consumers are weakly risk averse, (i.e. relaxing parametric assumptions on utility and ex-ante distributions of expenditure). Second, how ambiguous are the welfare consequences of choosing dominated plans? Both questions are important to understand the generality of previous results and to have a better understanding of the possible welfare implications of this behavior. The estimation of welfare bounds will allow us to assess the welfare ambiguity of choosing dominated plans.

We apply our method to study enrollees' choices of prescription drug insurance plans (PDPs) in markets created by Medicare Part D. Since its inception, Part D has been used to study consumer decision-making. For example, Abaluck and Gruber (2011, 2016); Ketcham *et al.* (2016, 2019), used mean-variance frontier analysis to determine the shares of enrollees selecting dominated plans. Under this approach, a distribution is dominated if it has a higher mean and a higher variance than another distribution.[2] As we show later, preferences for two or more moments are not nested under stochastic dominance axioms. Therefore, it is a competing theory. For this reason, we compare (empirically) the efficient sets of plans under both theories and we show the caveats of using preferences for two moments –e.g. mean-variance– when the assumption of normality of distributions doesn't hold.

We start by defining a "standard consumer" as a person whose preferences are consistent with first order stochastic dominance (monotonicity), and risk aversion in the sense that he dislikes mean-preserving spreads, (Rothschild and Stiglitz, 1970; Machina and Pratt, 1997). As we show in the following section, these two assumptions imply that our standard consumer will never choose plans that are stochastically dominated in second order. Figure 3.1 shows examples of theories that are consistent with stochastic dominance.

[2]If non-financial characteristics are taking into account –e.g. brand reputation–Abaluck and Gruber (2011) found that 70% of enrollees select dominated plans while, if non-financial characteristics are added to consumers' preferences, only 20% of them will be choosing dominated plans as shown by Ketcham *et al.* (2016).

Figure 3.1: Stochastic Dominance and Decision Theories

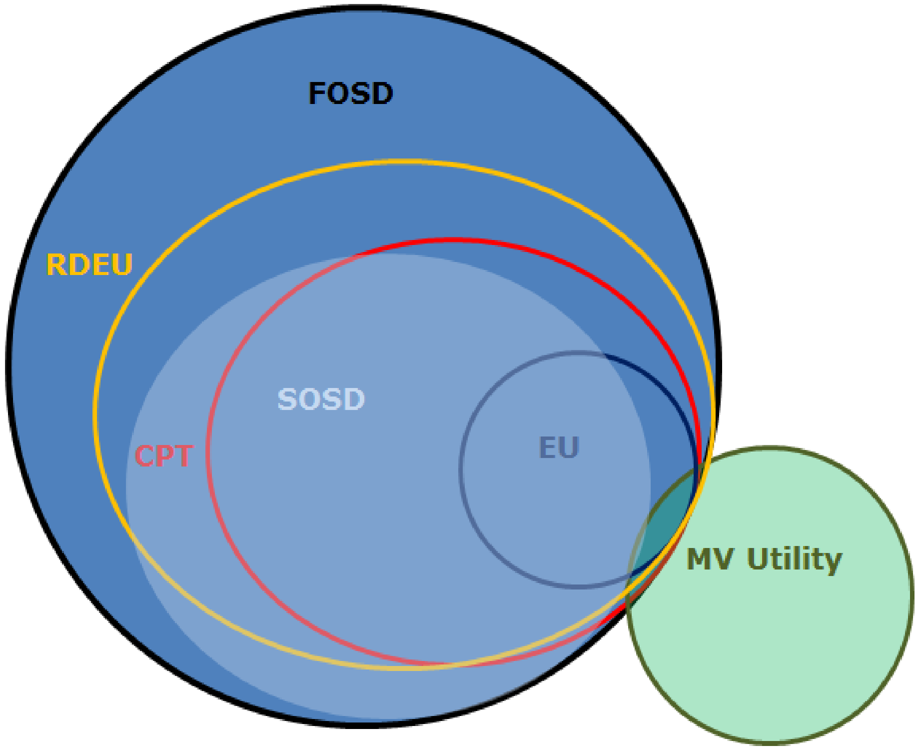


Figure 3.1: Figure 3.1 shows some of the valuation theories that are consistent with first–order stochastic dominance. Rank Dependent Expected Utility (RDEU), Expected Utility (EU), Cumulative Prospect Theory (CPT) and mean variance utility (MV). MV is consistent with FOSD only in the cases where it is consistent with EU. Some well–known examples of the this are quadratic utility and CARA utility coupled with normally distributed expenditures.

Figure 3.1, shows that stochastic dominance rankings are consistent with a wide range of valuation theories, e.g. expected utility, rank-dependent utility and cumulative prospect theory among others. Moreover, no assumptions on distributions of out of pocket expenditures are needed to define the efficient set of plans. This generality is key to distinguishing dominated choices from preference heterogeneity. The figure shows a few common valuation theories for illustrative purposes, but, almost all normative theories of decision-making under uncertainty are consistent with FOSD, Machina (1987). The subset contained under the label SOSD will take from each theory, preferences orderings that are consistent with risk aversion.

The limitation of stochastic dominance measures is that only categorize choices as being dominated or not. It does not quantify the "magnitude" of dominance. The intensity by which a plan is dominated determines the welfare consequences of choosing dominated plans. We develop an empirical measure of intensity by constructing willingness to pay-bounds like Bernheim and Rangel (2009). These bounds measure the lowest and highest welfare loss of standard consumers. In the context of "lotteries", these measures were first introduced by Mjelde and Cochran (1988) to calculate lower and upper bounds on the value of information of climate forecasts. We show that these measures can be interpreted as lower and upper willingness to pay bounds for efficient (non dominated) plans, and importantly that they contain information beyond expected utility theory. Finally, like Choi *et al.* (2014) we investigate which measures of consumer demographics are associated with larger welfare losses from choosing dominated plans.[3]

We find that mean-variance (MV) and second order stochastic dominance measures both imply that approximately, 70% of consumers select dominated plans based on financial characteristics (20% based on financial and non-financial characteristics). However, MV and SOSD measures differ in which sets of plans they label as dominated when we control for non-financial attributes. A second main finding is that the welfare ambiguity of choosing dominated plans is large. The average welfare loss captured by our upper bound is eight times higher than the lower bound, \$1,307 versus \$170 respectively. This implies that the average welfare loss is never smaller than 34% of the annual premium. Finally, while some consumers suffer welfare losses higher than \$2,000 per year, the probability of being in this group is not systematically correlated with observable characteristics such as income, education, and gender.

[3]Although designing policies to help consumers select better plans in insurance markets has been put into question since it could exacerbate adverse selection, Handel (2013). Targeting specific population groups may not impact insurance premiums.

Overall, our study advances knowledge in three areas. Relative to the literature that finds consumers selecting dominated plans in health insurance markets (Abaluck and Gruber, 2011; Sinaiko and Hirth, 2011; Handel, 2013; Ketcham *et al.*, 2016; Bhargava *et al.*, 2017), we provide welfare bounds of such behavior using a non-parametric method that is by definition, more general than state-wise domination. Relative to the experimental literature that detects violations of rational choice (Choi *et al.*, 2014; Apesteguia and Ballester, 2015; Heufer, 2014; Kourouxous and Bauer, 2019; Birnbaum and Navarrete, 1998; Birnbaum and Bahra, 2007), we are the first study that brings to the field a non-parametric method to assess consumers' decision-making under risk. Third, we contribute to the literature that assesses decision-making in Medicare Part D (Abaluck and Gruber, 2011; Ketcham *et al.*, 2016; Abaluck and Gruber, 2016; Ketcham *et al.*, 2019) by comparing stochastic dominance rankings with mean-variance utility analysis. The comparisons of these two methods in empirical contexts have been done in other markets (Porter *et al.*, 1972; Levy and Hanoch, 1970; Levy and Sarnat, 1970).

The paper proceeds as follows, in section 3.1 we develop the underlying theory behind our non-parametric method, section 3.2 explains how we implement this method in insurance settings, section 3.3 describes the institutional detail of Medicare Part D and the data, in section 3.4 we show our results and section 3.5 concludes.

3.1 Stochastic Dominance of Insurance Plans

3.1.1 Consumers' Preferences

In this section, we present a conceptual framework to guide our empirical work. Let *J* be the set of plans available for a consumer i, and j a specific plan that belongs to J. The value of plan *j* for individual *i* will be given by the welfare function:

$$U_i(q_j, F_{ij}) \tag{3.1}$$

where q_j is a vector capturing the plan's non–financial characteristics, such as customer service, brand reputation or pharmacy networks. The second component of the welfare function F_{ij} represents plan j's distribution of potential out of pocket (*oop*) expenditure plus premiums for person i. The i subscript reflects the fact that the distribution of *oop* expenditures will differ across individuals according to their health.

For a fixed q_j, we define the welfare function:

$$U_i(q_j, F_{ij}) = V_i(F_{ij}) \tag{3.2}$$

So that $V_i(.)$ captures how individual i feels about the financial characteristics of plan j, F_{ij}, conditional on non–financial characteristics. Given that we will define domination according to financial characteristics of plans (premiums and coverage design including cost–sharing), it is important to recognize that other characteristics may be also relevant for consumers. All these other characteristics are embedded in q_j.

To assess consumer decision-making we need a normative theory of rational choice. Ideally, this theory would represent a large set of preference orderings that are used in theoretical and empirical studies. For this, we will define a "standard consumer" whose preferences will conform with two general assumptions: first order stochastic dominance and risk aversion[4]:

Assumption 1: V_i is consistent with first order stochastic dominance (FOSD) preferences. Distribution G dominates distribution F in FOSD ($G \geq_{fsd} F$) if $G(x) \leq F(x)$

[4]Risk aversion is a natural assumption in our context. It is the very nature of these preferences that give rise to insurance markets. Risk averse individuals are willing to pay a sure amount of money (premium) to avoid certain types of risk.

for all x in the support of F and G with strict inequality for at least one x. V_i is consistent with FOSD if $G \geq_{fsd} F$ implies $V(G) > V(F)$. This is a type of monotonicty of preferences. In particular, for two monetary outcomes where $x_1 > x_2$, FOSD implies that $V(x_1) \geq V(x_2)$.

Assumption 2: Risk aversion: Let $F(x)$ be a mean-preserving spread of $G(x)$, then $V(G) \geq V(F)$. We adopt the definition of mean-preserving spreads in Machina-Pratt (1997). Under this definition F is a mean-preserving spread of G if there exists x' and x'' with $x'' > x'$ such that: F assigns at least as much probability as G to every sub-interval of $(-\infty, x')$, F assigns no more probability than G to every sub-interval of (x', x'') and F assigns at least as much probability as G to every sub-interval (x'', ∞). Under expected utility this condition is equivalent to the requirement that the utility function be concave. Alternatively, under rank-dependent utility, it is equivalent to requiring concavity of the utility function and convexity of the weighting function.

Having established our benchmark theory, we will define a dominated plan (distribution) as a plan that will never be chosen by our "standard consumer". We will give now the definition of second order stochastic dominance, that we use to find the set of dominated plans.

Definition 1: Distribution F dominates distribution G in SOSD if $\int f dF \geq \int f dG$ for all increasing, concave f.

The link between our "standard consumer" and SOSD is given by Machina and Pratt (1997). They showed that if F dominates G in SOSD, G can be obtained from F as a sequence of first order deteriorations[5] and/or mean-preserving spreads. Therefore, our

[5]The type of first order deterioration that they use is a leftward shift in probability. A distribution G is said to differ from F by a leftward shift of probability mass if there exists an outcome level x' such that: F assigns at least as much probability as G to every subinterval of (x', ∞), F assigns no more probability than G to every subinterval of $(-\infty, x')$.

standard consumer will never choose a plan that is dominated in SOSD.

In empirical studies, it is common to link stochastic dominance *only* with expected utility (EU). This would be unsatisfactory since we will be considering EU as the only rational benchmark. Recently, some empirical studies have used decision theories under uncertainty with non-linear probability weighting to rationalize insurance choices, (Barseghyan *et al.*, 2013b,a). Zilcha and Chew (1990) show that if we take into account all non-linear preference functionals that are consistent with our two assumptions the efficient sets remain unchanged relative to the efficient sets of risk–averse EU consumers. Therefore, choosing a dominated plan in SOSD could not be rationalized by a larger set of theories, e.g. cumulative prospect theory and rank dependent utility, provided that the consumer is averse to mean-preserving spreads. This result provides empirical tractability; if we are able to find the efficient sets of risk averse expected utility consumers, we will find the efficient sets of a much broader set of decision theories.

3.1.2 Welfare Bounds Under Second–Order Stochastic Dominance

Definition 1 shows that the main drawback of stochastic dominance is that it is a binary measure. Although it is useful to define consistency in a binary way, it provides no insight about the magnitude of any inconsistencies. For this purpose, we propose two money metrics that will bound the welfare implication of choosing dominated plans. Let ϑ be the set of all utility functions that satisfy assumptions 1 and 2. Assume that the random variable X with cumulative distribution function F dominates random variable Y with cumulative distribution G in SOSD. The lower bound δ_l and upper bound δ_u are given by:

$$\delta_l = \min_{V \in \vartheta} \min_{\delta \geq 0} \{\delta : V(G(Y - \delta)) > V(F(X))\} \tag{3.3}$$

$$\delta_u = \max_{V \in \vartheta} \min_{\delta \geq 0} \{\delta : V(G(Y - \delta)) > V(F(X))\} \tag{3.4}$$

δ_l can be interpreted as the minimum reduction in premiums needed so that the plan is not dominated. Intuitively, imagine a room full of people with preferences consistent with assumptions 1 and 2. Two plans are being offered, plan A and plan B with $A \geq_{ssd} B$. Every person in the room, then, prefers A over B. An auctioneer, in charge of selling the plans, starts decreasing the premium of the dominated plan until the first person announces that he now feels indifferent between the two plans. The lower bound δ_l captures this premium reduction. It is a money metric linked to the welfare of this first consumer. It is a lower bound since the rest of the room will require a higher reduction in premiums to be indifferent between the two plans. With the same logic, δ_u measures what the "last" person in the room will require to be indifferent between both plans. These measures were first introduced by Mjelde and Cochran (1988) to calculate bounds on the value of information of climate forecasts under expected utility. Bernheim and Rangel (2009) define similar measures for any type of preferences that can be applied in more general contexts.

It may be helpful to compare both measures with standard welfare measures that assume a parametric form of utility. One of the most common welfare measures used in applied and theoretical work is the certainty equivalent of constant absolute risk averse (CARA) consumers. Because of its tractability and closed form it has been used extensively in empirical studies, (Handel, 2013; Handel and Kolstad, 2015; Bhargava *et al.*, 2017; Handel *et al.*, 2019). In theoretical work, the use of CARA preferences to derive risk measures[6] has been proposed by Aumann and Serrano (2008). The certainty equivalents of exponential utilities are also used in the actuarial literature to define premiums. There were introduced by Gerber (1974) and they have many desirable properties. Some of these properties are independence, translation invariance, and continuity. See Kaas *et al.* (2008) for a

[6] Mathematically, a risk measure for a random variable $X \in \Omega$ is a function $\rho : \Omega \rightarrow \mathbb{R}$.

more comprehensive treatment of certainty equivalents and risk measures. The formula for $CE(.)$ for the case of CARA utility is given by:

$$CE_{\alpha}(X) = \frac{1}{\alpha}\log(M_X(\alpha)) \text{ with } M_X(\alpha) = E(e^{\alpha X})$$

More specifically, we will define:

$$\delta_{ra}^{\alpha} = CE_{\alpha}(G) - CE_{\alpha}(F) \tag{3.5}$$

where $CE_{\alpha}(.)$ is the certainty equivalent of the plan for a consumer with risk aversion parameter equal to α.[7] This difference in certainty equivalents is equivalent to the willingness to pay for the better plan, Eeckhoudt *et al.* (1997). In our welfare calculation we will set $\alpha = 0.0001$ which is the average level of absolute risk aversion estimated by Cohen and Einav (2007) using auto insurance data.

We will also calculate δ_{ra} for the special case when $\alpha \to 0$, δ_{rn}, the sure amount of money that a risk neutral consumer is willing to pay for the better plan. It can be shown that:

$$\delta_{rn} = E(G) - E(F) \tag{3.6}$$

where $E(.)$ is the expectation operator. Therefore, δ_{rn} captures the mean difference between the two plans; it is the average "overspending" under the dominated plan relative to the plan that dominates it. Using the definition of risk premium, $RP(X) = E(X) - CE(X)$. we can then rewrite equation 3.5 in the following way:

[7]More formally, for any random variable X, $CE(X)$ is the solution to the following equation, $E[u(w - X)] = u(w - CE(X))$.

$$\delta_{ra}^{\alpha} = \delta_{rn} + (rp_{\alpha}(G) - rp_{\alpha}(F)) \tag{3.7}$$

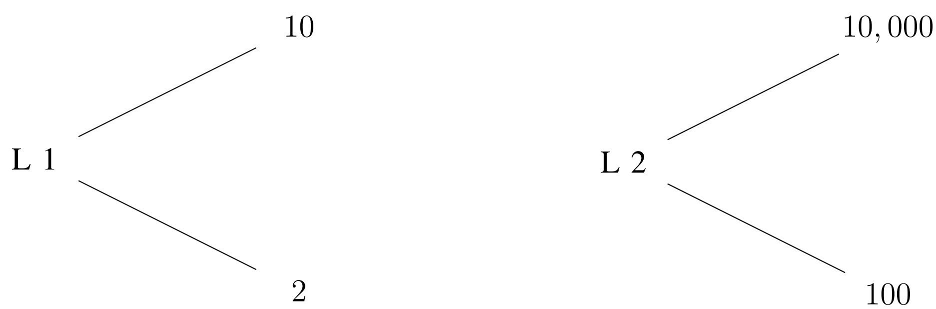
Depending on the difference between the risk premia, the willingness to pay of the exponential risk–averse consumer can be lower, equal or higher than the willingness to pay of the risk neutral consumer depending on the distribution of F relative to G. Finally, given the definitions of δ_l and δ_u it follows that $\delta_l \leq \delta_{ra}^{\alpha} \leq \delta_u$.

3.1.3 Mean–Variance Utility in Medicare Part D

There is a tradition in the Medicare Part D literature to assume two moment utility functions, e.g. mean-variance, as a normative benchmark to assess how consumers make decisions (Abaluck and Gruber, 2011, 2016; Ketcham *et al.*, 2016, 2019; Ho *et al.*, 2017).[8] This specific parametric form of utility is not nested under our non-parametric method. It is in general, not consistent with the notions of state wise and stochastic dominance.

Since Borch (1969), the choice of two-moment utility functions as a normative theory of rational choice has been questioned. Specifically, Borch showed that without additional restrictions on the lottery space or preferences, the indifference curves of mean–variance investors violate the consistency conditions of von Neumann and Morgenstern, and of monotonicity and risk aversion. Another pathological behavior of mean–variance preferences is featured in Ormiston and Schlee (2001). They show that a mean–variance investor could decide to invest more in a risky asset if the government decides to confiscate all the capital gains earned above some pre-specified rate of return. The following example shows one of the main drawbacks of this type of preferences.

[8] Abaluck and Gruber assume that preferences should be consistent with CARA utility and assume normally distributed out of pocket expenditures which is equivalent to mean-variance utility. While in Ketcham *et al.* (2016) the authors use mean–variance to characterize the financial aspects about plans and add a quality measure that summarizes other non–financial characteristics.



Under mean–variance the two lotteries cannot be ranked, both will be at the frontier, even if the second lottery pays more in every state of the world. They cannot be ranked because the second lottery has a higher variance. The example shows that the main problem of mean-variance utility is that the information contained in these two moments is not enough to characterize the problem that the decision-maker is facing. In the above example, there is no trade-off between the two distributions. The mean-variance utility is not consistent with the normatively compelling property of FOSD; i.e. the fact that a distribution dominates another distribution if the probability of getting an outcome higher than x is always higher than the other distribution, Levy (2016). In the health insurance context, a plan is dominated by FOSD if the probability of experiencing expenditures (premium + out-of-pocket) higher than x, is always higher relative to another plan. A rational consumer will avoid such plans.[9]

This criticism motivated researchers to define conditions under which mean–variance utility rankings are consistent with EU rankings, e.g. Meyer *et al.* (1987). Restricting the set of lotteries to the Normal Gaussian family is one example and assuming quadratic utility is another.[10] Beyond these specialized settings, the question of how well mean-variance

[9]State-wise domination measures avoid the problems of mean-variance utility rankings but their applicability is limited to specialized settings where state-wise dominated plans exist. Liu and Sydnor (2018) suggest that insurance plans that are state-wise dominated often occur in employer-sponsored settings. However, this does not appear to be true for federally subsidized health insurance plans.

[10]We tested the normality assumption with a 10% random sample of the distributions implied by plans and

efficient sets approximate stochastic dominance efficient sets is an empirical matter. Several studies compared mean-variance and stochastic dominance efficient sets in the context of financial portfolio returns, (Levy and Hanoch, 1970; Levy and Sarnat, 1970; Porter *et al.*, 1972). Porter *et al.* (1972), for example, found that the two efficient sets are close to each other when considering the monthly returns of 140 stocks during 1960-1963. Similar to these studies, we test how close are both sets when studying *oop* expenditures' distributions in insurance markets.

3.2 Implementing Stochastic Dominance with Insurance Data

While the theoretical assumptions used to define SOSD are well–established, implementing SOSD as a measure of the quality of consumers' insurance plan choices requires additional assumptions. Specifically, researchers must define the lotteries for each plan available to the consumer, F_{ij}. Measuring F_{ij} requires researchers to make decisions about the distribution of risks each person is insuring against, how each plan covers each risk, and whether consumers' responses to each risk depends on the plans' coverage.

This is innately difficult as only the realized outcome under the chosen plan is observed in data. We address all three requirements by combining claims data with the information that was available to consumers about every plan coverage rules for every risk and then assign people to different *ex ante* types. Specifically, we combine administrative information on each plan's coverage rules for every drug with each person's claims for every drug to determine what the person would have spent on that same bundle of drugs under each available plan.

This yields a single observation for each person–plan pair. To measure F_{ij}, we rely on the distribution of these observations across all of the people of a given type. To construct each type of risk we use the RxHcc Risk Adjustment Model developed by the Center of

persons (more than one thousand). We were able to reject normality in all tests at the one percent level.

Medicare and Medicaid Services (CMS). This model was created to adjust CMS's subsidies to insurance companies offering Part D plans. The scores are non-negative numbers normalized to one for the average risk score in the Medicare population. Individuals with higher scores have higher expenditure risk.[11]. The model uses data on chronic condition diagnoses and demographics, all of which we observe, to predict the score. We then define a risk type t as a set of individuals with the same risk score in year T-1 who live in the same CMS region. This approach to defining risk types is similar to the one employed by Handel (2013). The main advantage of defining risk types with risk scores is that it reduces the dimensionality problem of defining risk types with a larger set of state variables: gender, age, and chronic conditions. Therefore, the ex-ante distributions of *oop* expenditures of each plan and type at the beginning of year T (when plans are chosen) are constructed from the realized *oop* costs in year T of all beneficiaries that belonged to the same ex-ante risk type t.

Because the identification of dominated choices and welfare measures may be sensitive to researchers' definition of F_{ij}, we assess how our conclusions differ under alternative definitions of the distribution of risks faced by each person. In one extreme, we assume that they face the full set of risks observed in the data for a given year (i.e. there is only a single type), this is done for example in Sydnor and Liu (2019). At the other extreme, we assume that each person has private knowledge and represents their own type so that their distribution of risks is constructed from their experiences during the five years we observe them in the data. These two extremes are not meant to be realistic; but rather to set logical bounds on the scope to which this analytical decision can influence the results. We formalize our previous discussion with the following assumption.

Assumption 3: Consumers can not choose F_{tj}. It is fixed for every plan j and type t. In other words, there is no moral hazard.

[11]See Robst *et al.* (2007) for more details about the model.

This is essentially the same as the approaches used in Abaluck and Gruber (2011), Abaluck and Gruber (2016), Ketcham *et al.* (2016), and Ketcham *et al.* (2019). The no moral hazard assumption in these studies is justified by the small drug-specific price elasticities estimated in the literature and the high persistence of drug use, both of which are indicators of moderate moral hazard, (Abaluck *et al.*, 2018).

With the *oop* expenditure of each beneficiary in hand, the construction of empirical CDF for plan j and type t is standard:

$$\hat{F_{tj}}(x) = \frac{1}{n_t}\sum_{t=1}^{n_t} 1(X_{tj} \leq x) \tag{3.8}$$

Where n is the number of people of the given type t, x is a non–negative number that belongs to the support of the distribution of *oop* expenditure and $1(X_i \leq x)$ is an indicator function for whether realization *t* in plan j is less than x.[12] We will assume that the distributions of *oop* expenditure implied by each plan and type belong to the bounded and common support $[a, b]$.[13] If a plan has a smaller (realized) support we will define the density function of that plan to be zero outside this range. The following claim simplifies the empirical analysis of SOSD.

Claim. Distribution F dominates distribution G by SOSD if and only if $\int_{-\infty}^{x} F(s)ds \leq \int_{-\infty}^{x} G(s)ds$ for all x and $F \neq G$.

To assess stochastic dominance among empirical distributions we need an empirical

[12]Like Abaluck and Gruber (2011) I require the minimum cell size to be 200 individuals. This guarantees that the empirical CDFs will be at an epsilon distance of 0.10 from the true CDF within a confidence level of 95%,Florens *et al.* (2007). The caveat of this requirement is that 70% of the sample is dropped. Most enrollees belong to risk types with less than 200 individuals. This method of constructing cells may present a trade-off between credibly approximating the true CDF and external validity of my in-sample results.

[13]Here a will be the smallest realization of *oop* expenditure among all plans and types while b will be the maximum of this same variable. As is standard when dealing with empirical CDFs, we discretize the support in r pieces.

analog of this claim. With this in mind, for plan j and type t, we define:

$$I^1(x, F_{jt}) = F_{jt}(x) \text{ and for } s, \geq 2 \;\; I^s(x, F_{jt}) = \int_a^x I^{s-1}(u, F_{jt})$$

Davidson *et al.* (2000) show that for all s, $I^s(x, F_{jt})$ can be written as:

$$I^s(x, F_{jt}) = \frac{1}{(s-1)!} \int_a^x (x-u)^{s-1} dF_{jt}(u) \tag{3.9}$$

The distribution of plan j stochastically dominates the distribution of plan j' for type t in order s if and only if $I^s(x, F_{jt}) \leq I^s(x, F_{j't})$ for all x. If we insert the empirical CDF into expression 3.9 we obtain the empirical analog of the claim:

$$I_2(x; \hat{F_{jt}}) = \frac{1}{n_t} \sum_{i=1}^{n_t} 1(X_{ij} \leq x)(x - X_{ij}) \tag{3.10}$$

Equation 3.10 is just the plug-in estimator of equation 3.9 when $s = 2$. In our preferred specification, we will assume that the theoretical distribution can be set to equal its empirical analog. This allows us to use expression 3.10 to determine stochastic dominance without the need of a statistical test. We feel comfortable making this assumption because our goal is to judge how far the dominated distribution is from the one that dominates in economic terms.

3.3 Evaluating People's Medicare Prescription Drug Insurance Plan Choices

3.3.1 Medicare Part D

Medicare Part D was first launched in 2006 and it was the largest expansion of Medicare since its inception. Like the Medicare Advantage markets that came before and the Affordable Care Act exchanges that came after, Part D relied on private companies competing for enrollees within a government–created marketplace. While some Medicare Advantage plans and employer-sponsored retiree plans provide prescription drug coverage, traditional Medicare did not, leaving around one–third of people on Medicare without drug insurance. Part D was established by law in 2003 and launched in 2006 to address this missing market. In 2019, 45 million of the more than 60 million people covered by Medicare are enrolled in Part D plans. Of this total, more than half (56%) are enrolled in stand-alone prescription drug plans (PDP) and more than 4 in 10 (44%) are enrolled in Medicare Advantage drug plans (MA-PD), (KFF, 2019). Medicare Part D is a voluntary benefit; beneficiaries can choose to enroll in either a stand-alone PDP to supplement traditional Medicare or a MA-PD, mainly HMOs and PPOs, that cover all Medicare benefits including drugs.

Beneficiaries with low incomes are eligible for assistance with Part D plan premiums and cost sharing. Through the Part D Low-Income Subsidy (LIS) program, additional premium and cost-sharing assistance are available for Part D enrollees with low income levels. To encourage take-up in this group, CMS automatically enrolls beneficiaries who are eligible for LIS in PDP plans although they can decide to switch if they are not satisfied with the plan.

CMS divides the country into 34 regions (markets) to sell PDP plans and 26 regions for Medicare Advantage plans. In each region, different sets of plans are offered. Beneficiaries can enroll in two main periods: the Initial Enrollment Period which coincides with the enrollment period for newly Medicare Part B beneficiaries[14] and the Open Enrollment Period which is meant for people already enrolled in traditional Medicare (Part A and Part B) or enrollees who want to switch plans. This is held from October 15th to December 7th of each year. The default for new Medicare beneficiaries is to be uninsured. They must enter the market and actively choose a plan to become insured. Their choice becomes their automatic default plan for the following year. They will be re-enrolled in the same plan

[14]This is a seven-month window for people who turn 65. The window begins 3 months before the month you turn 65, includes the month you turn 65 and ends with the 3 following months.

unless they actively switch plans, opt out of the market during the annual open enrollment period, or if their plan exits the market the following year.

Insurance companies that want to participate have to sell plans that meet certain requirements in terms of coverage and benefits. The Standard Plan parameters are set by CMS each year and insurance companies can sell any plan that is actuarially equivalent to the Standard Plan and they can also sell plans with enhanced benefits. A Standard Plan is characterized by four attributes: monthly premiums, annual deductible, initial coverage limit and an out of pocket catastrophic threshold. The range between the deductible and the initial coverage limit is known as the Initial Coverage Period and the enrollee is responsible to pay 25% coinsurance for the drugs covered by the plan. The range between the initial coverage limit and the catastrophic threshold is known as the Coverage Gap or "doughnut hole" in which enrollees have none or very limited coverage until they reach the catastrophic threshold. Once they reach the catastrophic threshold, they face a 5% coinsurance rate. An insurance company can sell an actuarially equivalent plan to the Standard Plan. These plans are characterized by a premium and the same deductible of standard plans. However, insurance companies can change the cost-sharing structure. For example, they can use co-payments instead of co–insurance rates in the Initial Coverage Period. Finally, insurer companies can sell "Enhanced Plans" which the actuarial value must exceed the actuarial value of standard plans. An example of this is zero deductible plans. In 2016, 58% of beneficiaries in PDP plans were enrolled in a basic plan (standard or actuarially equivalent plans) and 42% were enrolled in enhanced plans, Hoadley (2016). Finally, among non–financial characteristics, PDPs can differ in terms of customer service, preferred pharmacy networks, mail order pharmacy access and drug management utilization rules.[15]

[15]For example, the beneficiary may need a plan's approval before it will cover a particular drug.

3.3.2 *Data and Sample*

We link three CMS data sets. The first is a 10% random sample of administrative data for Medicare beneficiaries age 65 and over between 1999 and 2013. For those on traditional Medicare (rather than Medicare Advantage), these data include information about whether and when each person received a diagnosis for a large set of chronic conditions. We combine this information to construct the risk scores, known as the RxHCC index, that CMS uses to adjust payments to PDPs based on the individual's expected drug spending.

The second set of data includes prescription drug claims for a random 20% sample of those age 65+ who enrolled in a PDP without a low–income subsidy at any point between 2006 and 2010. By combining these claims with CMS information on each plan offered in each region, including the specific coverage design of the plan such as cost–sharing and tiering rules, we construct an estimate of what each person would have spent under each plan available to them. This is accomplished with the "cost calculator" developed by Ketcham et al (2015). In this cost calculator, we assume that patients will substitute between generic versions of the same active ingredient but will not substitute between active ingredients nor change their quantities in response to cost–sharing.

Third, we link to panel survey data from the Medicare Current Beneficiary Survey (MCBS) between 2005 and 2011. The MCBS provides detailed information on demographics, knowledge, health insurance-related question, health status and health perception among other modules.

3.4 Results

3.4.1 *Choice Quality Under Various Measures*

Table 3.2 reports the average sizes of efficient sets and the shares of consumers choosing efficiently under three criteria: FOSD, SOSD, and mean-variance. For each criterion,

Table 3.1: Summary Statistics

Variables	statistic
# admin sample	905,871
# mcbs sample	2,465
age (mean)	76
male (%)	41
married (%)	51
white (%)	95
# plans (mean)	50
# brands (mean)	23
chosen premium (mean)	396
experienced *oop* (mean)	966
average premiums of available plans	496
range of premiums of available plans	22 - 1,628
average of OOP of available plans	1,028
range of OOP of available plans	0 - 34,694

Notes: Table 3.1 shows summary statistics for the final sample of enrollees. The observations correspond to person-years. # admin sample refers to the final sample of enrollees who belongs to the administrative data and Medicare Part D. # mcbs sample refers to the final sample who belongs to the MCBS data and Medicare Part D.

the average size of efficient sets is taken over risk types. We also calculate how the efficient sets and shares of consumers choosing efficiently change after controlling for non-financial attributes. Specifically, we control for non-financial plan quality by allowing consumers to have unrestricted preferences over insurance companies and plan-specific star ratings reported by CMS.[16] These two attributes have previously been found to help predict consumer choices, e.g. (Abaluck and Gruber, 2016; Ketcham *et al.*, 2016). We incorporate them into the second and third rows of Table 3.1 by limiting the set of potentially dominating plans to those that have equal or higher star ratings (middle row) or are sold by the same insurer (bottom row).

Comparing the sizes of the efficient sets within each row help us understand how "effective" are the different method as prescriptive theories. Effectiveness is measured as the

[16]The Centers for Medicare Medicaid Services (CMS) created a Five Star Quality Rating System that rates Part D plans. Ratings are between 1 and 5, 5 being the highest, for health plan quality based on measurements of customer satisfaction and quality of care the plan delivers.

Table 3.2: Efficient Sets and Efficient Choices

	FOSD		SOSD		Mean-variance	
	Efficient set	Efficient choice	Efficient set	Efficient choice	Efficient set	Efficient choice
All Plans	0.57	0.67	0.21	0.33	0.18	0.24
Quality	0.65	0.77	0.33	0.47	0.30	0.45
Brand	0.91	0.96	0.71	0.80	0.74	0.80

Notes: Table 3.2 shows the average size of efficient sets and the share of consumers choosing efficiently under different criteria, e.g. FOSD, SOSD, and mean-variance. For the case of efficient sets, the average is taken across risk types. The table shows these measures for different sets of plans: all plans, same or higher quality, same brand.

size of the efficient set relative to the feasible set. This concept of effectiveness is used by Levy (2016) to test how many prospects can be ranked as additional assumptions are made about investors' preferences. For example, the efficient set under FOSD criteria shows the set of plans that should be chosen by consumers when only monotonicity is assumed. If this set is identical to the feasible set (i.e. a share of 1), it means that more assumptions are needed to rank plans. By adding the assumption that consumers are risk averse, the efficient set will shrink (i.e. moving from FOSD columns to SOSD columns). Therefore, the size of the efficient set under SOSD relative to FOSD can be interpreted as what a risk-manager (researcher) gains in terms of narrowing his recommendation at the cost of adding one more assumption and possibly misspecifying preferences. As we move down in the table the, efficient sets expand by construction because as we control for quality attributes fewer plans will be compared.

Many interesting observations can be made from table 3.2. First, the average sizes of efficient sets are 57%, 21% and 18% under FOSD, SOSD, and mean-variance when comparing all plans. Comparing these statistics to previous studies, Levy and Sarnat (1970), found that the sizes of efficient sets for American mutual funds (based on their annual rates of return) were roughly 70%, 20%, and 17%, for FOSD, SOSD, and mean-variance respectively. Second, the sizes of efficient sets under SOSD and mean-variance are very similar even after controlling for non-financial attributes. Both criteria are very effective in the sense that many plans can be ranked with few assumptions. The share of consumers selecting efficient plans is also similar under both criteria. Both measures imply that roughly 70% of consumers select dominated plans in Medicare Part D. This number gets reduced to 20% when controlling for brands. Finally, while the SOSD and mean-variance measures have nearly identical implications for the fractions of consumers choosing dominated plans, it is not clear from Table 3.1 whether the two measures coincide in their assessments of which plans are dominated.

Figure 3.2 indicates the extent to which the SOSD and mean-variance measures agree or disagree labeling a distribution (plan + type) as dominated.

Figure 3.2: Dominated Distributions: SOSD vs MV

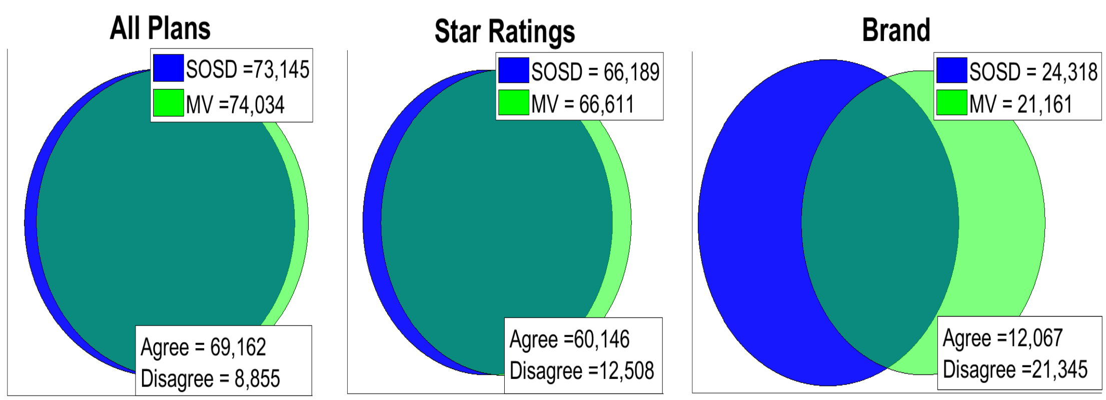


Figure 3.2: Figure 3.2 compares the inefficient sets under mean–variance and SOSD for the five years in the sample. It shows the number of distributions that are in each set, the number of distributions that are in the intersection of both sets, and the number of distributions where there is disagreement.

Figure 3.2 shows the inefficient sets under mean–variance and SOSD, pooling data over all five years in the sample, without controlling for non–financial characteristics (all plans), when controlling for star rating, and when controlling for the insurer selling the plan. It shows the number of distributions that are in each set, the number of distributions that are in the intersection of both sets, and the number of distributions where there is disagreement. To understand if mean-variance efficient sets are approximating SOSD efficient sets it is crucial to compare both measures for different sets of plans. We know from Table 3.2 that the sizes of the inefficient sets mean-variance and SOSD are roughly 80 and 70 percent when comparing all plans and plans with equal or higher star ratings. Therefore, the intersections of both sets in the left and middle panels of Figure 3.2 will be large. If mean-variance rankings approximate SOSD rankings, the relative sizes of the sets of plans where there is disagreement should be fairly constant. When comparing all plans the number of distributions where both measures disagree relative to the number of plans that are dominated under mean-variance is approximately 12% (9/74). This share increases up to 19% (13/67) when controlling for star ratings, and up to 100% (21/21) when we compare plans within the same brand. Thus, as we control for more attributes the sizes of the inefficient sets shrink and judgments about which plans are dominated under each theory increasingly diverge.

The choice of one method over another as a normative benchmark will depend on the axioms of rational choice that we are willing to maintain. Among all axioms that characterize rational choice, consistency with FOSD is typically considered a requirement, Machina (1987). With this in mind, Figure 3.3 assesses the empirical importance of the theoretical criticism of mean-variance rankings as being inconsistent with FOSD. As in the previous figure, we compare both criteria for different sets of plans.

Figure 3.3: Dominated Distributions: FOSD vs MV

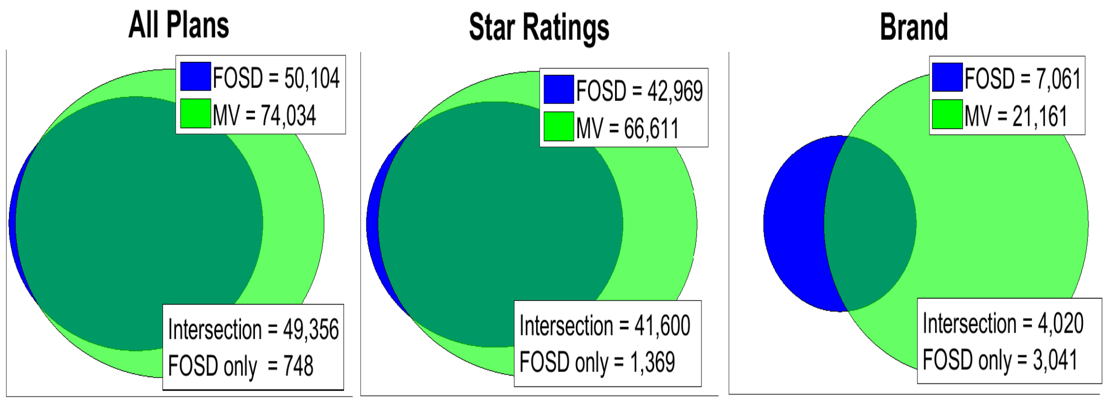


Figure 3.3: Figure 3.3 compares the inefficient sets under mean–variance and FOSD for the five years in the sample. It shows the number of distributions that are in each set, the number of distributions that are in the intersection of both sets, and the number of distributions where there is disagreement.

Mean-variance would be more compelling as a tool to assess decision-making if the set of plans that are dominated by FOSD are in general a subset of the plans that are dominated by mean-variance. To test whether the mean-variance criterion displays this form of empirical consistency, we calculate the ratio of plans that are dominated by FOSD and mean-variance (intersection) to set of plans that are dominated by FOSD. The higher this statistic, the more consistent it is mean-variance with FOSD. For the case of SOSD, this number is always one, the upper bound. The three ratios are 0.98 (49/50), 0.98(42/43) and 0.57 (4/7) when we compare all plans, plans of the same or higher quality based on star rating, and plans within the same brand. Interestingly, only in the last case, mean-variance rankings do not imply FOSD for the vast majority of cases. There are 43 percent of distributions that are dominated under FOSD that the mean-variance method assigns to the efficient frontier. Overall, the results show that mean-variance rankings are similar to SOSD rankings when comparing larger sets of plans. When we compare fewer plans, for example, plans sold by the same insurance company, both methods diverge in labeling dominated plans. Moreover, for this set, mean-variance rankings assign to the efficient frontier, plans that are dominated by FOSD.

Stepping back, another observation that can be made from table 3.2 is that no matter which set of plans we compare, the share of beneficiaries selecting efficient plans is always larger than the size of efficient sets. Thus, on average, consumers are clearly experiencing better outcomes than if they were to be randomly assigned to plans. What is not obvious from the table is whether enrollees do relatively better when we control for non-financial attributes (moving down any column) simply because the set of comparison plans gets reduced or because they are actually paying attention to non-financial attributes. We can start to test this by calculating: $e_t = \frac{\text{SED}_t}{\text{SI}_t}$. Where SED_t is the share of enrollees of type t that are selecting dominated plans and SI_t is the size of the dominated set for type t. If $e_t < 1$, it means that enrollees of type t are doing better relative to making a random choice, conditional on whether we account for preferences over quality, whereas, if $e_t > 1$ it means that their choice is worse relative to randomizing.

Figure 3.4: Efficiency Relative to Random Choices

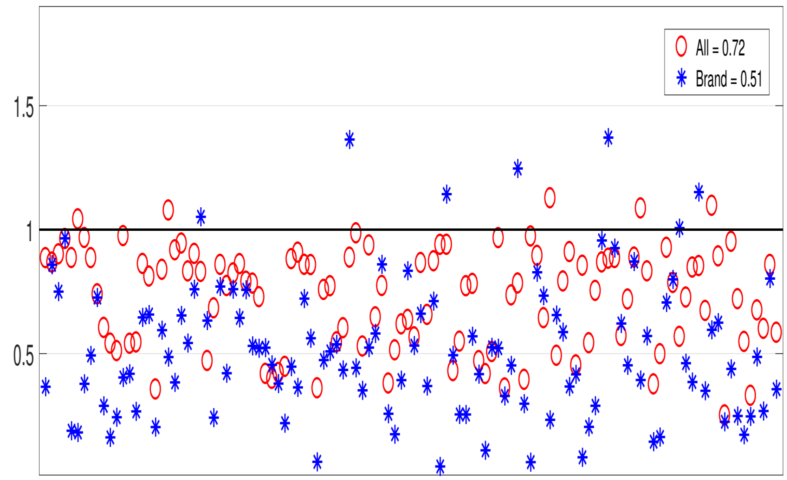


Figure 3.4: Figure 3.4 shows the efficiency measure (y-axis) for a random sample of types (x-axis). Each type is associated with two markers. The red circle shows the efficiency measure of each type when all plans are compared. The blue star shows the efficiency measure of each type when we control for non–financial characteristics at the brand level.

Figure 3.4 shows for a random sample of 20% different types (x-axis) the resulting e_t for two sets of plans. The circle measures how "efficient" are consumers, relative to a random choice when comparing choices among all plans. The star shows the same measure when comparing plans of the same brand. For most risk types, the star is below the circle. One explanation is that consumers do not compare all plans when deciding, perhaps because of strong preferences over insurance companies or because search costs preclude consumers from comparing the entire universe of plans. The means of e_t are approximately 0.7 and 0.5 when the "consideration" sets are composed of all plans and plans within the same brand respectively. We interpret figure 3.4 in positive terms; consumers' choices are more consistent with stochastic dominance when we control for brand effects. Whether consumers should or shouldn't pay attention to brands is a normative question. In principle, preferences over brands could be capturing omitted attributes that are valuable for consumers, while at the other extreme they could be used by consumers as a simple and potentially flawed heuristic to reduce the effort of comparing plans.

3.4.2 Heterogeneity in Choice Quality

The following two figures show the average welfare loss experienced by consumers who chose dominated plans in SOSD when comparing all plans (67%) and plans within the same brand (20%).

Figure 3.5: Welfare Loss Measures (All plans)

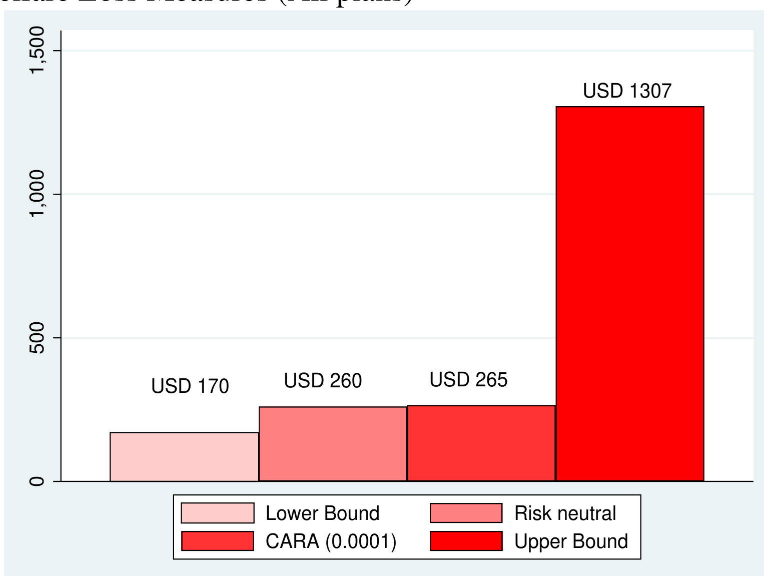


Figure 3.5: Figure 3.5 shows the average welfare loss of the four welfare measures introduced in section 2. The average is taken across consumers who selected dominated plans in SOSD.

Figure 3.5 shows the average welfare loss of the four welfare measures introduced in section 2. The average is taken across consumers who selected dominated plans in SOSD. The difference between the lower and upper bound is quite striking. The upper bound is approximately 8 times higher, implying that consumers could be willing to pay between \$170 and \$1,307 annually to choose an efficient plan, depending on their true preferences. Equally striking is the fact that the welfare loss is at least 34% of the annual premium that consumers face during this period (approximately \$500). Interestingly, the willingness to pay of risk neutral consumers and CARA consumers with absolute risk aversion equal to

0.0001 are relatively closer to the lower bound.

Finally, the next figure shows the same four measures but when we compare plans within the same brand.

Figure 3.6: Welfare Loss Measures (Brand)

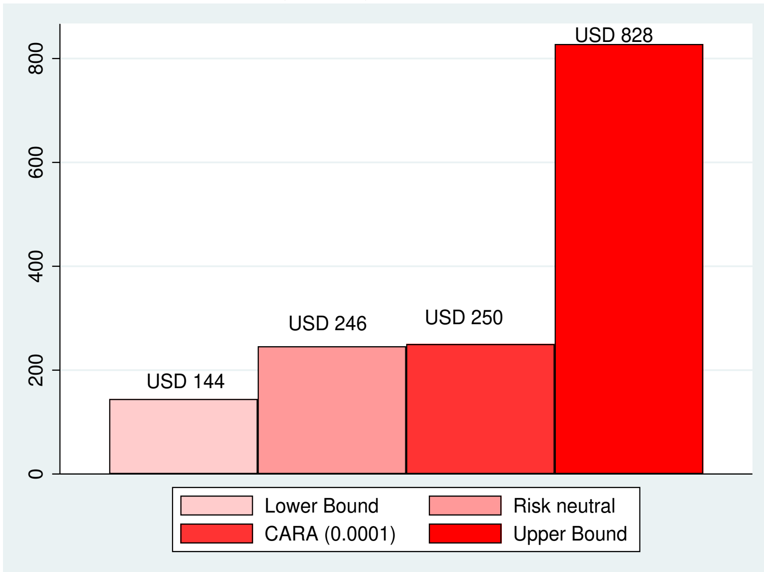


Figure 3.6: Figure 3.6 shows the average welfare loss of the four welfare measures introduced in section 2. The average is taken across consumers.

All our previous qualitative conclusions apply equally to the case when we control for brand effects. Comparing between Figure 3.6 and 3.5 allows us to draw conclusions that can inform discussion of the normativeness of brand effects in consumer preferences. Specifically, Figure 3.6 reveals that most of the welfare loss reported in Figure 3.5 is also experienced when consumers choose dominated plans within a given brand. The four welfare measures calculated under this last set represent 85% (lower bound), 95% (risk neutral), 94% (CARA) and 63% (upper bound) of the welfare loss when all plans are compared. Therefore, even if consumers use brands as a heuristic to reduce plan comparisons. Most of the welfare loss comes from not being able to select the best plan within each brand. The fact that consumers appear to have trouble when comparing plans within the same brand is consistent with previous studies that find inconsistent choices in employer-sponsored health insurance setting when no more than 5 plans are offered (Handel, 2013; Bhargava *et al.*, 2017). Finally, although from the intensive margin perspective there seem note to be differences when comparing all plans and plans sold by the same insurer. There are still larger differences at the extensive margin with 70% and 20% of consumer choosing dominated plans under each set.

We now show how these losses are distributed across consumers, and how they relate to socioeconomic attributes and plan characteristics. The next figure shows the distribution of the four welfare measures across consumers who selected dominated plans. In all subsequent analyses, we analyze the welfare losses when the comparison set includes all plans.

Figure 3.7: Distribution of Welfare Losses

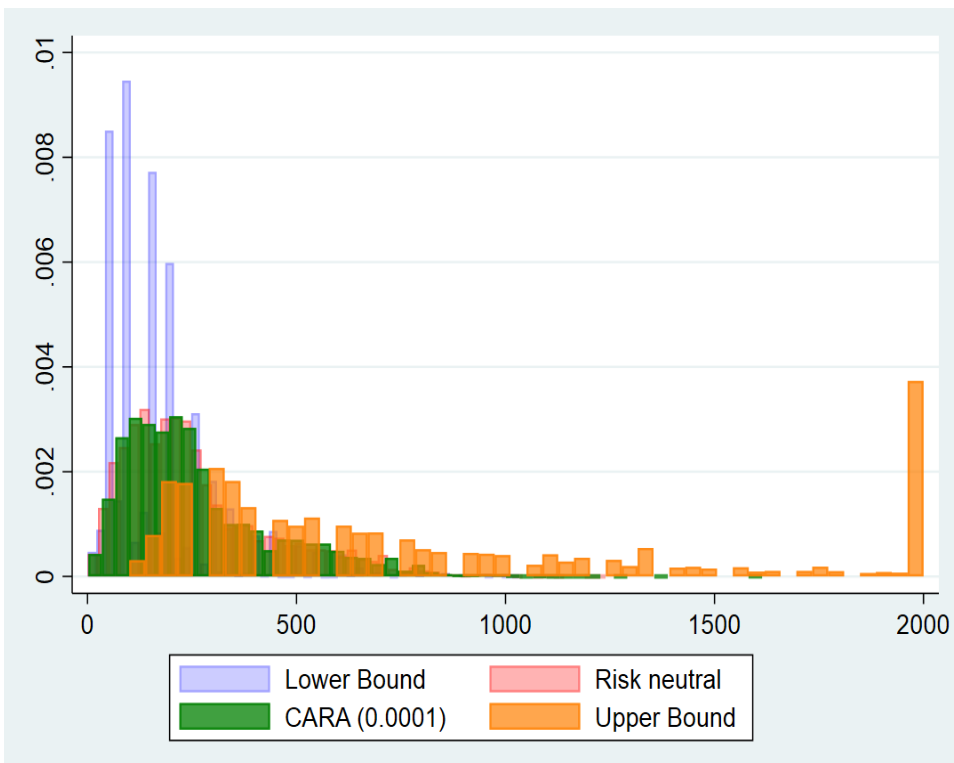


Figure 3.7: Figure 3.7 shows the distribution of welfare losses of the four welfare measures introduced in section 2.

The four distributions show that most consumers who choose dominated plans experience welfare losses similar to the average of the population. However, the long tails of the four distributions reveal that a small portion of consumers suffers much larger welfare losses. Identifying common characteristics of these consumers or aspects of the choice architecture that contribute to these losses could be important for targeting specific population groups and tailoring policies. To this end, we analyze the conditional association between dominated choices and demographics. Although we make no attempt to establish a causal link between the two, we believe that these correlations can provide relevant information to inform policy.

We first study the conditional association between dominated choices and our welfare measures with demographics and socioeconomic conditions such as age, gender, working status, income, marital status, race, clinically diagnosed chronic conditions, and self-assessed measures of health. As mentioned in section 4, we observe the diagnosis dates of more than 30 chronic medical conditions. To make the analysis more parsimonious we group them into four categories. Group 1 corresponds to chronic conditions that have below-average prevalence and below-average annual prescription drug costs. Examples are prostate and breast cancer. Group 2 corresponds to chronic conditions with above-average prevalence and below-average costs such as diabetes and anemia. Group 3 corresponds to conditions with below-average prevalence and above-average costs such as Alzheimer's and schizophrenia. Finally, Group 4 corresponds to conditions with above-average prevalence and above-average costs. It is important to understand that the three[17] categories are not mutually exclusive for a given individual. An individual can possess none, one, or more than one chronic condition in each group. Figure 3.8 shows the resulting groups according to our categorization of chronic conditions.

[17]In principle, our definition generates four categories for chronic conditions but we don't observe any chronic conditions falling in the last category, Group 4.

Figure 3.8: Chronic Conditions Group

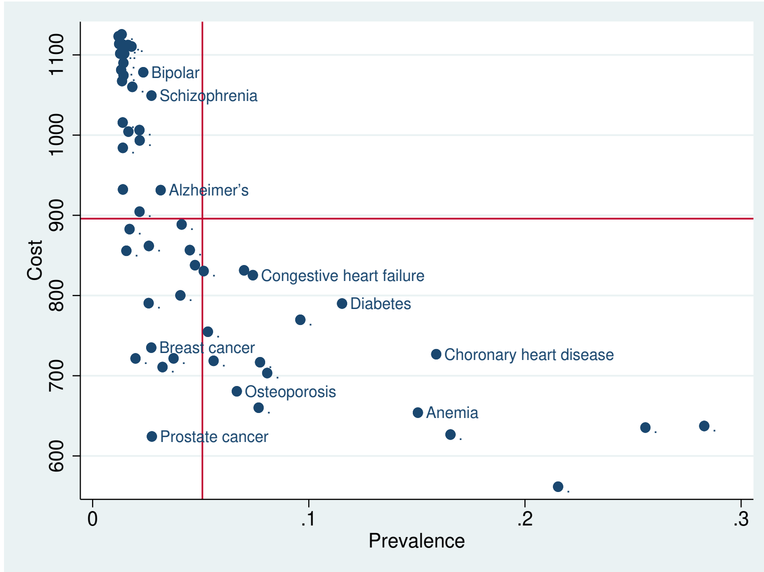


Figure 3.8: Figure 3.8 shows the resulting groups according to our categorization of chronic conditions.

Table 3.3 shows regression based estimates for the conditional association between SOSD and demographic conditions. The first column reports the marginal effects from a logit model where the dependent variable is an indicator for whether the choice is dominated. A positive marginal effect indicates that the likelihood of selecting a dominated plan is higher. The second column shows regression based estimates for the conditional association between mean-variance and demographic conditions. We showed previously how similar are empirically, mean-variance efficient sets and stochastic dominance efficient sets. The comparison of column 1 and column 2 will give us information on how similar are both methods in characterizing the "average" enrollee who is more likely to choose dominated plans. The third column reports the coefficients from an OLS regression where the dependent variable is the welfare loss of a CARA consumer. It indicates, conditional on selecting a dominated plan, which demographic characteristics have stronger conditional associations with the size of the welfare loss. Columns four and five report estimates from a similar regression based on the lower and upper bound welfare measures respectively.

If we compare the first two columns, the sign and size of marginal effects of both models are very similar. This is not surprising since efficient sets are almost the same when all plans are compared. Therefore, both methods coincide when characterizing consumers who are more likely to select dominated plans. For example, both methods agree that singles, e.g. widowed and divorced enrollees, are more likely to select dominated plans relative to married enrollees. In terms of race, black enrollees have higher chances of selecting dominated plans relative to white enrollees (11 percentage points). The variable "no default" refers to consumers who don't have a default plan on their menu. For example, consumers who are new to the market or consumers who are forced to switch because their chosen plan is not available next year. Interestingly, these consumers are less likely to choose dominated plans relative to consumers whose last year plan is available. This is consistent with the literature on inertia, Handel (2013); Polyakova (2016); Ketcham *et al.* (2019). The variables "excellent", "very good", "good", "fair", and "poor" refer to self-assessed measures of health. There is no systematic correlation with these variables and choosing dominated plans. However, conditional on choosing a dominated plan, people who perceive their health status as "excellent" and "very good" experience higher welfare losses according to the upper bound welfare measure. This could suggest that enrollees who choose dominated plans because they overestimated their health status suffer higher welfare losses. Finally, enrollees who choose dominated plans and don't have chronic conditions of Group 2, also suffer higher welfare losses relative to consumers who have chronic conditions belonging to this group. A possible explanation is that the uncertainty that they face is higher relative to enrollees who deal with this type of chronic condition.

Table 3.4 shows the conditional association of the same measures with characteristics of the "menu" and plans. Some of the attributes that are included in the regression are enrollment year, number of plans, number of brands, premiums, and deductibles. The

SOSD and mean-variance measures for the dependent variable generally have the same sign and similar magnitudes for marginal effects. The signs on the plan attributes are generally in the expected direction. Higher premiums or deductibles, all else equal, are positively associated with dominated choices. With the same logic, plans that offer coverage in the gap, all else equal, are negatively associated with dominated choices. When the market was first introduced in 2006, consumers were more likely to choose dominated plans than in subsequent years. This is consistent with the hypothesis of consumers learning, Ketcham *et al.* (2012).

Table 3.3: Correlation: SOSD - Demographics

Variables	ME SOSD	ME MV	CARA	Lower B	Upper B
age	0.00273	0.00285	1.348*	1.885*	25.34
	(0.00)	(0.00)	(0.81)	(1.13)	(60.17)
male	-0.0137	-0.018	10.46	10.93	-569.3
	(0.03)	(0.02)	(10.62)	(14.85)	(588.50)
working	-0.00965	-0.0118	-2.212	27.07	132.7
	(0.04)	(0.03)	(13.23)	(16.76)	(652.20)
income [30k -50k]	0.0338	0.014	-5.287	-2.394	-68.86
	(0.03)	(0.03)	(12.77)	(18.23)	(612.60)
income $>$ 50k	-0.0549	-0.0269	-13.42	-34.1	-815.4
	(0.04)	(0.04)	(16.39)	(27.30)	(728.50)
widowed	0.0684**	0.0600**	-4.165	-10.01	87.19
	(0.03)	(0.03)	(12.96)	(18.16)	(757.60)
divorced	0.0914**	0.0703	-27.07	-27.08	2461
	(0.05)	(0.04)	(16.59)	(24.60)	(1506.00)
black	0.113*	0.208***	-78.65***	-14.14	2053
	(0.07)	(0.04)	(21.34)	(18.48)	(2487.00)
asian	-0.0382	0.041	-48.84	-117.7***	3,105***
	(0.21)	(0.16)	(77.32)	(34.94)	(1174.00)
no default	-0.122***	-0.027	-13.56	14.15	-222.4
	(0.03)	(0.03)	(13.13)	(20.73)	(790.30)
group1	-0.114	-0.148	55.37	15.92	-1,723*
	(0.11)	(0.10)	(46.60)	(39.94)	(1005.00)
group2	0.0148	-0.108	-11.49	-36.8	-2,222**
	(0.06)	(0.07)	(24.07)	(24.35)	(938.10)
group3	0.000775	-0.206	190.4**	13.7	1147
	(0.14)	(0.15)	(86.07)	(49.89)	(1014.00)
excellent	-0.0301	0.0194	-39.55***	-0.53	1,740*
	(0.04)	(0.03)	(14.09)	(22.30)	(935.80)
very good	-0.023	-0.00663	-0.373	7.237	2,141**
	(0.03)	(0.03)	(13.45)	(17.85)	(891.80)
fair	-0.022	0.0317	-14.37	-8.545	968.6
	(0.04)	(0.03)	(16.65)	(21.86)	(658.70)
poor	-0.0764	-0.0678	-17.36	7.854	1,208
	(0.05)	(0.05)	(23.46)	(30.18)	(1,008)
Constant			112.1*	-35.26	-1,641
			(65.07)	(90.08)	(4,599)
Observations	1,744	1,744	1,138	220	220
R-squared			0.06	0.097	0.199

Notes: Table 3.3 shows regression based estimates for the conditional association between SOSD and demographic conditions. The first column reports the marginal effects from a logit model where the dependent variable is an indicator of whether the choice is dominated. The second column shows regression based estimates for the conditional association between mean-variance and demographic conditions. The third column reports the coefficients from an OLS regression where the dependent variable is the welfare loss of a CARA consumer. It indicates, conditional on selecting a dominated plan, which demographic characteristics have stronger conditional associations with the size of the welfare loss. Columns four and five report estimates from a similar regression based on the lower and upper bound welfare measures respectively.

When comparing plans within the same brand, the characterization of consumers who select dominated plans starts differing between the SOSD and mean-variance measures. This is shown in table 3.5. Stochastic dominance rankings identify people who are divorced, with annual income between 30 and 50 thousand dollars, and people in group 3 as being more likely to choose dominated plans within the same brand. The size of the marginal effect in the last group is striking. Enrollees who have chronic conditions in group 3 (e.g. Alzheimer's and schizophrenia) have higher probability of choosing dominated plans within the same brand by 38 percentage points.

Table 3.4: Correlation: SOSD - Menu

Variables	ME SOSD	ME MV	CARA	Lower B	Upper B
2007	0.000717	-0.0695***	-64.81***	-75.66***	759.6
	(0.03)	(0.03)	(16.88)	(22.20)	(810.10)
2008	-0.188***	-0.279***	-124.2***	-100.6***	146.4
	(0.05)	(0.04)	(21.46)	(29.92)	(1290.00)
2009	-0.264***	-0.271***	-200.5***	-144.3***	498.6
	(0.05)	(0.05)	(20.63)	(27.30)	(1481.00)
2010	-0.490***	-0.539***	-253.3***	-189.7***	603.2
	(0.04)	(0.04)	(21.31)	(28.85)	(1001.00)
#plan	-0.00068	0.00674***	-0.352	-0.606	-57.07
	(0.00)	(0.00)	(1.20)	(1.46)	(60.50)
#brand	0.00599	-0.0199***	7.191***	5.866**	115.1
	(0.01)	(0.00)	(2.07)	(2.37)	(90.93)
premium	0.00201***	0.00106***	0.613***	0.632***	1.763
	(0.00)	(0.00)	(0.03)	(0.06)	(1.45)
deductible	0.000741***	0.000825***	0.631***	0.183***	0.105
	(0.00)	(0.00)	(0.04)	(0.04)	(2.48)
gap brand	-0.671***	-0.756***	-228.9***	-631.5***	-1597
	(0.01)	(0.01)	(45.64)	(50.94)	(1182.00)
gap any	-0.219***	-0.470***	42.83***	-67.66***	-744.2
	(0.06)	(0.04)	(12.72)	(22.68)	(690.80)
no default	-0.0396*	-0.0383	6.302	-3.071	219.9
	(0.02)	(0.02)	(9.13)	(14.27)	(459.50)
Constant			-352.4***	-317.3***	938.6
			(54.84)	(74.59)	(2567.00)
Observations	2,752	2,752	1,820	345	345
R-squared			0.487	0.586	0.02

Notes: Table 3.4 shows regression based estimates for the conditional association between SOSD and characteristics of the "menu" and plans. The first column reports the marginal effects from a logit model where the dependent variable is an indicator of whether the choice is dominated. The second column shows regression based estimates for the conditional association between mean-variance and menu characteristics. The third column reports the coefficients from an OLS regression where the dependent variable is the welfare loss of a CARA consumer. It indicates, conditional on selecting a dominated plan, which menu and plan characteristics have stronger conditional associations with the size of the welfare loss. Columns four and five report estimates from a similar regression based on the lower and upper bound welfare measures respectively.

Table 3.5: Correlation: SOSD - Demographics

Variables	ME SOSD	ME MV
age	0.00194	-0.000618
	(0.00)	(0.00)
male	0.00213	-0.0219
	(0.02)	(0.02)
working	-0.0175	-0.033
	(0.03)	(0.03)
income [30k -50k]	0.0415*	-0.00979
	(0.03)	(0.03)
income > 50k	-0.0359	-0.0415
	(0.03)	(0.03)
widowed	-0.00507	-0.0101
	(0.02)	(0.02)
divorced	0.0765*	0.0115
	(0.04)	(0.04)
black	-0.0235	-0.0311
	(0.06)	(0.06)
asian	-0.037	0.155
	(0.15)	(0.19)
no default	-0.0406*	-0.00236
	(0.02)	(0.02)
group2	-0.0506	-0.0829**
	(0.04)	(0.04)
group3	0.386***	-0.0594
	(0.14)	(0.09)
Observations	1,744	1,744

Notes: Table 3.5 compares regression based estimates for the conditional association between SOSD and demographic conditions with the same estimates using mean-variance. The table shows results when we only compare plans sold by the same insurance company.

Finally, figures 3.9 and 3.10 show regional variation in the probability of choosing dominated plans based on the SOSD and mean-variance measures, across states. Recall that on average, 67% of consumers select dominated plans by SOSD and 76% in mean-variance. The maps show that both measures have similar predictions for which states have relatively lower and higher rates of dominated choices.

Figure 3.9: Extensive Margin by Region: SOSD

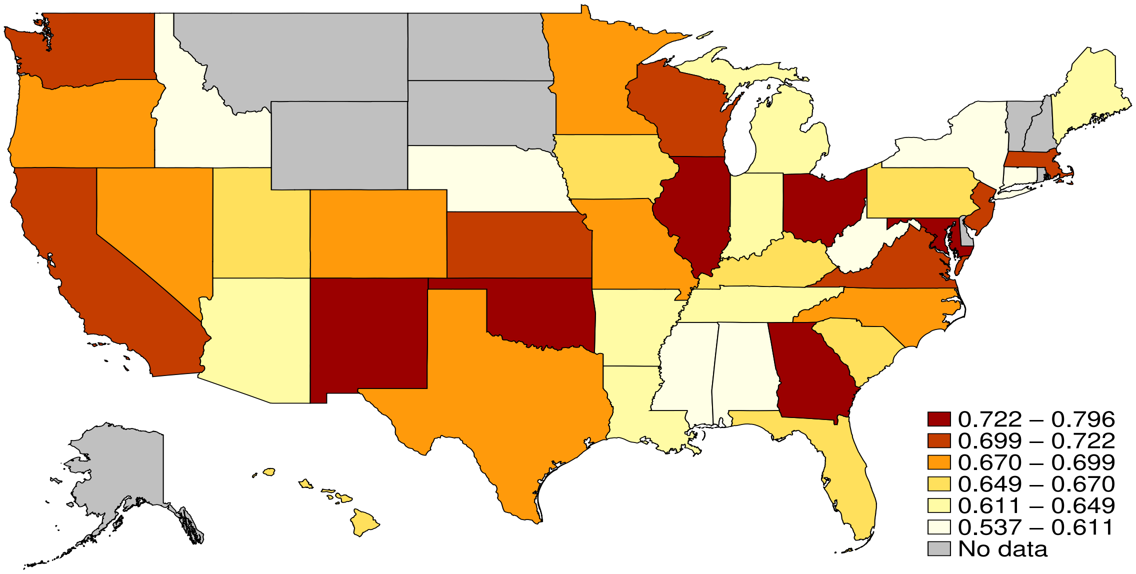


Figure 3.9: Figure 3.9 show regional variation in the probability of choosing dominated plans based on the SOSD measure across states.

Figure 3.10: Extensive Margin by Region: MV

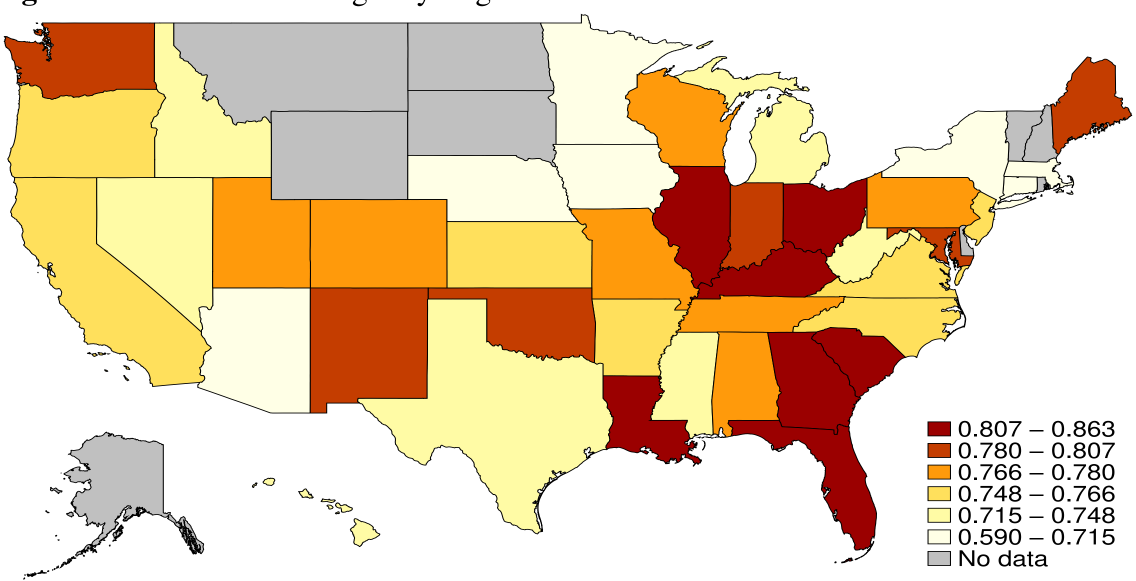


Figure 3.10: Figure 3.10 show regional variation in the probability of choosing dominated plans based on the MV measure across states.

3.5 Conclusions

Overall, we find that mean-variance (MV) and second order stochastic dominance measures show similar results when we don't control for non-financial attributes. Both methods imply that approximately, 70% of consumers select dominated plans and coincide when characterizing consumers who are more likely to select dominated plans. However, MV and SOSD measures differ in which sets of plans they label as dominated and who choose them when we control for non-financial attributes. When only comparing plans sold by the same insurance company, enrollees seem to behave more rationally in the sense that their choices conform more with stochastic dominance axioms. Enrollees who have chronic conditions with below-average prevalence and above-average costs such as Alzheimer's and schizophrenia have a higher probability of choosing dominated plans within the same brand by 38 percentage points.

Regarding the welfare implications of choosing dominated plans, we showed that the welfare ambiguity of this behavior is large. The average welfare loss captured by our upper bound is eight times higher than the lower bound, $1,307 and $170 respectively. This implies that the average welfare loss is never smaller than 34% of the annual premium. This difference doesn't change much even when we only compare plans sold by the same insurance company. This result sheds light on the need for additional sources of information to elicit true preferences if we want to credibly assess consumer's welfare in these markets.

Chapter 4

CONCLUSIONS AND FUTURE RESEARCH

Making decisions in health care markets is difficult. The mix of actors involved—physicians, hospitals, insurance companies and public programs—together with the inherent uncertainty of health care requires consumers to exert high cognitive effort when navigating these markets. This may make the "value of information" especially high with individuals seeking advice and information from family, friends, peers, and specialists.

Chapter 2 of this essay illustrated that individual choices are influenced by others. It helped advance prior literature that assessed decision-making in insurance markets by relaxing the assumption that consumers behave as independent agents. Married couples' choices of insurance plans are highly correlated. In future research it would be interesting to investigate whether peer-effects extend beyond immediate family members and close friends. In particular, many US citizens live in communities that provide some degree of assistance to residents. From communities where seniors are totally independent to communities where they are fully assisted, retirement communities provide an ideal "laboratory" to investigate how peers and specialists affect individual decision-making.

In addition to being an ideal setting to study the effect of social interactions on individual choices, retirement communities are a fascinating and understudied institution. There is already a large literature on "valued added" by teachers and schools that aims to measure how they affect education and labor market outcomes (e.g. Chetty et al., 2014). Similarly, we can investigate the value added by peers and specialists in retirement communities when it comes to choosing health insurance plans. Population aging makes it especially important to study this issue. The US Census Bureau projects that by 2035 senior citizens will exceed children in population size. Understanding whether seniors' welfare could be improved by informational economies of scale or other policy mechanisms that could leverage social interactions in retirement communities may be important to inform housing and urban policies and also have important fiscal implications for the taxpayer burden of Medicare programs.

In chapter 3 of this essay I advanced prior literature by using a non-parametric method to assess decision-making quality in insurance markets. It relies on the rational axioms of stochastic dominance. These axioms are normativelly compelling for economists as descriptions of rational consumer behavior but they may be difficult to explain to consumers who are looking for guidance on how to choose health insurance plans in real markets. Policies that can deliver information about the financial implications of insurance plan choices in a simple manner could help to inform consumers' choices in these markets. Star ratings is a example. In the case of health insurance plans, existing methods for developing star ratings are based on consumer satisfaction surveys. In principle, the economic content of star ratings could be extended to include information from stochastic dominance rankings. There is an interesting conceptual link between star ratings and stochastic dominance rankings. Star ratings provide a unanimous signal about plan rankings for all consumers, similar to the way that stochastic dominance rankings provide a unanimous (partial) ordering of financial lotteries. It would be interesting to test whether CMS's star ratings for Medicare Part D plans are consistent with stochastic dominance rankings. If not, a second research question would be to consider how CMS could redesign the star rating system to conform with stochastic dominance rankings. The challenge is to improve the economic rigor of the star rating system while continuing to present the information in an intuitive way for consumers who are unfamiliar with stochastic dominance concepts.

Taken all together, the results of both chapters reveal that the assumptions that we make about stochastic terms in utility functions are at least as important as the assumptions we make on the deterministic theory. When estimating the welfare loss of choosing dominated plans under risk neutrality and CARA, both magnitudes are between 150 and 250 dollars per year. How can this be consistent with estimated welfare losses of inertia and pooling above 1,000 dollars? When researchers use stochastic choice or random utility to estimate the welfare loss of inertia or pooling, the assumptions made on ϵ_{ij} will also impact the size of these measures. In most applied models of choices in insurance markets, preferences are assumed to have a random component that is type I EV or normal. Importantly, these random terms are assumed to be *iid* across people and time (Abaluck and Gruber, 2011; Handel, 2013; Handel and Kolstad, 2015; Polyakova, 2016; Ketcham *et al.*, 2019). This is important since under the standard *iid* assumption of ϵ_{ij} across time, the error term can not contribute to explain inertia. Therefore, when measuring the welfare cost of inertia in the space of the deterministic theory, it will be higher under the assumption of independent random terms compared to a model that allows for dependencies. The welfare bounds estimated in chapter 3 show the contribution of the deterministic theory for the welfare losses. We already saw that the estimates of inertia in different studies with very different deterministic theories are still very similar. Understanding how robust are these measures to other assumptions about ϵ_{ij} is a promising area for future research.

Finally, my dissertation research combines insights from the two canonical frameworks for measuring consumer preferences: revealed preference and stated preference. While stated preferences methods are often better at controlling for cofounders, revealed preference methods are often thought to be more informative about how consumers behave in real-world scenarios that are difficult to replicate on a survey or in the laboratory. I believe that much can be learned with a careful and creative research design that combines the two methods. The availability of large-scale surveys like the Health and Retirement Study (HRS) combined with original survey data from the field offer a fruitful avenue to explore questions that seek to measure consumer preferences in different environments that matter for health economics. Applications where I see potential for combining the two frameworks to develop new insights include eliciting risk aversion parameters, estimating the value of a statistical life, and estimating the willingness to pay for disease prevention.

REFERENCES

Abaluck, J. and J. Gruber, "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program", American Economic Review **101**, 4, 1180–1210, URL `http://pubs.aeaweb.org/doi/10.1257/aer.101.4.1180` (2011).

Abaluck, J. and J. Gruber, "Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance", American Economic Review **106**, 8, 2145–2184, URL `http://pubs.aeaweb.org/doi/10.1257/aer.20130778` (2016).

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

Addoum, J. M., "Household portfolio choice and retirement", Review of Economics and Statistics **99**, 5, 870–883 (2017).

Akerlof, G. A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism", The Quarterly Journal of Economics **84**, 3, 488–500 (1970).

Allcott, H. and D. Taubinsky, "Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market", American Economic Review **105**, 8, 2501–2538 (2015).

Apesteguia, J. and M. A. Ballester, "A measure of rationality and welfare", Journal of Political Economy **123**, 6, 1278–1310 (2015).

Apesteguia, J. and M. A. Ballester, "Monotone Stochastic Choice Models: The Case of Risk and Time Preferences", Journal of Political Economy **126**, 1, 74–106, URL `https://www.journals.uchicago.edu/doi/10.1086/695504` (2018).

Aumann, R. J. and R. Serrano, "An economic index of riskiness", Journal of Political Economy **116**, 5, 810–836 (2008).

Barseghyan, L., F. Molinari, T. O'Donoghue and J. C. Teitelbaum, "Distinguishing probability weighting from risk misperceptions in field data", in "American Economic Review", vol. 103, pp. 580–585 (2013a).

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

Barseghyan, L., J. Prince and J. C. Teitelbaum, "Are Risk Preferences Stable across Contexts? Evidence from Insurance Data", American Economic Review **101**, 2, 591–631, URL `http://pubs.aeaweb.org/doi/10.1257/aer.101.2.591` (2011).

Barsky, R. B., F. T. Juster, M. S. Kimball and M. D. Shapiro, "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study", The Quarterly Journal of Economics **112**, 2, 537–579, URL `https://academic.oup.com/qje/article-lookup/doi/10.1162/003355397555280` (1997).

Bernheim, B. D., S. Dellavigna and D. Laibson, *Handbook of Behavioral Economics - Foundations and Applications 2* (2019).

Bernheim, B. D., A. Fradkin and I. Popov, "The welfare economics of default options in 401(k) plans", in "American Economic Review", vol. 105, pp. 2798–2837 (American Economic Association, 2015).

Bernheim, B. D. and A. Rangel, "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics [*]", Quarterly Journal of Economics **124**, 1, 51–104, URL `https://academic.oup.com/qje/article-lookup/doi/10.1162/qjec.2009.124.1.51` (2009).

Beshears, J., J. J. Choi, D. Laibson and B. C. Madrian, "How are preferences revealed?", Journal of Public Economics **92**, 8-9, 1787–1794, URL `https://www.sciencedirect.com/science/article/abs/pii/S0047272708000728` (2008).

Bhargava, S., G. Loewenstein and J. Sydnor, "Choose to lose: Health plan choices from a menu with dominated options", Quarterly Journal of Economics (2017).

Birnbaum, M. H. and J. P. Bahra, "Gain-loss separability and coalescing in risky decision making", Management Science **53**, 6, 1016–1028 (2007).

Birnbaum, M. H. and J. B. Navarrete, "Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence", Journal of Risk and Uncertainty **17**, 1, 49–79 (1998).

Bone, J., "Risk-sharing CARA individuals are collectively EU", Economics Letters **58**, 3, 311–317 (1998).

Borch, K., "A Note on Uncertainty and Indifference Curves", The Review of Economic Studies **36**, 1, 1 (1969).

Borghans, L., B. H. H. Golsteyn, J. J. Heckman and H. Meijers, "Gender Differences in Risk Aversion and Ambiguity Aversion", Journal of the European Economic Association **7**, 2-3, 649–658, URL `https://academic.oup.com/jeea/article-lookup/doi/10.1162/JEEA.2009.7.2-3.649` (2009).

Bundorf, M. K., J. Levin and N. Mahoney, "Pricing and welfare in health plan choice", American Economic Review **102**, 7, 3214–3248 (2012).

Cain, C. and W. Derek, "Maiden Names, on the Rise Again", New York Times. **www.nytimes.com/2015/06/28/upshot/maiden-names-on-the-rise-again.html**, URL `www.nytimes.com/2015/06/28/upshot/maiden-names-on-the-rise-again.html` (2015).

Carey, C., “Technological Change and Risk Adjustment: Benefit Design Incentives in Medicare Part D”, American Economic Journal: Economic Policy **9**, 1, 38–73, URL http://pubs.aeaweb.org/doi/10.1257/pol.20140171 (2017).

Chandra, A., B. Handel and J. Schwartzstein, “Behavioral economics and healthcare markets”, Handbook of Behavioral Economics: Applications and Foundations 1 **2**, 459–502, URL https://www.sciencedirect.com/science/article/pii/S2352239918300241 (2019).

Chiappori, P. and B. Salanie, “Testing for Asymmetric Information in Insurance Markets”, Journal of Political Economy **108**, 1, 56–78, URL https://www.journals.uchicago.edu/doi/10.1086/262111 (2000).

Chiappori, P.-A., “Collective Labor Supply and Welfare”, Journal of Political Economy **100**, 3, 437–467, URL https://www.journals.uchicago.edu/doi/10.1086/261825 (1992).

Chiappori, P.-A. and M. Mazzocco, “Static and Intertemporal Household Decisions”, Journal of Economic Literature **55**, 3, 985–1045, URL http://pubs.aeaweb.org/doi/10.1257/jel.20150715 (2017).

Choi, S., S. Kariv, W. Müller and D. Silverman, “Who Is (More) Rational?”, URL https://www.jstor.org/stable/42920858 (2014).

Cohen, A. and L. Einav, “Estimating Risk Preferences from Deductible Choice”, American Economic Review **97**, 3, 745–788, URL http://pubs.aeaweb.org/doi/10.1257/aer.97.3.745 (2007).

Courbage, C., G. Montoliu-Montes and B. Rey, “How vulnerable is risk aversion to wealth, health and other risks? An empirical analysis for Europe”, Working Papers (2018).

Cutler, D. M., A. Finkelstein and K. McGarry, “Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance”, American Economic Review **98**, 2, 157–162, URL http://pubs.aeaweb.org/doi/10.1257/aer.98.2.157 (2008).

Cutler, D. M. and S. J. Reber, “Paying for health insurance: The trade-off between competition and adverse selection”, Quarterly Journal of Economics **113**, 2, 433–466 (1998).

Davidson, R., J.-Y. Duclos, R. Davidson and J.-Y. Duclos, “Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality”, Econometrica **68**, 6, 1435–1464, URL http://econpapers.repec.org/article/ecmemetrp/v{_}3a68{_}3ay{_}3a2000{_}3ai{_}3a6{_}3ap{_}3a1435-1464.htm (2000).

de Meza, D. and D. C. Webb, “Advantageous Selection in Insurance Markets”, The RAND Journal of Economics **32**, 2, 249–262, URL http://doi.wiley.com/10.2307/2696408 (2001).

Decker, S. and H. Schmitz, “Health shocks and risk aversion”, Journal of Health Economics **50**, 156–170, URL https://www.sciencedirect.com/science/article/abs/pii/S0167629616302260 (2016).

Dellavigna, S., "Psychology and economics: Evidence from the field", Journal of Economic Literature **47**, 2, 315–372 (2009).

Denuit, M., J. Dhaene, M. Goovaerts and R. Kaas, *Actuarial Theory for Dependent Risks* (John Wiley & Sons, Ltd, Chichester, UK, 2005), URL http://doi.wiley.com/10.1002/0470016450.

Diamond, P. A. and J. A. Hausman, "Contingent Valuation: Is Some Number Better than No Number?", Journal of Economic Perspectives **8**, 4, 45–64, URL http://pubs.aeaweb.org/doi/10.1257/jep.8.4.45 (1994).

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp and G. G. Wagner, "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences", Journal of the European Economic Association **9**, 3, 522–550, URL https://academic.oup.com/jeea/article-lookup/doi/10.1111/j.1542-4774.2011.01015.x (2011).

Eeckhoudt, L., P. Godfroid and C. Gollier, "Willingness to pay, the risk premium and risk aversion", Economics Letters **55**, 3, 355–360 (1997).

Eika, L., M. Mogstad and B. Zafar, "Educational Assortative Mating and Household Income Inequality", Journal of Political Economy p. forthcoming (2019).

Einav, L., A. Finkelstein, R. Kluender and P. Schrimpf, "Beyond Statistics: The Economic Content of Risk Scores", American Economic Journal: Applied Economics **8**, 2, 195–224, URL http://dx.doi.org/10.1257/app.20150131 (2016).

Einav, L., A. Finkelstein, I. Pascu and M. R. Cullen, "How General Are Risk Preferences? Choices under Uncertainty in Different Domains", American Economic Review **102**, 6, 2606–2638, URL http://pubs.aeaweb.org/doi/10.1257/aer.102.6.2606 (2012).

Ericson, K. M. M., "Consumer Inertia and Firm Pricing in the Medicare Part D Prescription Drug Insurance Exchange", American Economic Journal: Economic Policy **6**, 1, 38–64, URL http://pubs.aeaweb.org/doi/10.1257/pol.6.1.38 (2014).

Fadlon, I. and T. H. Nielsen, "Family Health Behaviors", American Economic Review **109**, 9, 3162–3191, URL https://pubs.aeaweb.org/doi/10.1257/aer.20171993 (2019).

Fang, H., M. Keane and D. Silverman, "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market", Journal of Political Economy **116**, 2, 303–350, URL https://www.journals.uchicago.edu/doi/10.1086/587623 (2008).

Fernández, R., N. Guner and J. Knowles, "Love and money: A theoretical and empirical analysis of household sorting and inequality", Quarterly Journal of Economics **120**, 1, 273–344 (2005).

Finkelstein, A. and K. McGarry, "Multiple dimensions of private information: Evidence from the long-term care insurance market", American Economic Review **96**, 4, 938–958 (2006).

Fleurbaey, M. and E. Schokkaert, "Equity in Health and Health Care", in "Handbook of Health Economics", vol. 2, pp. 1003–1092 (Elsevier B.V., 2011).

Florens, J.-P., V. Marimoutou and A. Péguin-Feissolle, "Econometric modeling and inference: Themes in modern econometrics", Cambridge University Press p. 518, URL http://books.google.com/books?hl=en{&}lr={&}id=MrkUeviIvYkC{&}oi=fnd{&}pg=PA3{&}dq=cointegration+nonparametric{&}ots=eRsS1pLdnM{&}sig=bC0Rs-uO97J823NaKBbZ9Yj4bsg{%}5Cnpapers2://publication/uuid/BF814B9A-1FC6-411F-B80F-585F867666F9 (2007).

Fonseca, R., K. J. Mullen, G. Zamarro and J. Zissimopoulos, "What Explains the Gender Gap in Financial Literacy? The Role of Household Decision Making", Journal of Consumer Affairs **46**, 1, 90–106 (2012).

Gerber, H. U., "On Additive Premium Calculation Principles", ASTIN Bulletin **7**, 3, 215–222 (1974).

Gollier, C., Gollier and Christian, *The economics of risk and time*, vol. 1 (MIT Press, 2001), URL https://econpapers.repec.org/bookchap/mtptitles/0262572249.htm.

Handel, B. R., "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts", American Economic Review **103**, 7, 2643–2682, URL http://pubs.aeaweb.org/doi/10.1257/aer.103.7.2643 (2013).

Handel, B. R. and J. T. Kolstad, "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare", American Economic Review **105**, 8, 2449–2500, URL http://pubs.aeaweb.org/doi/10.1257/aer.20131126 (2015).

Handel, B. R., J. T. Kolstad and J. Spinnewijn, "Information frictions and adverse selection: Policy interventions in health insurance markets", Review of Economics and Statistics **101**, 2, 326–340 (2019).

Harless, D. W., C. F. Camerer, D. Harless and C. F. Camerer, "The Predictive Utility of Generalized Expected Utility Theories", Econometrica **62**, 6, 1251–89, URL https://econpapers.repec.org/article/ecmemetrp/v{_}3a62{_}3ay{_}3a1994{_}3ai{_}3a6{_}3ap{_}3a1251-89.htm (1994).

Harsanyi, J. C., "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility", Journal of Political Economy **63**, 4, 309–321, URL https://www.journals.uchicago.edu/doi/10.1086/257678 (1955).

Heiss, F., D. McFadden and J. Winter, "Regulation of private health insurance markets: Lessons from enrollment, plan type choice, and adverse selection in Medicare Part D", Tech. rep., National Bureau of Economic Research, Cambridge, MA, URL http://www.nber.org/papers/w15392.pdf (2009).

Heufer, J., "Nonparametric comparative revealed risk aversion", Journal of Economic Theory **153**, 1, 569–616 (2014).

Hey, J. D. and C. Orme, “Investigating Generalizations of Expected Utility Theory Using Experimental Data”, Econometrica **62**, 6, 1291–1326, URL https://www.jstor.org/stable/2951750?origin=crossref (1994).

Hippisley-Cox, J., C. Coupland, M. Pringle, N. Crown and V. Hammersley, “Married couples’ risk of same disease: cross sectional study.”, BMJ (Clinical research ed.) **325**, 7365, 636, URL http://www.ncbi.nlm.nih.gov/pubmed/12242177http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC126307 (2002).

Ho, K., J. Hogan and F. Scott Morton, “The impact of consumer inattention on insurer pricing in the Medicare Part D program”, The RAND Journal of Economics **48**, 4, 877–905, URL http://doi.wiley.com/10.1111/1756-2171.12207 (2017).

Hsu, J., J. Huang, V. Fung, M. Price, R. Brand, R. Hui, B. Fireman, W. Dow, J. Bertko and J. P. Newhouse, “Market Watch: Distributing $800 billion: An early assessment of medicare part D risk adjustment”, Health Affairs **28**, 1, 215–225 (2009).

Kaas, R., M. Goovaerts, J. Dhaene and M. Denuit, *Modern Actuarial Risk Theory* (Springer Berlin Heidelberg, 2008).

Ketcham, J. D., N. V. Kuminoff and C. A. Powers, “Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program: Comment”, American Economic Review **106**, 12, 3932–3961, URL http://pubs.aeaweb.org/doi/10.1257/aer.20131048 (2016).

Ketcham, J. D., N. V. Kuminoff and C. A. Powers, “Estimating the Heterogeneous Welfare Effects of Choice Architecture”, International Economic Review **60**, 3, 1171–1208, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/iere.12382 (2019).

Ketcham, J. D., C. Lucarelli, E. J. Miravete and M. C. Roebuck, “Sinking, Swimming, or Learning to Swim in Medicare Part D”, The American Economic Review **102**, 2639–2673, URL http://www.jstor.org.ezproxy1.lib.asu.edu/stable/41724667 (2012).

Ketcham, J. D., C. Lucarelli and C. A. Powers, “Paying Attention or Paying Too Much in Medicare Part D”, American Economic Review **105**, 1, 204–233, URL http://pubs.aeaweb.org/doi/10.1257/aer.20120651 (2015).

Kimball, M. S., C. R. Sahm and M. D. Shapiro, “Imputing Risk Tolerance From Survey Responses”, Journal of the American Statistical Association **103**, 483, 1028–1038, URL https://www.tandfonline.com/doi/full/10.1198/016214508000000139 (2008).

Kling, J. R., S. Mullainathan, E. Shafir, L. C. Vermeulen and M. V. Wrobel, “Comparison Friction: Experimental Evidence from Medicare Drug Plans”, The Quarterly Journal of Economics **127**, 1, 199–235, URL https://academic.oup.com/qje/article-lookup/doi/10.1093/qje/qjr055 (2012).

Kourouxous, T. and T. Bauer, “Violations of dominance in decision-making”, (2019).

Kunreuther, H. C., M. V. Pauly and S. McMorrow, *Insurance and behavioral economics: Improving decisions in the most misunderstood industry*, vol. 9780521845 (Cambridge University Press, 2013).

Layton, T. J., “Imperfect risk adjustment, risk preferences, and sorting in competitive health insurance markets”, Journal of Health Economics **56**, 259–280, URL `https://www.sciencedirect.com/science/article/abs/pii/S0167629617303211` (2017).

Levy, H., “Stochastic Dominance Decision Rules”, in “Stochastic Dominance”, pp. 41–124 (Springer International Publishing, Cham, 2016), 3 edn., URL `http://link.springer.com/10.1007/978-3-319-21708-6{_}3`.

Levy, H. and G. Hanoch, “Relative Effectiveness of Efficiency Criteria for Portfolio Selection”, The Journal of Financial and Quantitative Analysis **5**, 1, 63–76 (1970).

Levy, H. and M. Sarnat, “Alternative Efficiency Criteria: An Empirical Analysis”, The Journal of Finance **25**, 5, 1153–1158 (1970).

Liu, C. and J. R. Sydnor, “Dominated Options in Health-Insurance Plans”, NBER Working Paper Series p. 55, URL `http://www.nber.org/papers/w24392` (2018).

Machina, M. and J. Pratt, “Increasing Risk: Some Direct Constructions”, Journal of Risk and Uncertainty **14**, 2, 103–127, URL `http://link.springer.com/10.1023/A:1007719626543` (1997).

Machina, M. J., “Choice Under Uncertainty: Problems Solved and Unsolved”, Journal of Economic Perspectives **1**, 1, 121–154 (1987).

Madrian, B. C. and D. F. Shea, “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior”, The Quarterly Journal of Economics **116**, 4, 1149–1187, URL `https://academic.oup.com/qje/article-lookup/doi/10.1162/003355301753265543` (2001).

Mata, R., R. Frey, D. Richter, J. Schupp and R. Hertwig, “Risk Preference: A View from Psychology”, Journal of Economic Perspectives **32**, 2, 155–172, URL `https://pubs.aeaweb.org/doi/10.1257/jep.32.2.155` (2018).

Mazzocco, M., “Saving, Risk Sharing, and Preferences for Risk”, American Economic Review **94**, 4, 1169–1182, URL `http://pubs.aeaweb.org/doi/10.1257/0002828042002516` (2004).

Meyer, J., Meyer and Jack, “Two-moment Decision Models and Expected Utility Maximization”, American Economic Review **77**, 3, 421–30 (1987).

Mjelde, J. and M. Cochran, “Obtaining Lower and Upper Bounds on the Value of Seasonal Climate Forecasts as a Function of Risk Preferences”, Western Journal of Agricultural Economics **13**, 2, 285–293 (1988).

Ormiston, M. B. and E. E. Schlee, "Mean-Variance Preferences and Investor Behaviour", The Economic Journal **111**, 474, 849–861, URL https://academic.oup.com/ej/article/111/474/849-861/5059887 (2001).

Pauly, M. V. and Y. Zeng, "Adverse selection and the challenges to stand-alone prescription drug insurance.", Frontiers in health policy research **7**, 55–74, URL http://www.ncbi.nlm.nih.gov/pubmed/15612335 (2004).

Polyakova, M., "Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D", American Economic Journal: Applied Economics **8**, 3, 165–195, URL http://pubs.aeaweb.org/doi/10.1257/app.20150004 (2016).

Pope, R. D. and J.-P. Chavas, "Producer Surplus and Risk", The Quarterly Journal of Economics **100**, Supplement, 853–869, URL https://academic.oup.com/qje/article-lookup/doi/10.1093/qje/100.Supplement.853 (1985).

Porter, R. B., J. E. Gaumnitz, R. B. Porter and J. E. Gaumnitz, "Stochastic Dominance vs. Mean-Variance Portfolio Analysis: An Empirical Evaluation", American Economic Review **62**, 3, 438–46 (1972).

Pratt, J. W., "Risk Aversion in the Small and in the Large", Econometrica **32**, 1, 122–136, URL https://www.jstor.org/stable/1913738?origin=crossref (1964).

Richter, A., J. Ruß and S. Schelling, "Insurance customer behavior: Lessons from behavioral economics", Risk Management and Insurance Review **22**, 2, 183–205, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/rmir.12121 (2019).

Robst, J., J. M. Levy and M. J. Ingber, "Diagnosis-based risk adjustment for medicare prescription drug plan payments.", Health care financing review **28**, 4, 15–30, URL http://www.ncbi.nlm.nih.gov/pubmed/17722748http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4195001 (2007).

Rothschild, M. and J. Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", The Quarterly Journal of Economics **90**, 4, 629–649, URL https://academic.oup.com/qje/article-lookup/doi/10.2307/1885326 (1976).

Rothschild, M. and J. E. Stiglitz, "Increasing risk: I. A definition", Journal of Economic Theory **2**, 3, 225–243 (1970).

Samuelson, W. and R. Zeckhauser, "Status quo bias in decision making", Journal of Risk and Uncertainty **1**, 1, 7–59, URL http://link.springer.com/10.1007/BF00055564 (1988).

Sinaiko, A. D. and R. A. Hirth, "Consumers, health insurance and dominated choices", Journal of Health Economics (2011).

Trivedi, P. K. and D. M. Zimmer, "Copula Modeling: An Introduction for Practitioners", Foundations and Trends in Econometrics **1**, 1, 1–111, URL http://www.nowpublishers.com/article/Details/ECO-005 (2006).

von Gaudecker, H.-M., A. van Soest and E. Wengstrom, "Heterogeneity in Risky Choice Behavior in a Broad Population", American Economic Review **101**, 2, 664–694, URL http://pubs.aeaweb.org/doi/10.1257/aer.101.2.664 (2011).

Wilson, R., "Informational Economies of Scale", The Bell Journal of Economics **6**, 1, 184–195, URL https://www.jstor.org/stable/3003221?origin=crossref (1975).

Zilcha, I. and S. H. Chew, "Invariance of the efficient sets when the expected utility hypothesis is relaxed", Journal of Economic Behavior and Organization **13**, 1, 125–131 (1990).