NBER WORKING PAPER SERIES

# REGULATION OF INSURANCE WITH ADVERSE SELECTION AND SWITCHING COSTS: EVIDENCE FROM MEDICARE PART D.

Maria Polyakova

Working Paper 21541
http://www.nber.org/papers/w21541

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2015

This paper is the first chapter of my MIT dissertation. I am indebted to Amy Finkelstein, Nancy Rose, Jonathan Gruber, and Stephen Ryan for their guidance and support throughout this project. I thank Jason Abaluck, Isaiah Andrews, Andreas Breiter, Francesco Decarolis, Glenn Ellison, Sara Fisher Ellison, Kate Easterbrook, Michael Greenstone, Martin Hackmann, Nathan Hendren, Gaston Illanes, Ray Kluender, Brad Larsen, Anna Mikusheva, Manisha Padi, Maxim Pinkovskiy, Jim Poterba, Adam Sacarny, Brad Shapiro, Annalisa Scognamiglio, Stefanie Stantcheva, Michael Whinston, Joachim Winter and seminar participants at MIT, Georgetown, Yale, UIC, University of Wisconsin-Madison, Stanford, Toulouse School of Economics, Bocconi, Sciences Po, WUSTL, NBER, ARIA RTS, iHEA, IIOC, for very helpful comments, suggestions and discussions. I also thank Jean Roth at NBER and Sarah Brunsberg at ResDAC for their help in navigating the Medicare data, as well as the Max Planck MEA for their hospitality. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D.
Maria Polyakova
NBER Working Paper No. 21541
September 2015
JEL No. H0,H50,H51,I1,I13,L51,L78

**ABSTRACT**

I take advantage of regulatory and pricing dynamics in Medicare Part D to empirically explore interactions among adverse selection, switching costs, and regulation. I first document novel evidence of adverse selection and switching costs within Part D using detailed administrative data. I then estimate a contract choice and pricing model in order to quantify the importance of switching costs for risk-sorting, and for policies that may affect risk sorting. I first find that in Part D, switching costs help sustain an adversely-selected equilibrium and are likely to mute the ability of ACA policies to improve risk allocation across contracts, leading to higher premiums for some enrollees. I then estimate that, overall, decreasing the cost of active decision-making in the Part D environment could lead to a substantial gain in consumer surplus of on average $400-$600 per capita, which is around 20%-30% of average annual per capita drug spending.

Maria Polyakova
Department of Health Research & Policy
Stanford University
Redwood Building T111
150 Governor's Lane
Stanford, CA 94305
and NBER
maria.polyakova@stanford.edu

# 1 Introduction

Outsourcing a public health insurance benefit to private insurers creates a familiar trade-off between potential efficiency gains from competition and potential efficiency losses from adverse selection. We know much less, however, about the trade-offs from other market imperfections in competitive insurance settings. In this paper, I utilize the institutional environment of Medicare Part D prescription drug insurance program, to empirically analyze the role of switching frictions, focusing on how they may interact with the allocation of risks across contracts, and with the outcomes of policies that regulate the contract space.

Medicare Part D is a \$100 billion public insurance program introduced in 2006 that currently enrolls approximately 36 million beneficiaries.[1] This heavily subsidized benefit is administered entirely by private insurers that are extensively regulated. The insurers offer a variety of plans, making individual contract choice a prominent feature of the program.

Motivated by the dynamic evolution of the regulatory and pricing environment of Medicare Part D, I use the detailed administrative records on enrollment, health risk, and spending in the program, to explore how switching costs interact with adverse selection and with the regulatory interventions that change the relative generosity of plans. The paper argues that, conceptually, whether switching costs ameliorate or exacerbate adverse selection depends crucially on how both the relative prices *and* the relative coverage generosity of contracts evolve in comparison to the initial conditions. Quantitatively, I estimate a large switching friction, and that implementing a nudging policy in the current Part D environment is likely to lower the degree of adverse selection, and result in on average $\$400 - \$600$ per capita gain in consumer surplus from better matching of individuals to contracts and more efficient pricing. My results also highlight that switching costs may mute the ability of policy instruments, such as changes in minimim coverage requirements, to redistribute risks. I estimate that in the presence of switching frictions, the Affordable Care Act policy of increasing coverage by "filling the donut hole," will not equalize risks across contracts, foregoing a \$17 gain in average per capita consumer surplus.

[1] https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-04-30.html

The paper is structured as follows. The key pieces of Part D's institutional environment are summarized in Section 2. Section 2 also describes the sample of the nationwide administrative data from the Centers of Medicare and Medicaid. In Section 3, I turn to descriptive analysis.

The data suggest that there exists both cross-sectional and dynamic selection in Medicare Part D. Surprisingly, given the extent of policy attention to this concern, adverse selection in Medicare Part D has received little attention in the empirical literature. To fill this gap, I document several pieces of evidence for the presence of adverse selection. First, I present the results of the cross-sectional positive correlation tests in the spirit of Chiappori and Salanie (2000). These tests suggest a large degree of asymmetric information: the most generous plans attracted individuals whose annual drug spending was more than a standard deviation above the spending in the least generous plans. Although it is typically hard to disentangle selection from moral hazard in the positive correlation tests, Medicare Part D setting provides me with a way of addressing this problem. I use a rich set of moral hazard-free ex ante diagnostic information about the beneficiaries to recompute the positive correlation tests that still reveal similar patterns. Consistent with these cross-sectional results is also the evidence of dynamic adverse selection, which we see in the rapid unraveling of the most generous contracts.

I then turn to the question of whether choices in Medicare Part D exhibit switching frictions, which may be an important factor in understanding the over time dynamics of the risk-allocation patterns highlighted above. Medicare Part D's institutional setting and administrative records provide a fruitful setting for exploring inertia in contract choices. First, there is substantial variation in the choice set that is available in the program over time. Second, every year, new beneficiaries enter the program, as they become eligible for Medicare upon turning 65. Using these two sources of variation, I show that choices of the "continuing" cohorts persistently reflect the market conditions of the year when they first entered, while the choices of the otherwise similar newly entering cohorts reflect current market conditions.

With the model-free evidence of adverse selection and switching costs in hand, I proceed to a contract-valuation choice model of individual preferences for insurance plans in Section 4. I estimate a flexible choice model that admits observed and unobserved heterogeneous preferences, and allows me to quantify switching frictions, adverse selection, and the individuals' willingness to pay for different contract features conditional on their risk. The estimation results, discussed in Section 5.1, suggest that switching costs are large and critical for explaining enrollment paths over time. They further suggest that health risk plays an important role in determining individual choices, which is consistent with the model-free evidence of adverse selection.

Counterfactual analyses in Section 5.2 utilize the estimated choice model to explore the complex interactions among adverse selection, switching costs, and minimum standard policy regulations in the Medicare Part D setting. In the first counterfactual, I find that shutting down the inertia channel in Part D, decreases the difference in the average risk between the least and the most generous contracts by 26%. In the second counterfactual that simulates the expansion of the minimum required benefit in Part D as envisioned under the Affordable Care Act (ACA), I show that costly switching mutes the ability of policy instruments to redistribute risks. These results may appear unintuitive, as in a standard model of adverse selection where only premiums adjust over time, we would expect that active choices of rational consumers are likely to lead to more acture adverse selection. Indeed, Handel (2013) provides an empirical example of switching costs restricting a selection spiral in an employer-provided health insurance setting. In general, however, the direction of the effect of switching costs on selection is ambiguous, and depends on the development of relative premiums and relative coverage over time. Determining the direction of the interaction between selection and choice stickiness is especially complex in exchange-style health insurance settings. In such settings, a large number of competing plans are differentiated both vertically and horizontally, their prices and characteristics are affected by a host of regulatory interventions, and, due to complex subsidy and risk-adjustment provisions, as well as strategic pricing, the relative premiums of plans may not perfectly reflect the actuarial differentials among them. The combination of these factors implies that in such markets the relative prices of more comprehensive contracts may up or down, and their characteristics may adjust relative to the other contracts in a way that favors less acute sorting.

To assess the quantitative relevance of these mechanisms in Medicare Part D, I conduct several welfare analyses in Section 5.3. I find that independently of whether we believe that the estimated switching friction is a real cost or not, there is a substantial gain in consumer surplus to be realized from active decision-making and better matching. I estimate a gain in average per capita consumer surplus on the order of $\$400 - \$600$ or $20\% - 30\%$ of the average annual drug expenditure. This represents an upper bound of the possible surplus gain, as it assumes that we completely eliminated the stickiness in choices. These numbers are in stark contrast to my estimate of consumer surplus gain from the reduction in adverse selection, which I estimate to be at least an order of magnitude lower. The latter result largely reflects how differences in risks across more and less generous contracts are translated into consumer premium differences under the existing regulatory environment in Medicare Part D. Overall, my results suggest that, conditional on the observed environment, decreasing the cost of decision-making in a setting like Part D is likely to have a significantly net positive effect on consumer surplus.

The analysis in this paper is related to several literatures. First, the paper is related to the growing literature on Medicare Part D. It is closest to the analyses in Ho et al. (2015), who identify inattention as a key force for the observed choice stickiness, and analyze its implications for strategic supply-side pricing. A lot of the Part D literature has focused on assessing the rationality of individual decisions (Heiss et al. 2010, 2013; Abaluck and Gruber 2011, 2013; Kesternich et al. 2013; Ketcham et al. 2012; Kling et al. 2012; Ketcham et al. 2015). Heiss et al. (2009) note that adverse selection may be an important concern in Part D. Several papers have suggested that switching costs may be present in Part D. Ericson (2014) documents evidence of insurer pricing strategies consistent with the presence of inertia, while Abaluck and Gruber (2013) and Miller and Yeo (2015a) include consumer inertia in their models. Ketcham et al. (2012), on the other hand, suggest that inertia is not of key concern in Part D. Some of the other work analyzing Medicare Part D has looked at the impact of Part D on prescription drug consumption (Yin et al. 2008; Duggan and Scott Morton 2010); the role of subsidies(Decarolis, 2015; Miller, 2015; Decarolis et al., 2015); the welfare of reducing choice (Lucarelli et al., 2012) and of adding a public option (Miller and Yeo, 2015b); mergers (Chorniy et al., 2014); risk adjustment (Carey, 2015); and the moral hazard response to non-linearities of the contracts (Einav et al., 2015). Duggan et al. (2008) provide an extensive overview of the Part D's design.

Second, the paper is related to the growing empirical literature that analyzes asymmetric information, regulation, contract design, and welfare in both employer-provided and public health insurance settings. Einav et al. (2010) provide a systematic overview of the literature in this vein. The paper builds upon the insights in the work on the interaction between adverse selection and minimum standards in Finkelstein (2004), as well as on the interaction between adverse selection and inertia in an employer-provided insurance setting in Handel (2013).

Third, the paper also relates to the growing literature on inertia and defaults in a variety of public finance settings: among recent examples are Chetty et al. (2014), Beshears et al. (2014), Hastings et al. (2013) and Luco (2013), who document such patterns in retirement savings accounts, and Nosal (2012), who estimates switching costs in Medicare Advantage. Finally, methodologically and conceptually, the paper is related to a broad literature that assesses the impact of switching costs on market outcomes in a variety of settings (Farrell and Klemperer 2007 provides a survey), as well as the vast literature on regulation (Joskow and Rose 1989; Armstrong and Sappington 2007).

# 2 Institutional Setting and Data

### Basics of Medicare Part D

Medicare is a public health insurance program for the elderly and disabled in the U.S. Until 2006, standard Medicare insurance covered hospital and physician services, but not prescription drugs. In 2006, Medicare Part D prescription drug coverage was launched as part of the Medicare Modernization Act of 2003, becoming the largest expansion of Medicare since its introduction in 1965. While Medicare bears the greater share of Part D costs, the actual administration of the Rx benefit and part of the actuarial risk have been outsourced to private insurers. Part D insurers cover around 32 million beneficiaries (Hoadley et al., 2012). The program has cost the government around \$60-70 billion annually since 2006.[2]

Medicare Part D coverage is voluntary. Eligible individuals have to actively enroll in one

[2] See Medicare Baseline Projection reports by the Congressional Budget Office: https://www.cbo.gov/publication/44205.

of more than 30 stand-alone Rx plans offered in their state of residence during annual open enrollment period or when they first become eligible, e.g. turn 65.[3] Once enrolled, beneficiaries pay premiums on the order of \$400 – \$500 a year, and in return insurers pay for prescription drug purchases subject to a deductible, cost-sharing, and coverage limits. Beneficiaries stay in their chosen plan for a year and may change their choice during the open enrollment period next year. If beneficiaries make no active changes, CMS continues enrolling them into the same plan. The facts that individuals self-select into plans and have a "default" plan if they do not take any action after their first enrollment will be important for my analysis.

**Data – baseline sample**

I utilize administrative CMS data that comprises a 20% random sample of Medicare beneficiaries nationwide for years 2006-2009. The data provides demographic and health information about the beneficiaries, records the characteristics of all Part D plans available in each region of the country, as well as the enrollment choices of the beneficiaries, and their subsequent prescription drug claims. I make a number of restrictions to the original sample of 40.3 million beneficiary-year observations to isolate the part of the market where 65 year old and older enrollees self-sort into a cleanly observable set of stand-alone prescription drug contracts. These restrictions bring the sample down to 5.3 million individual-year observations.[4] I further construct a panel sub-sample of the baseline sample that contains individuals, whose choices and utilization can be observed from the first year they enter the program to 2009. The panel sub-sample has 3.7 million beneficiary-year observations on approximately 1 million unique individuals.

Table 1 provides the summary statistics of the observed demographic and risk related variables for the full sample, the baseline sample and the panel sub-sample. The individuals in

[3]Part D is split into 34 geographic markets. They trace out state boundaries, combining some less densly populated states into one market. All beneficiaries that live in the same Part D region face the same premiums and contracs features.

[4]More details on data construction are available in Appendix; Table A.1. The two biggest restrictions are: 1) excluding beneficiaries that enrolled in Medicare Advantage plans that combine inpatient, outpatient, and prescription drug coverage; 2) excluding individuals that were eligible for low-income subsidies from the government, as a lot of these individuals did not make their own contract choices, and faced different premiums and cost-sharing.

the baseline sample are on average 76 years old, 64% female, predominantly white (95%), with annual average drug spending of about $1,900. The panel sub-sample looks very similar, albeit a year younger on average and with slightly lower annual spending. In comparison to the full sample, the baseline sample has individuals that are older, more often white and more often female. Overall, the baseline sample has somewhat healthier individuals when compared to the full Medicare population.

**Regulatory environment and the contract space**

The Medicare Part D's contract space is strongly influenced by a minimum standard regulation. Medicare has designed a so-called Standard Defined Benefit (SDB) for the Part D program, and insurers are required to provide coverage with at least the same actuarial value. The SDB has an unusual design that features a relatively low deductible, flat co-insurance rate of 25% up to the initial coverage limit (ICL) and subsequent "donut hole", or coverage gap, that has a 100% co-insurance until the individual reaches the catastrophic coverage arm of the contract. Figure 1 illustrates what these cost-sharing features imply for individual out-of-pocket spending. Insurers are allowed to adjust and/or top up the SDB contract design. The resulting plans are highly multidimensional contracts that vary on a variety of characteristics, such as cost-sharing thresholds, co-pay and co-insurance levels, and coverage in the "donut" hole.

Despite the multi-dimensionality of contracts, there are three stylized facts about this market that allow me to simplify the description of the contract space. First, each insurer in practice offers the same me

nu of 2-3 contracts in all Part D regions in which it operates. Second, insurers tend to adjust only premiums, but not other contract features across different geographic regions. Third, insurers tend to keep the "types" of contracts in their menu fixed over time. Using these three stylized facts, I classify all contracts into four types. Contracts that offer the SDB level of deductible and no coverage in the gap are classified as *Type 1* contracts, while contracts that reduce the deducitble (typically all the way to zero) and offer no coverage in teh gap are classified as *Type 2* contracts. *Type 3* contracts offer a reduced deductible and partial coverage in the gap, while *Type 4* contracts offer a reduced deductible and full coverage in the gap. In addition to the cross-sectional complexity of the contract space, there is substantial over time variation in characteristics and premiums. The time-series changes are the outcome of both the market-driven adjustments of contract characteristics by insurers, as well as of annual changes in SDB regulation. Figure 1 demostrates that SDB regulation increased its generosity over time for higher spenders. For example, an individual who spent $3,000 on prescription drugs in an SDB plan, would have paid $1,500 out of pocket in 2006, but $1,196 in 2009.

# 3 Descriptive evidence

## 3.1 Adverse selection in Medicare Part D

The theory of adverse selection postulates that more comprehensive insurance contracts are likely to attract a pool of individuals with higher health risks. In practice, the literature does not always find evidence of adverse selection in insurance markets (Finkelstein and McGarry, 2006), which suggests that the presence of selection is ultimately an empirical question. In this section, I document novel empirical evidence for the presence of adverse selection in Medicare Part D. To begin with, I present the cross-sectional correlation test as described in Chiappori and Salanie (2000) using the ex-post realized drug expenditures. This exercise detects the presence of asymmetric information and follows the well-established testing literature reviewed in Einav et al. (2010). I then provide two pieces of additional evidence that help disentangle selection from moral hazard, which is a common concern in health insurance. First, I repeat the Chiappori and Salanie (2000) test using ex-ante information about individuals' health summarized in a risk score index. Second, I document patterns strongly consistent with the presence of a death spirals in the early years of the program. It is important to note that in this section I evaluate and define adverse selection as the inherent sorting of risks across contracts. I do not explicitly consider whether this inherent risk-sorting is ex-post successfully mitigated in insurers profits though a multitude of risk-adjustment and reinsurance policies that the federal government has put in place to combat selection in Part D.[5] We can think about the exercise as testing for whether there is in fact a reason for the government to be implementing these policies.

Panel A of Figure 2 illustrates the positive correlation test graphically. It plots the average realized drug spending in each region by contract type in 2006. We can see the stark differences in the expenditures in the more and less generous contracts; for example, *Type 4* contracts have realized average annual spending on the order of $\$4,000$ in all regions, as compared to $\$1,500$ in the *Type 1* contracts. Column (1) in Panel A of Table 2 presents the formal specification of this positive correlation test. Since insurers are allowed to price the same contracts differently in different regions, I control for region fixed effects. The regression specification takes the following form:

$$Y_{irt} = \alpha_r + \delta_t + \sum_{k=2}^{k=4} \beta_k \mathbf{1}\{ContractType_{it} = k\} + \epsilon_{irt} \tag{1}$$

where $i$ indexes individuals, $r$ indexes regions and $t$ indexes years. I use realized total drug spending as the outcome variable in Column (1). The results suggest that more generous contracts have higher spenders. For instance, contracts with full gap coverage have individuals with realized drug spending that is more than a standard deviation higher than in the plans with minimum standard coverage.

Since the correlation test that uses the realized spending as the outcome variable captures both adverse selection and moral hazard, I repeat the exercise using risk scores as the outcome variable in Column (2). Risk scores are constructed using ex-ante diagnostic information from Medicare's inpatient and outpatient claims, and therefore do not reflect any effects of Part D plan structure on spending.[6] Using the risk score measure gives qualitatively similar results, although the magnitude of differences is smaller. To give a more meaningful interpretation to

[5]There are three key policies that the government uses to ex-post adjust insurer profitability from bad risk pools. First, government subsidies to insurers are adjusted to account for enrollees' risk scores. Second, at very high spending level,s the government directly picks up 80% of the bill, leaving the insurer liability limited to 15% with the remaining 5% paid by the patient. Thirds, so-called risk corridors ensure that insurance companies do not suffer large overall losses at the end of one fiscal year of operations.

[6]CMS uses risk scores to adjust subsidy payments to insurers. This implies, if risk adjustment system were perfect, the implications of adverse selection documented here on premiums would have been mitigated by CMS payments across insurers. As noted above, in that case we could think about the exercise as testing for whether there is in fact a reason for the government to be implementing these policies.

the results with risk scores, in Column (3) I project drug spending onto risk scores and use "monetized" risk scores as the outcome variable in the correlation test. Column (3) thus gives us the *lower bound* on how much of the estimated differences in Column (1) can be attributed to adverse selection on ex ante observed (and thus potentially priceable) risk. Panel B of Figure 2 illustrates the moral hazard-free results graphically. The figure plots the empirical CDF of the ex-ante risk, as measured by monetized risk scores, in different types of contracts in year 2006. We observe that the empirical CDF of risks in more generous contracts is shifted towards having more mass of higher risks.[7]

In addition to the presence of cross-sectional adverse selection, over time development of the contract space in Medicare Part D is consistent with the presence of a rarely observed selection spiral. Despite the extensive efforts of the regulator to incentivize private insurers to offer Part D contracts with full gap coverage, *Type 4* contracts were not available in the market from 2008 onwards. In Panel B of Table 2, I document that the average premiums in *Type 4* contracts increased by 93% from 2006 to 2007, as compared to under 30% premium increase in other contract types. The average expenditure in these contracts increased from \$3,852 to \$4,377, and enrollment dropped from 5% to 1% of the market in the first two years. In 2008 onwards there were no *Type 4* contracts.

## 3.2 Switching frictions in Medicare Part D

In general, documenting inertia or switching frictions in observational data is challenging, since we need to distinguish between the "structural" state dependence and unobserved persistent individual heterogeneity (Heckman, 1981a,b; Heckman and Singer, 1986; Heckman, 1991; Honoré, 2002; Honoré and Tamer, 2006; Dube et al., 2010). Several features of the Medicare Part D environment, however, render themselves well to such analysis. First, due to the regulatory changes, entry and exit of plans, as well as substantial market-driven over time adjustment in prices and characteristics of contracts (some of it strategic, see Decarolis (2015)), there is pronounced non-stationarity in the observed contract space of the program. Second, the Part D

[7]The graphical results of Panel A and B for other years of the data are presented in the Appendix.

environment allows observing all individuals making their first choice in 2006, as well as younger individuals making their first choices in years 2007-2009 from the adjusted contract menus. In this section, I utilize these two features of the Part D environment to provide model-free evidence of choice behavior consistent with the presence of switching frictions.[8] This approach to documenting switching frictions is in the spirit of Handel (2013)'s approach to documenting inertia in employer-provided health insurance.

I find four descriptive patterns in the data consistent with the presence of significant switching frictions in Part D. First, as I document in Panel B of Figure 3, in each year of the program about 90% of individuals enroll in their "default" plans; for individuals whose default plans significantly change their financial characteristics (and thus their type) this probability is still around 80%. Since premiums and contract characteristics change substantially from year to year even if plans do not change their "type", the high persistence in choices suggests that switching costs may be present; this evidence alone, however, could just point to very persistent preferences.

As the second piece of evidence, I compare the choices of the newly entering and existing enrollees in different years. The results are recorded in Table 3. I focus on the individuals that can be tracked from their first entry continuously to 2009 and whose default plans' types were not changed by insurers throughout the observed years. This isolates individuals whose choices are not conflated with substantial supply-induced re-classification of plan types. Two patterns are pronounced in the data and consistent with the idea that switching frictions play an important role. First, enrollment shares over time for a given cohort tend to be closely related to the choices and market conditions of the first year in which the cohort entered the program. Second, the choices of different cohorts in the same year are very different, even if cohorts differ only by one year of birth. For example, comparing the choices in 2008 of the cohort that entered in 2006 and the cohort that entered in 2008, we see that the 2008 cohort is almost twice less likely to enroll in the least generous *Type 1* plan than the 2006 cohort in

[8] Section **??** then presents a more formal discussion of the issues related to the identification of the state dependence parameter in dynamic discrete-choice panel data models.

2008: 10% vs 19 % enrollment share. Another persistent difference in choices is visible for the 2007 cohort, which in 2007 was much more likely to select the *Type 3* plan with partial gap coverage than cohorts entering in other years.

Comparing the choices of different cohorts by their choice of different insurer brands rather than contract types paints a similar picture. Figure 3 records the enrollment shares of the two biggest insurers in the sample for each year 2006-2009. The enrollment shares are shown separately for the 65 year olds, who are entering the program anew, and older enrollees with incumbent plans. We see a striking difference in the 2009 choices. In this year, one of the insurers ("Insurer B") - which substantially increased its premiums in 2009 - lost almost all of its market share with the new enrollees. Only about 5% of the 65 year olds chose to enroll in the plans offered by this insurer. Among the continuing cohorts, its enrollment share also fell, but not nearly as dramatically. It remained higher than 20%, implying that in 2009 the existing cohorts were four times more likely to be enrolled with Insurer B than the new enrollees.

# 4 Empirical model

The descriptive evidence in Section 3 has documented that adverse selection and switching frictions are present in Medicare Part D. This evidence alone, however, doesn't allow quantifying the economic significance of these market imperfections, or the extent to which they may impact regulatory policies. To allow for such quantification, I formulate an econometric model of how individuals choose which contract to enroll in, and how insurers adjust contracts' premiums as a function of their risk pool. The model specifies the choice decision as a function of the individuals' information about their health risk, their incumbent insurance contract, and heterogeneity in individuals' preferences for different features of the contracts. Importantly, the model does not attempt to identify different sources of inertia, but rather essentially quantifies how much more likely beneficiaries are to select their incumbent plan conditional on the charactersitics of that plan, their choice set, and their idiosyncratic preferences. The choice model takes a contract-valuation approach rather than a realized utility approach. This implies that I project all contracts into a set of discrete characteristics and specify a stochastic indirect utility function over these charactersitics, rather than projecting contract characteristics into the mean and variance of spending and specifying a utility function over spending. Einav, Finkelstein, and Levin (2010) extensively discuss the advantages and diasadvnatages of both approaches.[9]

**Specification.** Each year $t$ an individual $i$ who lives in region $r$ and is enrolled in the Medicare Part D stand-alone prescription drug program chooses among $J_r$ plans offered by $B$ insurers. $J_r$ varies by region, as some plans are only offered in a subset of markets, and for most plans premiums vary across markets. The insurance contracts in each $J_r$ can be projected into the same set of observable characteristics. I let individual $i$'s utility from choosing plan $j$ (where "plan" is region-specific, so $r$-indexing is suppressed) in year $t$ be given by:

$$u_{ijt} = -\alpha p_{jt} + \beta_{it}\phi_{jt} + \gamma_{it}\mathbf{1}\{\text{Default}\}_{ijt} + \epsilon_{ijt} \tag{2}$$

$$\epsilon_{ijt} \sim \text{iid Type 1 EV}$$

Utility thus depends on the annual premiums charged by the plan in a given region in a given year $p_{jt}$, the characteristics of the plan $\phi_{jt}$, and whether $j$ was a default plan for individual $i$ in year $t$, where the default plan is usually the incumbent plan from $t-1$.[10] Individuals are assumed to choose a plan that gives them the highest utility. An important assumption that is implicit in this formulation of the utility function is that individuals are myopic in their choice of plans. Further, this formulation assumes that beneficiaries choose the option with the highest "perceived" utility, which may not necessarily correspond to the "best" plan for them from actuarial risk-protection perspective (Abaluck and Gruber (2011, 2013)). For the analysis of risk-allocation and choices in counterfactual scenarios, however, this "perceived" utility is exactly the object of interest.

[9]The contract valuation approach is common in the literature on health insurance; among recent applications, are Carlin and Town 2010; Lustig 2011; Bundorf et al. 2012; Starc 2014; Ericson and Starc 2015.

[10]In cases where plans were renewed or consolidated, CMS would default individuals in the same (if renewed) or designated new (if consolidated) plan if individuals took no action to change their choices; I record these cases as having "defaults".

I let the characteristics component $\phi_{jt}$ of the utility function include the following:

$$\begin{aligned} \beta_{it}\phi_{jt} &= \beta_{1it}\text{Deductible}_{jt} + \beta_{2it}\text{ICL}_{jt} + \beta_{3it}\mathbf{1}\{\text{Partial coverage in gap}\}_{jt} + \\ &+ \beta_{4it}\mathbf{1}\{\text{Tiered Cost Sharing}\}_{jt} + \beta_{5it}\mathbf{1}\{\text{LIS eligible plan}\}_{jt} + \beta_{6i}\mathbf{1}\{\text{Brand}\}_{j} \end{aligned} \tag{3}$$

This specification assumes that conditional on insurer, differences in plans can be accounted for by the deductible level, the initial coverage limit and a set of indicators on whether or not the plan offers partial coverage in the gap, whether the plan uses fixed dollar co-payments or co-insurance percentage, and whether the plan is eligible to enroll individuals with low-income subsidy. The included characteristics capture a substantial amount of variation among plans, since many features that are not explicitly included, such as the quality of customer service, pharmacy network quality, and the generosity of drug formularies tend to be insurer-level rather than plan-level characteristics.

Individual preferences for health insurance may reflect both individual health risk, as well as horizontal tastes and risk aversion that may or may not be correlated with expected drug spending. To capture these features of insurance demand in the model, I include rich observed and unobserved heterogeneity in the specification of marginal utility. First, I allow individual preferences for contract features to depend on the demographics and the individual's health risk. I use the full set of demographics observed in the data - age, gender, and race; as well as proxies for expected spending - risk scores and an additional flag for having end-stage renal disease diagnosis, which identifies especially high risks. Vector $D_{it}$ records this information: $D_{it} = \{\text{age}_{it}, \text{gender}_i, \text{race}_i, \text{risk score}_{it}, \text{esrd indicator}_{it}\}$. The preferences are allowed to depend on risk scores, but not on the realized spending. The advantage of this specification is twofold. First, it takes an ex-ante perspective and does not impose that beneficiaries have the exact knowledge of their ex-post realized spending. Instead, it assumes that beneficiaries have a general understanding that certain diagnostic groups on average cause higher drug spending. Second, it removes the concern about moral hazard in the choice model.[11] In addition to

[11] To verfiy that this choice of specification does not drive the results, I have estimated the model with ex-post spending as part of the mean of random coefficients, with and without risk scores as additional mean shifters.

the rich set of observed heterogeneity, I allow for unobserved heterogeneity in the model, by including random coefficients on the deductible, the initial coverage limit, and the indicator for partial coverage in the gap. The distribution of the random coefficients is specified to be normal.[12]The unobserved heterogeneity allows for any private information about health risk not captured in risk scores, for the heterogeneity in risk aversion, or any other individual-specific preferences (or "mistakes") for certain contract features. All in all, the coefficients on the contract characteristics and the lagged dependent variable are specified as follows:[13]

$$\beta_{it} = \pi^{\beta} D_{it} + \psi_i^{\beta}, \text{ where } \psi^{\beta} \sim N(\psi^{\beta}, \sigma^2), \text{ for} \beta_1, \beta_2, \beta_3 \tag{4}$$

$$\beta_{it} = \pi^{\beta} D_{it} + \psi^{\beta} \text{ for} \beta_4, \beta_5, \beta_6 \tag{5}$$

$$\gamma_{it} = \pi^{\gamma} D_{it} + \psi^{\gamma} \tag{6}$$

Assuming that an individual chooses the plan that maximizes his or her utility, the model lets us calculate the probability of a beneficiary choosing different plans in his or her choice set as a function of parameters. The maximum likelihood estimation approach is then used to find the values of the parameters that best rationalize the observed choices. Since in the data I can track the same individuals making several consecutive choices, the estimation utilizes this panel structure, explicitly modeling the probability of a sequence of choices. While assuming the extreme value distribution for the taste shocks produces a closed-form probability expression conditional on the realization of the random coefficients, the unconditional probability involves integrating out the normally-distributed random coefficients. The latter implies that there is no analytic closed-form solution for the probability integral that is part of the log-likelihood

The results are very similar to the baseline specification.

[12]Variations on this specification, such as specifying more flexible heterogeneity on brand fixed effects, or premiums, or imposing a log-normal distribution of random coefficients, do not significantly alter the main estimates, although adding more unobserved heterogeneity does tend to reduce the estimate of the switching friction. Hence, an important caveat to the current specification is that allowing for substantially more unobserved heterogeneity may further reduce the exact magnitude of the switching friction estimate. Appendix Table A.2 reports the results of the specification checks.

[13]Note that for $\beta_6$, the baseline model includes demographic interactions on 2 largest brands, the more flexible specification, reported in the Appendix Table A.2 includes full set of observed and unobserved heterogeneity in insurer brands.

function. Thus, the model is estimated using a simulated maximum likelihood (MSL) procedure as described in Train (2003, 2009) and Hole (2007).

**Identification.** The identification of the parameters relies on several unique features of the data. First, in the estimation of the switching friction parameter $\gamma$, we have to consider two issues: distinguishing between the "spurious" versus "structural" state-dependence and the initial conditions problem.[14] The inclusion of the unobserved individual heterogeneity through random coefficients into the model addresses the first issue in a way that is standard in the literature. The assumption is that the random coefficients capture individual-specific persistence in preferences, while the lagged dependent variable parameter captures the "true" state-dependence. Moreover, in 2007, 2008, and 2009, there were cohorts of 65 year olds that first became eligible for Medicare and entered the Part D program anew without switching costs. This implies that in years 2007-2009 of the data I observe individuals choosing with and without switching costs from the same menu of contracts. The latter feature greatly aids in separating the persistent individual heterogeneity from the switching friction. The second issue of the initial conditions problem does not arise in the current setting, as I observe the first choices for all individuals in the estimation sample, because the year of the program's launch is recorded in the data.

In addition, the descriptive evidence in Sections 2 and 3 suggests that there is substantial variation in the prices and the characteristics of plans in each year of the program. Such variation is important, since if the environment were very stable, we couldn't expect to observe any changes in choices either with or without costly switching.[15] The cross-sectional variation in the non-price features offered by different plans (such as zero deductible) is generated by the insurers' strategy of offering menus of several vertically-differentiated contracts. This strategy is

[14]The concern in the first issue is that the lagged dependent variable in the utility function, which is capturing the switching cost, will be correlated with an individual-specific preference parameter. To illustrate, in a generic binary non-linear dynamic panel model, this would imply that in $y_{it} = 1\{\beta x_{it} + y_{i,t-1}\gamma + \alpha_i + \epsilon_{it} > 0\}$, $y_{i,t-1}$ is a function of $\alpha_i$. Thus, if $\alpha_i$ is unaccounted for and left in the unobserved part of the utility function, the identification assumptions about $\epsilon_i$ will be violated. The literature on the non-linear dynamic panel data discusses the two broad approaches to this problem - assuming a parametric distribution for the unobserved individual heterogeneity ("random effects"), or trying to difference out the individual effects without functional form assumptions ("fixed effects"), where the latter approach encounters a lot of challenges given the non-linear nature of the model. Honoré (2002) and Honoré and Tamer (2006) provide excellent discussions.

[15]Dube et al. (2010) discuss the importance of observing variation in the choice set for the estimation of demand models with lagged dependent variables.

stable over time, which ameliorates that concern about contemporaneous responses to aggregate demand shocks. Further, a lot of time-series variation in contract features is generated by the changes in the minimum standard policy that annually adjusts the standard deductible and initial coverage limits. The variation in premiums stems from two sources. First, insurers set different relative prices for the two or three contract types in their contract menus. Second, insurers set different prices for the same contracts in different geographic regions. All these features of the data combined, aid to the estimation of the switching friction in this setting - a task that is typically challenging in observational choice data settings. The intuition of the identification argument is similar to that in Handel (2013) and Ho et al. (2015).

Naturally, the rich cross-sectional and time-series variation in the premiums of insurance contracts comes from observational data and not pricing experiments, suggesting that endogeneity concerns are warranted. The stochastic component of the utility function $\epsilon_{ij}$ may include unobserved characteristics of contracts that are correlated both with premiums and individual choices, leading to an omitted variables bias. For example, an insurer could be advertising a particular contract in its menu more than other contracts and setting the price of this contract higher/lower because of that. Note that conditioning on the insurer-specific fixed effects does not resolve this issue if insurers are advertising a particular contract in their menu. One example of such situation in the Medicare Part D setting is the endorsement of selected contracts by a well-known third party. Several contracts in the portfolio of one insurer were endorsed by the American Association of Retired Persons. It would be natural to assume that the AARP endorsement both leads individuals to select this contract with a higher probability and allows the insurer to raise prices either to exploit the less elastic demand, or to cover costs that may arise from the marketing relationship with the AARP. Having such unobserved characteristics in the stochastic portion of the random utility specification violates the assumption of no correlation between the observed and unobserved components of utility. The standard approach in the Medicare Part D literature (Abaluck and Gruber (2011); Heiss et al. (2013)) has been to assume that the rich observed characteristics capture all the relevant information about choices. In this paper I utilize an instrumental variables strategy based on the observations of contracts' expenditures in the administrative data to improve upon this approach.

To correct for potential endogeneity, we need an instrumental variable that affects contract premiums, but is not correlated with the contract characteristics that are not observed in the utility function. Since the costs of insurance contracts depend almost entirely on the prescription drug claims submitted by their enrollees, I use mean lagged realized claims as an instrument for current premiums. As Medicare regulation suggests, and the pricing model below confirms empirically, contract prices are strongly conditionally and unconditionally correlated with lagged mean realized claims in the contract. To utilize the lagged mean realized claims in each contract as an instrumental variable, we need to assume that the variation in this variable is independent of the unobserved contract characteristics conditional on the observed contract characteristics. For example, we need to assume that the AARP endorsement does not affect the level of realized risks in the contract. This assumption appears plausible, as long as we believe that the unobserved characteristics do not screen risks. The latter is indeed very likely, as the reduced-form analysis of Section 3.1 suggests that the largest risk-screening happens primarily on the gap coverage margin. To operationalize the instrumental variables estimation, I utilize the control function approach.[16]

**Empirical model of contract pricing.** To analyze how risk allocation may change in response to nudging policies or to increases in minimum standard requirements, I need to account for how insurers may adjust premiums of their contracts in response to the hypothesized changes in the environment. I utilize a hedonic-style approach to estimate each insurer's emprical policy

[16] Petrin and Train (2009). Formally, the premium for contract $j$ is a function of observed contract characteristics $\phi_j$, a variable that affects premiums, but doesn't otherwise affect the choice decisions $z_j$ , and the remaining unobserved term $\kappa_j$: $p_j = f(\phi_j, z_j, \kappa_j)$. The endogeneity concern arises if $\kappa_j$ is correlated with $\epsilon_{ij}$ in the utility function. Assuming linearity and additive separability of the unobserved component we have: $p_j = \lambda_\phi \phi_j + \lambda_z z_j + \kappa_j$, which is the first stage familiar from a linear IV model. As the choice model is not linear in price, however, the 2SLS technique cannot be applied. One alternative is the control function approach. The idea of this approach is to empirically estimate $\kappa_j$ and condition on it (or its function) explicitly in the utility function. In practice, $\kappa_j$ is calculated as the residuals of the first-stage regression of premiums on the observed contract characteristics included in the utility function and the lagged claims instrument. In the second step, a linear control function $CF = \omega \hat{\kappa_j}$ is included into the utility function:$u_{ijt} = -\alpha p_{jt} + \beta_{it}\phi_{jt} + \gamma_{it}\mathbf{1}\{\text{Default}\}_{ijt} + \omega\hat{\kappa_j} + \epsilon'_{ijt}$. I assume that the stochastic part of the utility function $\epsilon'_{ijt}$ has an iid extreme value component that is independent of $\kappa_j$ with the remaining component distributed jointly normal with $\kappa_j$. This assumption returns a mixed logit model with mixing over the selected characteristics of the contract as well as the error component (Villas-Boas and Winer, 1999).

function. This reduced form approach does not specify an explicit model of strategic pricing incentives, such as Bertrand. It does, however, allow for a tractable way of accounting for how changes in risk allocation may impact premiums in an environment that is likely characterized by a Markov-style equilibrium, as pricing decisions depend on the risk pool accumulated by insurers in the previous period.[17] Furthermore, the hedonic specification captures the regulatorily allowable determinants of prices. When insurance plans submit their annual prices to the Part D program, Medicare requires them to "justify" the economic validity of these prices. Participating insurers have to provide information about the spending experienced by the current enrollees in a given plan in the previous year and how the plan projects these spending will change in light of any planned changes in plan characteristics (usually those driven by changes in the minimum standard regulation).[18] I therefore include the moments of the lagged spending distribution and the key financial characteristics of the plans as the primary components of this pricing regression. Medicare allows plans to include administrative costs and desired profit margins for the plans, which I assume are insurer-specific and so can be picked up by insurer fixed effects. Importantly, this empirical policy function captures insurer's pricing decisions conditional on the overall observed environment, such as the degree of risk adjustment, and the bidding mechanism that CMS uses to convert prices submitted by insurers to premiums faced by consumers(Decarolis et al., 2015). The full specification takes the following form:

$$\begin{aligned} E[Y_{jbt}|.] &= \alpha_b + \delta_r + M'_{jbt-1}\beta + \\ &+ \gamma_1 Ded_{jbt} + \gamma_2 ICL_{jbt} + \gamma_3 1\{PartialGap\}_{jbt} \end{aligned} \tag{7}$$

where $j$ indexes plans (where "plan" is region-specific), $b$ indexes insurers (brands), $r$ indexes

[17]See Decarolis et al. (2015) for a Bertrand model of pricing incentives in Part D. The Bertrand model encorporates explicit strategic incentives, but in its standard form needs to make assumptions about the full effectiveness of risk adjustment in mitigating risk-related pricing incentives.

[18]For example, from CMS 2007 regulation ("Call Letter"): "In order to prepare plan bids, PDP Sponsors will use HPMS to define their plan structures and associated plan service areas and then download the PBP and Bid Pricing Tool (BPT) software. For each plan being offered, PDP Sponsors will use the PBP software to describe the detailed structure of their benefit packages and the BPT software to define their bid pricing information. Each formulary submitted by April 17, 2006, must accurately crosswalk to a plan (or set of plans) defined during the bid process. The combination of the PBP and BPT for a plan comprises a bid."

34 Part D regions, $t$ indexes years.[19] $Y_{jbt}$ is the annual premium charged to the beneficiaries by plan $j$ of insurer $b$ in year $t$. Vector $M$ contains several moments of the distribution of drug expenditures experienced by plan $j$ in year $t-1$, including the mean, the standard deviation, the inter-quartile range and the tail percentiles. As Table A.5 reports in the Appendix, the moments of the lagged risk distribution and the key characteristics of the plans together with the region and insurer fixed effects account for 80% of the variation in the data on premiums over years 2007-2009. As expected, plans with higher lagged mean annual spending, or plans that offer a lower deductible, higher ICL or partial coverage in the gap charge higher premiums.

# 5 Results

## 5.1 Parameter estimates

Table 4 records the detailed parameter estimates of the baseline specification.[20] The estimates suggest that beneficiaires dislike paying premiums and deductibles, and that the insurer's brand plays an important role in enrollment decisions. The switching "cost" is estimated to be substantial on average, on the order of $\$1,000$. I estimate that the switching cost is increasing at the rate of \$74 for an additional unit of risk score, and at the rate of about \$3 per year of age. Hence, the level of inertia varies in a statistically, but not economically significant way with observable demographic and risk chracteristics. For example, a 75 year old white female with no end stage renal disease and an average health risk is estimated to face a switching "cost" of $\$1,164$; while an 80 year old white male with no ESRD and twice the expected risk is estimated to face a switching cost of $\$1,253$.[21]

[19]To alleviate a concern about potentially radically different pricing techniques in the first year of the program, I only utilize data on 2007-2009 in the pricing regression.

[20]Table A.2 compares parameter estimates from several specifications of the choice model. Models that include some unobserved heterogeneity and control function all give similar results.

[21]The estimated order of magnitude is roughly similar to other findings in the health insurance literature. Handel (2013) estimates the switching costs to be about $\$2,000$ in the context of employer-provided health insurance; Nosal (2012) estimates the switching cost in Medicare Advantage health plans for seniors to be about $\$4,000$. Using a different choice model, Abaluck and Gruber (2013) estimate the switching costs in Medicare Part D to be on the order of $\$600-\$700$, which is lower than what my estimates suggest in the willingness-to-pay terms, although they similarly find that beneficiaries are roughly 500% more likely to choose a plan that is the

Consistent with the reduced-form tests for adverse selection in Section 3, I find that beneficiaries with higher health risk value the generosity of coverage more than individuals with lower risks. For instance, individuals with a risk score that is twice the Medicare average, are willing to pay \$22 more, on average, for each additional \$100 of the initial coverage limit, about \$120 more to enroll in plans that have fixed co-pays rather than co-insurance, and about \$230 more, on average, to be in a plan that offers partial gap coverage, than otherwise observationally identical beneficiaries with average risk. There is substantial unobserved heterogeneity in the valuation of gap coverage - with a standard deviation estimate of \$248 - suggesting that conditional on the observationally the same health risk profile, individuals may have strong preferences for or against coverage in the gap. This heterogeneity would be consistent with the presence of private information about health risk, or substnatial differences in risk aversion, or in other preferences for this plan feature.

Overall, the parameter estimates capture the intuition of the reduced form patterns in the data - beneficiaries are substantially more likely to enroll in their default contracts, and some contract features are highly valued by enrollees with higher health risk. Appendix Section 7.3 discusses the quantitative aspects of the fit in detail.

## 5.2 Counterfactual simulations

**Simulation of risk allocation with costly and costless switching** Cases A and B in Table 5 and Figure 4 illustrate the results of a counterfactual nudging policy simulation that completely shuts down inertia. Hence, the results serve as the upper bound of the effect of reducing inertia. This counterfactual exercise is similar in spirit to the analysis in Handel (2013) in the employer-provided insurance setting. The counterfactual is implemented as a policy shock in year 2009 of the program. I report several moments of the data, aggregating them to the three types of plans. Note that while in practice the complete elimination of inertia may be impossible, we can imagine an intervention that dropped all existing coverage

"default" plan for them in a given year. Using aggregate data and a dynamic demand model, Miller and Yeo (2015a) estimate the switching cost in Part D to be $\$1,700$, which is higher than my estimates.

and required individuals to choose their plans anew.

I perform two simulations of the model. The first simulation, marked with A, simulates the model *with* estimated switching costs. I use the pricing function from Section 4 to predict 2009 premiums conditional on the enrollment patterns and risk allocation observed in 2008 data. Other contract charactersitics, pre-2009 choices, and the demographic and health risk of the beneficiaries are taken as given from the data.[22] This step creates a baseline that takes into account the simulation error. I then use this baseline instead of the actual observed 2009 outcomes as a comparison benchmark in analyzing the scenarios without switching costs. Baseline enrollment shares, risk sorting, and prices paid, closely resemble the earlier descriptive evidence. *Type 2* plans with reduced deductible and no gap coverage have the highest enrollment share - 71% and about average (in-sample) risk profile of $\$1,915$ expected annual spending. *Type 3* plans with partial gap coverage are adversely selected with the average expected spending of $\$2,386$. The difference in the average risk between the *Type 1* and *Type 3* plans is substantial - expected spending in type 3 plans is 31% or \$563 higher. *Type 3* plans, with just 9% enrollment share, have annual premiums that are on average \$400 higher than in plans without partial gap coverage and no deductible. Note that the average premiums are unweighted and reflect the (simulated) pricing levels on the market rather than enrollment patterns.

The next simulation sets the switching cost parameter $\gamma$ in the utility function to zero for all individuals, which eliminates the inertia channel starting with year 2007. Hence, I re-simulate individuals' choices starting with 2007. I allow insurers to respond to changes in the risk sorting patterns induced by the removal of switching costs in each year. Insurers are assumed to adjust their premiums according to the stylized recursive pricing model outlined above. To calculate new prices I use the simulation of demand for individual contracts without switching costs to calculate the new allocation of risks, which is then feeded into the pricing equation.

[22]As part of specification checks, I tested the alternative approach of simulating the model from 2006 onward rather than taking the observed lagged choices as given in 2009. While the simulation error accumulates starker over several simulation periods in this case, this doesn't change the analysis in a substantive way. The baseline approach pursued in the main text renders itself better to the interpretation of the switching cost reduction as a sudden policy shock in one year. This point is irrelevant for the simulations without switching costs, since lagged choices do not enter the utility function.

Shutting down the inertia channel and adjusting insurers' premiums leads to lower relative premiums of the most generous *Type 3* contracts, and a smoother distribution of risks across contract types. The enrollment-unweighted premium for the most generous *Type 3* plans with partial gap coverage that now enroll slighly lower average expected risk decreases by 5%, so does the unweighted premium for *Type 1* plans. Weighted premiums decrease more, as without a switching friction individuals enroll into cheaper plans. Risk sorting among plans becomes less acute and all of them move closer to the average. The relative average risk between the *Type 1* minimum standard plans and *Type 3* plans with partial gap coverage falls from \$563 to \$416, which is a 26% decrease in the risk difference relative to the baseline simulation. Figure 4 illustrates this result graphically. It plots the counterfactual risk CDFs by the type of plan for the baseline (Panel A) and the simulation without the inertia channel (Panel B). The graph demonstrates that the decrease in the relative risk between most and least generous contract types holds throughout the whole distribution of risks and not only for the mean.

**Simulation of "filling the gap" policy**. The observation that switching frictions alter the sorting of risks across contracts suggest that they may impact the effect of policies that regulate the contract space, and which we would naturally expect to change the allocation of risks on the market. Consider Medicare Part D's minimum standard regulation. Under the Affordable Care Act (ACA), the government envisions "filling the donut hole", which basically implies increasing the level of the minium defined benefit that insurers have to provide. Since the inception of the program, the Centers for Medicare and Medicaid Services has been concerned with the availability of Part D coverage that goes beyond the minimum coverage requirements. CMS recognized that such coverage may be valuable for beneficiaries with costly health conditions, but also that the existing risk-adjustment and reinsurance policies may not be sufficient to overcome adverse selection into such coverage. CMS has experimented with different reinsurance policies that were supposed to encourage private provision of coverage in the gap; nevertheless, as we have seen, plans with full coverage were not offered on the market after 2008. The donut hole coverage has become a prominent issue in public debate and in the media, especially as public health research was argueing that the "donut" hole could be harmful for beneficiaries with chronic conditions (Stuart et al., 2005; Zhang et al., 2009; Gu et al., 2010) Ultimately, the goal of closing the gap in coverage via regulatory decree became a part of the new health care law. This provision of the ACA is currently in the early implementation stage. A notable feature of this policy is that although the new regulation is substnatially changing the contract landscape, current insurance plans were not discontinued, and beneficiaries were not required to make brand new choices.

To understand how inertia may impact the effect of this regulation, I do two counterfactual simulations in which I force all plans to have the level of 2009 catastrophic coverage threshold ($\$6,145$) as their initial coverage limit (originally $\$2,700$). Increasing the initial coverage limit to the catastrophic coverage attachment point effectively corresponds to the idea of "closing the gap." This change implies that *Type 2* and *Type 3* plans become identical in terms of their key financial characteristics, as both of them have reduced deductible and are imposed to have the same coverage limit. Even though the financial characteristics of these contracts are now very close, there is still some residual horizontal differentiation across the individual plans within each group, such as which insurer offers the plans and whether the insurer utilizes a co-insurance or co-pay cost-sharing structure.

The results of the simulations are reported in cases C and D in Table 5 and Figure 5. I start the counterfactual policy analysis by quantifying how increasing the minimum required benefit would impact risk sorting across contracts in the status quo with costly switching. I then consider how switching costs are altering this impact. The first simulation, reported in row C, adjusts contract features and premiums to the new level of coverage, but keeps inertia. We see that the relative risk in contracts previously marked as *Type 3* decreases by 39% relative to the baseline; however, there is still substantial difference in risks across the three contract types. The average expected risk in previously *Type 3* contracts is higher than in previously *Type 1* and *Type 2* contracts by \$342 and \$255, respectively. Ex ante, however, we would expect that, since there is no gap coverage dimension of differentiation among plans anymore, this policy should completely eliminate the selection that took place on the gap coverage margin. Indeed, the results in the second simulation, reported in D, suggest that exactly this effect takes place when I shut down the inertia channel. In this scenario, all types of plans converge to having practically identical pool of risks of around $\$1,950$ in expected spending, while in the scenario with costly switching, the simulated policy had a very muted effect on the distribution of risks. Figure 5 plots the counterfactual risk CDFs for both simulation, demonstrating that the result holds throughout the whole distribution of risks and not only for the differences in mean risk. In essence, the presence of switching frictions sustains "artificial" adverse selection, despite the regulatory convergence in the generosity of contracts.

### 5.3 Welfare analysis

To assess the efficiency implications of the counterfactual allocations, I calculate several measures of changes in consumer surplus. The main results are reported in Table 6. In most cases, I use the compensating variation metric of change in expected consumer surplus as derived in Small and Rosen (1981). In general, the change in expected consumer surplus for individual $i$ from a change in the characteristics of the choice set from $J$ to $J'$ is (conditional on the individual-specific parameters of the utility function $\theta_i$):

$$\triangle E[CS_i|\theta_i] = \frac{1}{\alpha}\left[ln(\sum_{j'\in J'} exp(v_{ij'})) - ln(\sum_{j\in J} exp(v_{ij}))\right]$$

where $v_{ij}$ is the part of the utility function in Equation 2 without the unobserved portion $\epsilon$:

$$v_{ij} = -\alpha p_j + \beta_i \phi_j + \gamma_i \mathbf{1}\{\text{Default}\}_{ij}$$

In all but two calculations, I will treat the switching friction as welfare-neutral, setting $\gamma$ to zero for the purpose of surplus calculation. Since we don't observe the individual-specific parameters of the utility function, I use a sparse grids method (Heiss and Winschel, 2008) to integrate out individual heterogeneity based on the observed distribution of demographics $D$ and the estimates of the distributions of random coefficients. The change in expected consumer surplus then becomes:

$$\triangle E[CS_i] = \int \frac{1}{\alpha} \left[ ln(\sum_{j' \in J'} exp(v'_{ij'})) - ln(\sum_{j \in J} exp(v_{ij})) \right] dF(\psi_i) dF(D_i) \tag{8}$$

In the first benchmark calculation, I treat the estimated switching friction as a real cost, and calculate how consumer surplus would change if we removed this cost and allowed insurance premiums to adjust. Hence, I am comparing consumer surplus between simulations A and B of Table 5 When we remove the switching cost individuals now face a different choice set. First, $J$' differs from $J$ because premiums have changed. In addition, $v'$ differs from $v$, as the former does not include the switching friction. The results reported in Column (1) of Table 6 suggest that consumer surplus increases substantially by \$621 or 32% of the average annual drug spending in the sample. This large increase is not surprising, since I have estimated a substantial switching friction, and this calculation treats the whole friction as a welfare-relevant cost. I estimate that there is significant heterogeneity in the efficiency cost of the switching friction - the gain in consumer surplus increases monotonically with risk. This is intuitive, as for higher risk individuals the differences across contracts are likely to be larger, and hence the presence of a default is more important.

In the next calculation I attempt to assess the efficiency implication of the switching friction in a scenario where we do not consider the friction itself to be a real cost. In this case, the consumer surplus formula 8 does not apply, since if we do not treat the switching friction as part of the utility function, then we are not changing the characteristics of the choice set (apart from the re-pricing). Hence, to evaluate the matching value of removing the switching cost, I turn to the simulations of individual choices in the demand model. While the switching friction does not affect welfare, it does affect choice probabilities. Simulating choices with and without the switching friction, I can calculate how much money-metric utility individuals are foregoing when they are lead to choose a sub-optimal contract because of inertia. We expect that the foregone money-metric utility has to be lower on average than the estimated switching friction, as otherwise beneficiaries would have chosen the optimal contract even in presence of inertia. Column (2) reports the results. I estimate the value better matching at \$455, or 23.4% of the annual drug spending. In this simulation I kept premiums the same, at the baseline simulated levels. Adjusting premiums to the re-allocation of risks (as in row B of Table 5) results in an additional average gain of $10, as recorded in Column (3). Notably, there is large heterogeneity in how much beneficiaries gain (or loose) when premiums are adjusted. The gains are higher for higher-risk individuals, but overall the estimated standard deviation of the change is $84, suggesting that not all beneficiaries benefit when insurers adjust premiums in response to the re-allocation of risks.

In Column (4) of Table 6 I come back to calculating changes in expected consumer surplus, and evaluate the effect of removing inertia in the second counterfactual of "closing the gap." This calculation is similar to the one in Column (1), in that I take treat the switching friction as a real cost. The result is quite similar; the effect of removing the switching cost is estimated at $576. Overall, I find that whether we treat the switching friction as a real cost or welfare-neutral, the is a substantial gain in consumer surplus - on the order of $20\% - 30\%$ of annual drug spending - to be realized from better matching of beneficiaries to insurance contracts.

To compare the effect of better matching to the efficiency impact of adverse selection, I conduct four more evaluations. Note that a full evaluation of the welfare cost of adverse selection would require a specification of insurer's marginal cost curves, subsidy mechanisms, and the evaluation of equilibrium allocations with and without selection. Specifying the equilibria with and without selection is beyond the scope of this paper; hence, I provide several suggestive estimates. First, focusing on the effect of selection on prices, I calculate the change in expected consumer surplus that would occur if insurers all enrolled a similar pool of risks and charged the same premiums for their contracts. I specify the premiums to be 25% of the average expected risk in the whole sample, which leads to premiums that are roughly similar to the observed premiums in *Type 2* contracts. This stylized approach attempts to illustrate what effect "perfect" risk-adjustment could have on consumer surplus. As reported in Column (5), I find the effect - $26 or 1.3% of average spending - to be very modest relative to the matching value of inertia. The magnitude, however, is in line with the estimates of the welfare cost of adverse selection in other contexts, e.g. Einav et al. (2010); Bundorf et al. (2012). I estimate a large and intuitive gradient along the health risk dimension - the gain in consumer surplus from equal premiums is larger the higher the risk. For the top 1% of the risk distribution the gain is 4 times larger than the average.

A typical concern in the literature that estimates the welfare loss from adverse selection is that the loss may be substantially underestimated if we just focus on premiums, since the key efficiency problem may be that of "missing contracts" (Einav, Finkelstein, and Levin, 2010). To evaluate this possibility in the current context, in Column (6) I calculate the loss in consumer surplus from removing *Type 4* contracts that were still observed in 2006. Perhaps surprisingly, removing these contracts in 2006 results in very little loss in consumer surplus on average - about \$1. This loss, however, is 20 times higher for the highest risk population, for whom these contracts provide valuable coverage. This implies that while adverse selection may have lead to *Type 4* contracts "missing" in the Part D market, only a small fraction of the population had a meaningful valuation for these contracts.

In the last two surplus calculations, I evaluate the efficiency effect of the re-allocation of risks that occurs in the counterfactual simulations after I remove the switching friction. In both counterfactuals of Section 5.2, removing the switching friction leads to less acute sorting of risks between the more and less generous contracts. In the recursive model of contract premiums of Section 4, this change in risk sorting would affect premiums in year $t+1$. To evaluate the effect of the change in risk distribution on surplus, I compare the surplus in $t+1$ between the scenarios with and without the switching friction. To quantify this effect in Counterfactual 1, I simulate 2010 premiums using the risk allocations from simulations A and B of Table 5. I assume that the contract space in 2010 is exactly identical to that of 2009, and only change the premiums based on different risk-sorting that had occurred in 2009 with and without inertia. The case where I remove inertia in 2009, which decreases the differences in risks by 26%, has a positive, but relatively small effect on consumer surplus in 2010. I estimate the average surplus change in the sample to be \$7 with a standard deviation of \$12. Forward simulating the second counterfactual of filling the gap, we get a bigger increase in surplus, as differences in risk allocations with and without inertia and larger. I estimate that the average surplus in

2010 goes up by \$17. Note that in this case we do not estimate a risk-related gradient, which is again intuitive, since all contracts are assumed to provide full coverage in the gap. Taken together, the calculations suggest that removing inertia is likely to lead to a substantial gain in consumer surplus from better matching, and to a positive, but at least an order of magnitude smaller, gain in surplus from reduced adverse selection.

# 6 Conclusion

In this paper I have documented evidence of adverse selection and switching frictions in a highly regulated Medicare Part D environment using a parsimonious classification of the contract space and detailed administrative data. I have also shown that the wide-spread inertia in choices may have complex interactions with the allocation of risks and with policies related to risk allocation in this highly regulated environment. In particular, I have shown that in this environment the initial conditions led switching costs to support an adversely selected equilibrium over time, in the sense that different types of plans would have had more similar average risks if switching were costless. I have also argued that in the presence of switching costs, tightening the minimum standard requirement by "filling" the donut hole is unlikely to have the intuitive effect of balancing risks across different contract. In particular, the counterfactual analyses demonstrate that the issue of locked-in risks due to stickiness of choices despite the evolution of the contract space may have an important effect on the outcomes of ACA reforms in Medicare Part D. While I find a substantial degree of adverse selection in Part D, the current policy environment appears to restrain the pass-through of plans' risk profiles to consumer premiums, so that the consumer-side efficiency loss from adverse selection *among observed contracts* is likely to be moderate. I estimate this loss to consumer surplus to be on the oder of 1% of average annual per capita drug spending. This is in stark contrast to the estimated gain in consumer surplus on the order of $20\% - 30\%$ of average annual spending, which could be achieved from better matching if switching or paying attention were costless. Overall, my results support the idea that, conditional on the existing policies that mitigate the effects of adverse selection in the background, decreasing the cost of decision-making in the environment like Part D is likely to have a significantly net positive effect on efficiency.

More broadly, this paper argues that in considering the polices that may improve consumer choice in the increasingly common public health insurance settings with regulated competition, we have to take into account the nuanced interconnections of the different market imperfections with the regulatory instruments targeted at correcting them. The caution, of course, comes from the caveat that in this work I represented the reaction of insurers to policy changes in a very stylized way. Expanding the model to allow insurers to endogenously react to adverse selection and regulation in their choices of contract characteristics in a competitive setting provides a fruitful area for future research.

# References

Abaluck, J. and J. Gruber (2011). Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program. *The American Economic Review 101*(June), 1180–1210.

Abaluck, J. and J. Gruber (2013). Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance. *NBER Working Paper no. 19163*.

Armstrong, M. and D. E. M. Sappington (2007). Recent Developments in the Theory of Regulation. In M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization, Vol. 3*, Number 06, pp. 1557–1700. Elsevier B.V.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2014). Who Uses the Roth 401(k), and How Do They Use It? *NBER Discoveries in the Economics of Aging University of Chicago Press*, 411–440.

Bundorf, K., J. Levin, and N. Mahoney (2012). Pricing and Welfare in Health Plan Choice. *American Economic Review 102(7)*, 1–38.

Carey, C. (2015). Technological Change and Risk Adjustment: Benefit Design Incentives in Medicare Part D. *Mimeo, Cornell University*.

Carlin, C. and R. Town (2010). Adverse selection: The dog that didn't bite. *Mimeo, University of Pennsylvania*.

Chetty, R., J. N. Friedman, S. Leth-Petersen, T. Nielsen, and T. Olsen (2014). Active vs. Passive Decisions and Crowdout in Retirement Savings Accounts: Evidence from Denmark. *Quarterly Journal of Economics 129*(3), 1141–1219.

Chiappori, P.-A. and B. Salanie (2000). Testing for Asymmetric Information in Insurance Markets. *The Journal of Political Economy 108*(1), 56–78.

Chorniy, A., D. P. Miller, and T. Tang (2014). Mergers in Medicare Part D: Decomposing Market Power, Cost Efficiencies, and Bargaining Power. *Mimeo, Princeton University and Clemson University*.

Decarolis, F. (2015). Medicare Part D: Are Insurers Gaming the Low Income Subsidy Design? *American Economic Review 105(4)*, 1547–80.

Decarolis, F., M. Polyakova, and S. Ryan (2015). Welfare Effects of Supply-Side Regulation in Medicare Part D. *NBER Working Paper No. 21298*.

Dube, J.-P., G. J. Hitsch, and P. E. Rossi (2010). State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics 41*, 417–445.

Duggan, M., P. Healy, and F. Scott Morton (2008). Providing Prescription Drug Coverage to the Elderly: America's Experiment with Medicare Part D. *Journal of Economic Perspectives 22(4)*, 69–92.

Duggan, M. and F. Scott Morton (2010). The Effect of Medicare Part D on Pharmaceutical Prices and Utilization. *The American Economic Review 100*(1), 590–607.

Einav, L., A. Finkelstein, and M. Cullen (2010). Estimating Welfare in Insurance Markets using Variation in Prices. *Quarterly Journal of Economics 125*(3), 877–921.

Einav, L., A. Finkelstein, and J. Levin (2010). Beyond Testing: Empirical Models of Insurance Markets. *Annual Review of Economics 2*, 311–336.

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

Ericson, K. (2014). Consumer Inertia and Firm Pricing in the Medicare Part D Prescription Drug Insurance Exchange. *American Economic Journal: Economic Policy 6*(1), 38–64.

Ericson, K. and A. Starc (2015). Pricing Regulation and Imperfect Competition on the Massachusetts Health Insurance Exchange. *Review of Economics and Statistics 97*(3), 667–682.

Farrell, J. and P. Klemperer (2007). Coordination and Lock-In: Competition with Switching Costs and Network Effects. In M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization, Vol. 3*, Number 06, pp. 1967–2072. Elsevier B.V.

Finkelstein, A. (2004). Minimum standards, insurance regulation and adverse selection: evidence from the Medigap market. *Journal of Public Economics 88*(12), 2515–2547.

Finkelstein, A. and K. McGarry (2006). Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market. *American Economic Review 96(4)*, 938–958.

Gu, Q., F. Zeng, B. Patel, and L. Tripoli (2010). Part D coverage gap and adherence to diabetes medications. *The American Journal of Managed Care 16(12)*, 911–918.

Handel, B. R. (2013). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *The American Economic Review 103*(7), 2643–2682.

Hastings, J., A. Hortacsu, and C. Syverson (2013). Advertising and competition in privatized Social Security: The case of Mexico. *NBER Working Paper No. 18881*.

Heckman, J. (1981a). *Structural Analysis of Discrete Data with Econometric Applications*, Chapter Statistical Models for Discrete Panel Data, pp. 114–78. Cambridge, Mass.: MIT Press.

Heckman, J. (1981b). *Structural Analysis of Discrete Data with Econometric Applications*, Chapter The Incidental Parameters Problem and the Problem of Initial Condition in Estimating a Discrete Time-Discrete Data Stochastic Process, pp. 179–95. Cambridge, Mass.: MIT Press.

Heckman, J. (1991). Identifying the hand of past: Distinguishing state dependence from heterogeneity. *American Economic Review 81(2)*, 75–79.

Heckman, J. and B. Singer (1986). *Handbook of Econometrics*, Chapter Econometric Analysis of Longitudinal Data, pp. 1689–763. Handbooks in Economics series, book 2 Amsterdam; Oxford and Tokyo: North-Holland;.

Heiss, F., A. Leive, D. McFadden, and J. Winter (2013). Plan Selection in Medicare Part D: Evidence from Administrative Data. *Journal of Health Economics 32*, 1325–1344.

Heiss, F., D. McFadden, and J. Winter (2009). Regulation of Private Health Insurance Markets: Lessons from Enrollment, Plan Type Choice, and Adverse Selection in Medicare Part D. *NBER Working Paper No. 15392*.

Heiss, F., D. Mcfadden, and J. Winter (2010). Mind the Gap ! Consumer Perceptions and Choices of Medicare Part D Prescription. In D. Wise (Ed.), *Research Findings in the Economics of Aging*, pp. 413–481. The University of Chicago Press.

Heiss, F. and V. Winschel (2008). Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics 144*(1), 62–80.

Ho, K., J. Hogan, and F. Scott Morton (2015). The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program. *NBER Working Paper No. 21028*.

Hoadley, J., J. Cubanski, E. Hargrave, L. Summer, and J. Huang (2012). Medicare Policy data spotlight Medicare Part D: A First Look at Part D Plan Offerings in 2013. *KFF Report* (October 2012).

Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal 7-3*, 388–401.

Honoré, B. (2002). Nonlinear models with panel data. *Portuguese Economic Journal 1(2)*, 163–179.

Honoré, B. and E. Tamer (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica 74(3)*, 611–629.

Joskow, P. and N. Rose (1989). The Effects of Economic Regulation. In R. Schmalensee and R. Willig (Eds.), *Handbook of Industrial Organization, Vol. II*, pp. 1449–1506. Elsevier Science Publishers B.V.

Kesternich, I., F. Heiss, D. McFadden, and J. Winter (2013). Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D. *Journal of Health Economics* (2013).

Ketcham, J., C. Lucarelli, and C. Powers (2015). Paying Attention or Paying Too Much in Medicare Part D. *American Economic Review 105(1)*, 204–233.

Ketcham, J. D., C. Lucarelli, E. J. Miravete, and M. C. Roebuck (2012). Sinking, Swimming, or Learning to Swim in Medicare Part D. *The American Economic Review 102*(6), 2639–2673.

Kling, J. R., S. Mullainathan, E. Shafir, L. C. Vermeulen, and M. V. Wrobel (2012). Comparison Friction: Experimental Evidence from Medicare Drug Plans. *The Quarterly Journal of Economics 127*(1), 199–235.

Lucarelli, C., J. Prince, and K. Simon (2012). The Welfare Impact of Reducing Choice in Medicare Part D: A Comparison of Two Regulation Strategies. *International Economic Review 53*(4), 1155–1177.

Luco, F. (2013). Switching Costs and Competition in Retirement Investment. *Mimeo, Northwestern University*.

Lustig, J. (2011). Measuring Welfare Losses from Adverse Selection and Imperfect Competition in Privatized Medicare. *Unpublished mimeo*.

Miller, D. P. (2015). Risk Adjustment and Low Income Subsidy Distortions in Medicare Part D. *Mimeo, Clemson University*.

Miller, D. P. and J. Yeo (2015a). Discrete Choice Models to Estimate Switching Costs: An Application to Medicare Part D. *Mimeo, Clemson University and Singapore Management University*.

Miller, D. P. and J. Yeo (2015b). The Consequences of a Public Health Insurance Option: Evidence From Medicare Part D. *Mimeo, Clemson University and Singapore Management University*.

Nosal, K. (2012). Estimating Switching Costs for Medicare Advantage Plans. *Mimeo, University of Mannheim*.

Petrin, A. and K. Train (2009). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research XLVI*.

Small, K. and H. S. Rosen (1981). Applied Welfare Economics with Discrete Choice Models. *Econometrica 49*(1), 105–130.

Starc, A. (2014). Insurer Pricing and Consumer Welfare: Evidence from Medigap. *Rand Journal of Economics 45*(1), 198–220.

Stuart, B., L. Simoni-Wastila, and D. Chauncey (2005). Assessing The Impact Of Coverage Gaps In The Medicare Part D Drug Benefit. *Health Affairs*.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

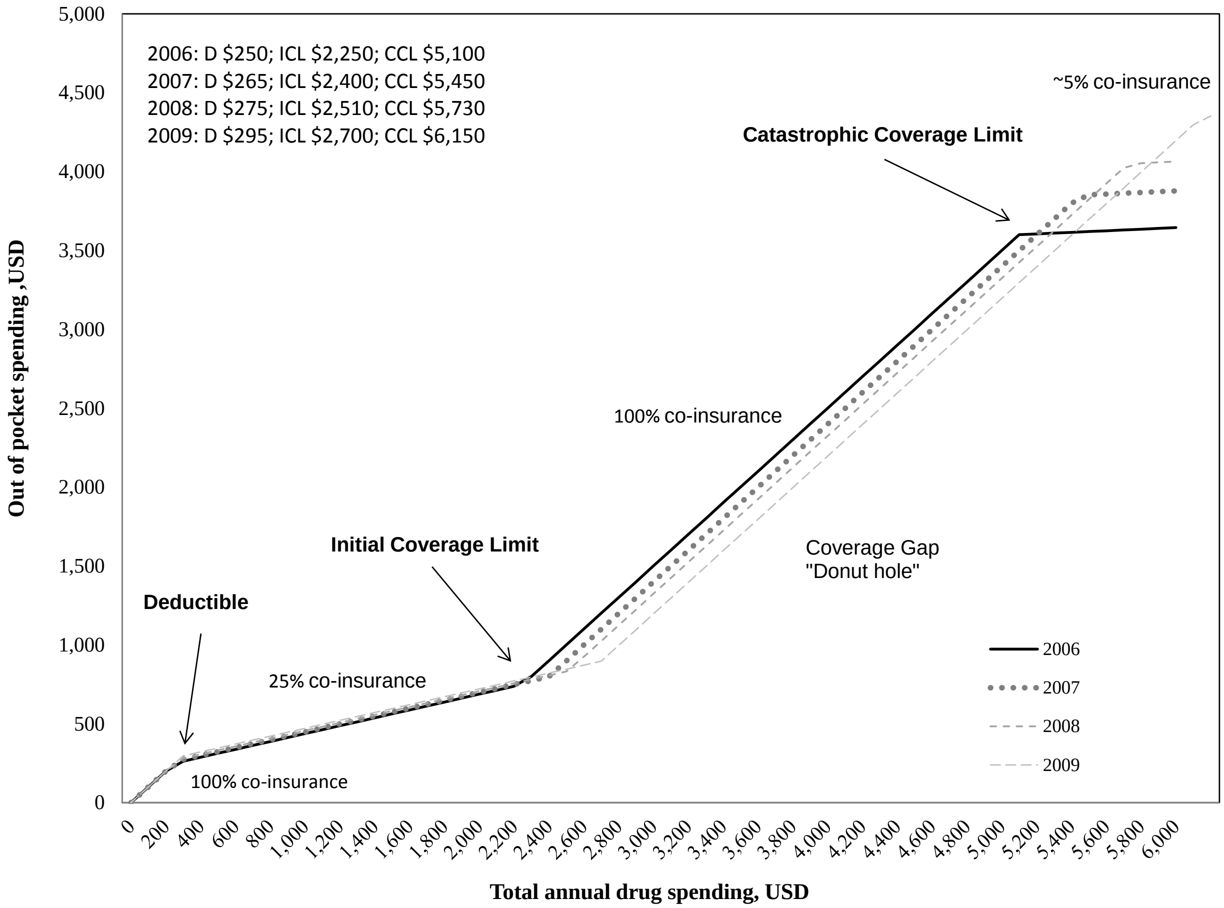
Train, K. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge University Press.

Villas-Boas, J. M. and R. Winer (1999). Endogeneity in brand choice models. *Management Science 45(10)*, 1324–1338.

Yin, W., A. Basu, J. X. Zhang, A. Rabbani, D. O. Meltzer, and G. C. Alexander (2008). The Effect of the Medicare Part D Prescription Benefit on Drug Utilization and Expenditures. *Annals of Internal Medicine 148*, 169–177.

Zhang, Y., J. Donohue, J. Newhouse, and J. Lave (2009). The Effects Of The Coverage Gap On Drug Spending: A Closer Look At Medicare Part D. *Health Affairs 28(2)*, 317–325.

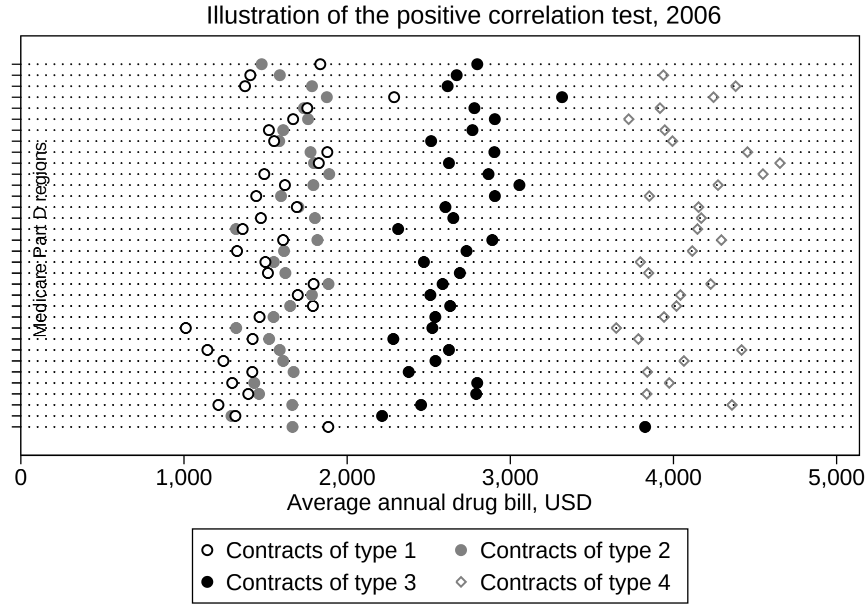
Figure 1: **Minimum benefit policy in Medicare Part D: shape of the Standard Defined Benefit in 2006-2009**



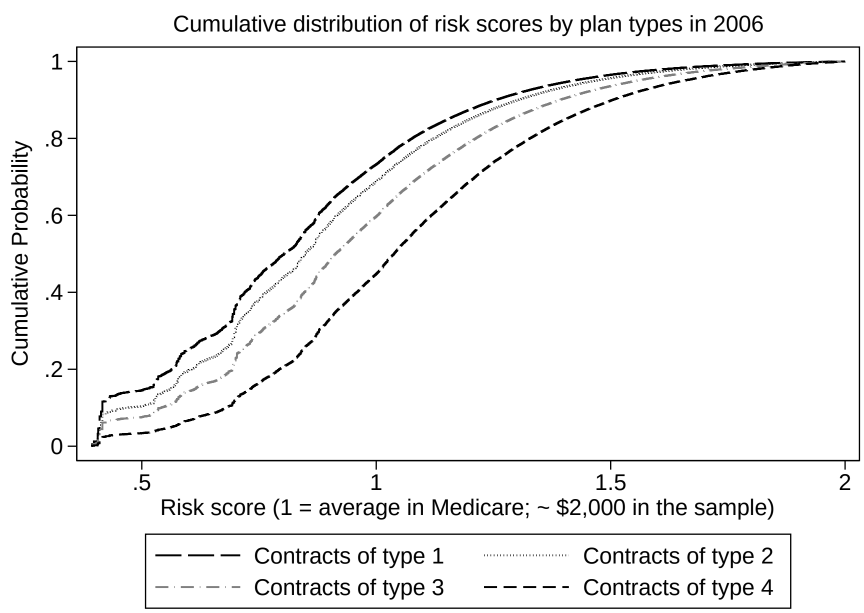

Insurers in the Medicare Part D program are required to provide coverage that gives at least the same actuarial value as the Standard Defined Benefit (SDB). The SDB design features a deductible, a co-insurance rate of 25% up to the initial coverage limit (ICL) and the subsequent “donut hole” that has a 100% co-insurance until the individual reaches the catastrophic coverage arm of the contract. The graph illustrates these features of the SDB by mapping the total annual drug spending into the out-of-pocket expenditure. Consider an individual, who in 2006 was in an SDB contract, and purchased prescription drugs for a total of $3,000. Out of this amount, the individual would pay the deductible of $250, then 25% of the next $2,000 up to the ICL of $2,500, and then 100% of the next $750 in the gap, for a total out of pocket spending of $1,500. As the figure illustrates, the generosity of the SDB changed over time. For example, an individual spending $3,000 on drugs in 2009 would face the out-of-pocket expenditure of less than $1,200.

Figure 2: **Model-free evidence of adverse selection**

*Panel A:* ***Positive correlation test***

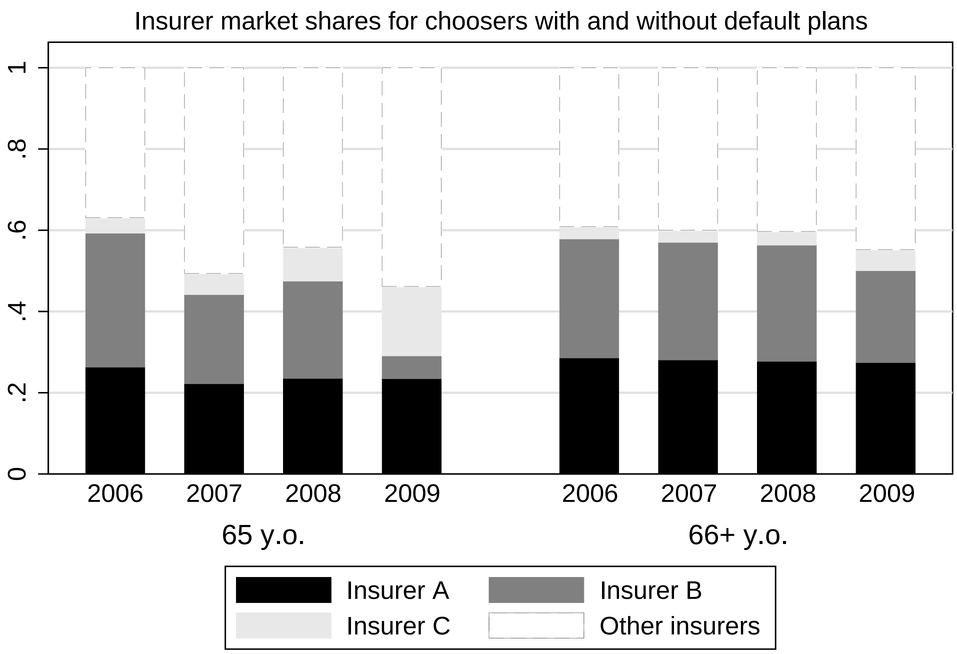



*Panel B:* ***Distribution of risks by type of plan using risk score measure***




*Panel A: A*verage annual drug expenditures in different types of Medicare Part D plans are calculated separately within each geographic region. Baseline sample. The spending of individuals with enrollment shorter than 12 months (primarily 65 year olds) was extrapolated to the full year. *Panel B:* Baseline sample. Risk scores are based on lagged diagnostic information and not on drug expenditures. The stochastic ranking of the distribution functions visualizes the riskier pool of enrollees recruited by plans with more generous coverage.

Figure 3: **Model-free evidence of switching frictions: insurer brands**




Based on the working sample, may not coincide with agggregate market reports

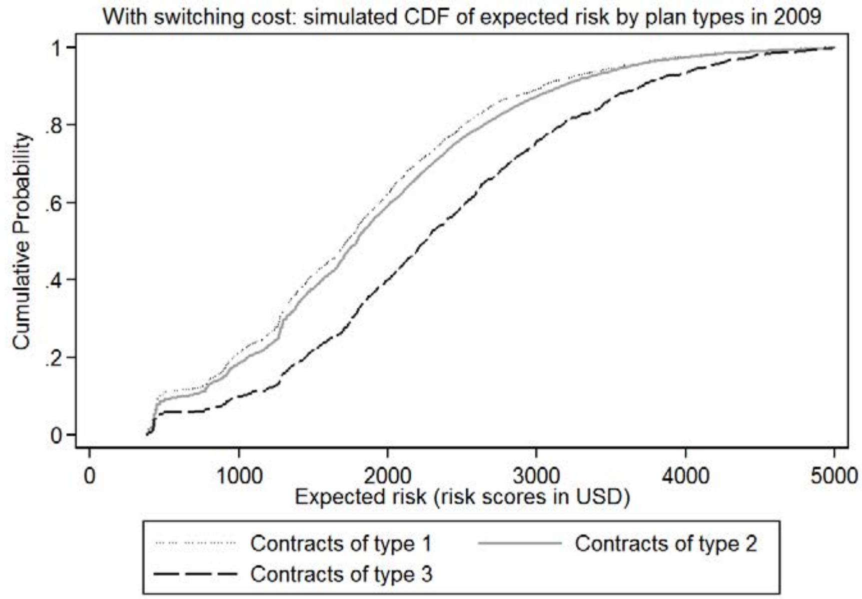
Baseline sample. We observe that the choices of the 65 year old individuals newly entering the program, who by definition do not have incumbent plans, are much more volatile and responsive to the market conditions over time than choices of the individuals in the older cohorts, who usually have the default option of their incumbent plan available. Insurer "identities" here are constructed using contract encryption in the administrative data. Because of the data encryption, separate insurers may have been identified with error. The corresponding commercial identities of the insurance companies are not known to the author.

Aggregate share of continuing enrollees choosing the same plan as in $t-1$ (includes all insurers and all plans):

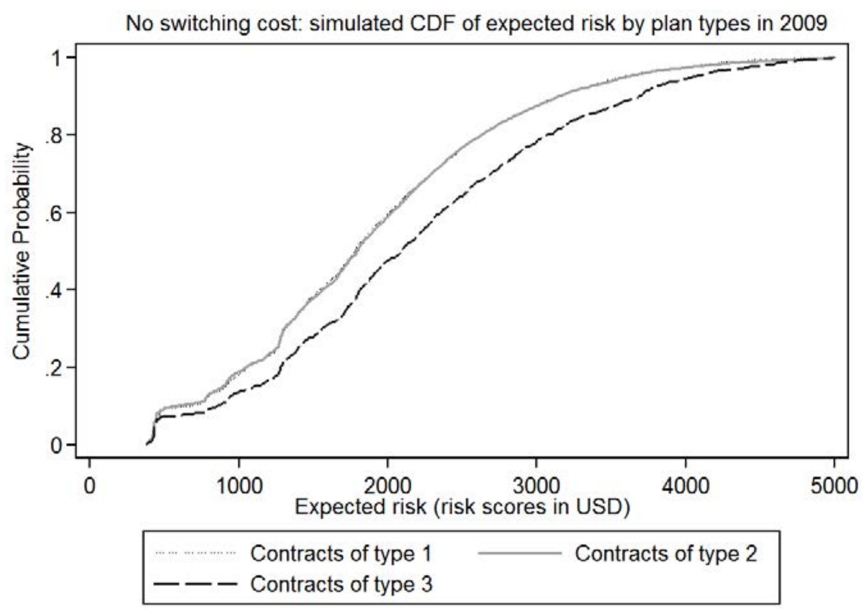
| | 2007 | 2008 | 2009 |
|---|---|---|---|
| Probability of choosing default plan for 66+ y.o. enrollees | 89.9 % | 88.7 % | 89.1 % |
| $N$ | 1,089,978 | 1,162,545 | 1,194,036 |

Figure 4: **Counterfactual risk allocation without switching frictions**

- *Panel A: Baseline simulated distribution of risks*



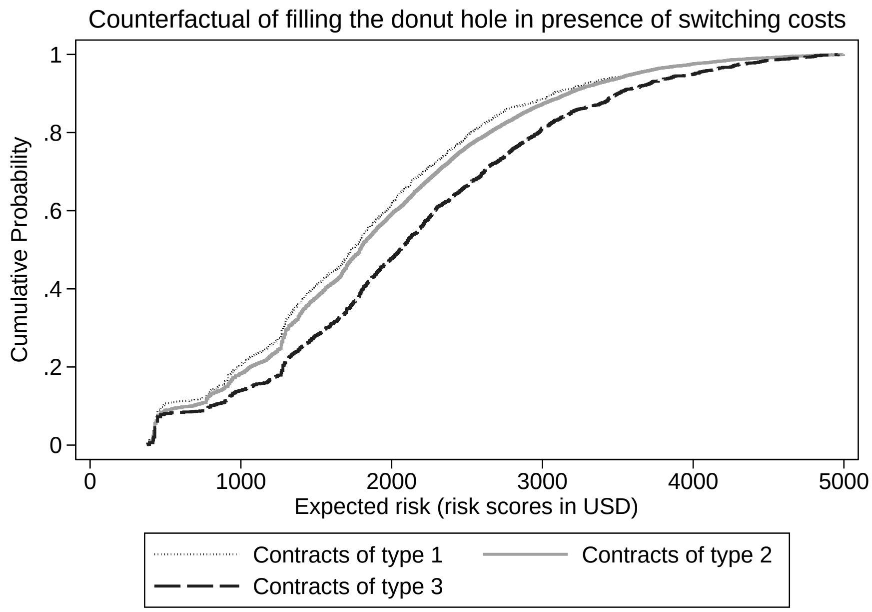

- *Panel B: Distribution of risks without switching costs (with endogenous re-pricing)*



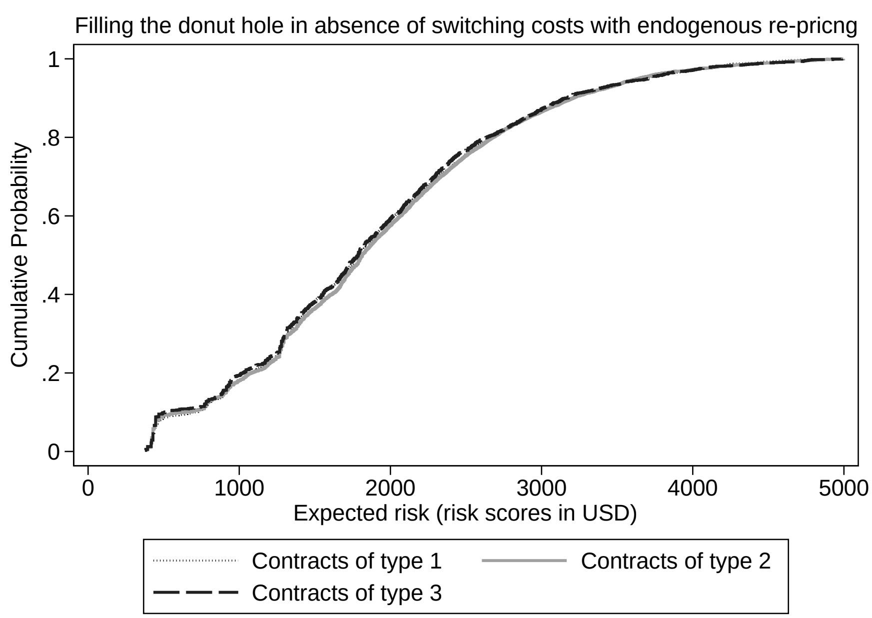

In-sample simulation of the choice model. *Panel A* keeps the estimated switching cost parameter $\gamma$, and predicts enrollment and pricing in the model. *Panel B* sets $\gamma = 0$, and re-simulates both enrollment and prices. Prices are re-simulated from year 2008 onwards, as this is the first year where the reallocation of risks induced by zero switching friction would affect recursively specified pricing function.

Figure 5: **Switching frictions and the ACA policy of "filling" the donut hole**

- *Panel A: Distribution of risks* ***with*** *switching friction*




- *Panel B: Distribution of risks* ***without*** *switching friction*




In-sample simulation of the choice model with and without switching frictions under the counterfactual policy of raising the minimum standard to completely fill the "donut" hole as envisioned under the Affordable Care Act. Both *Panels* are based on simulated prices that reflect the filling of the donut hole.

Table 1: **Summary statistics: full sample; baseline sample; panel sub-sample**

| | Full sample | Baseline sample | Panel sub-sample |
|---|---|---|---|
| **2006** | | | |
| $N$ | 9,999,538 | 1,221,252 | 871,818 |
| Age ($\sigma$) | 72 (12) | 76 (8) | 75 (7) |
| Female | 0.59 | 0.65 | 0.65 |
| White | 0.84 | 0.95 | 0.96 |
| ESRD | 0.01 | 0.003 | 0.001 |
| Risk score* | n/a | 0.89 (0.34) | 0.86 (0.31) |
| Annual drug spending ($\sigma$) | n/a | 1,518 (1,899) | 1,449 (1,733) |
| **2007** | | | |
| $N$ | 10,176,611 | 1,307,966 | 911,403 |
| Age ($\sigma$) | 72 (12) | 76 (8) | 75 (7) |
| Female | 0.59 | 0.63 | 0.65 |
| White | 0.84 | 0.95 | 0.96 |
| ESRD | 0.01 | 0.003 | 0.002 |
| Risk score ($\sigma$) | n/a | 0.90 (0.35) | 0.88 (0.32) |
| Annual drug spending ($\sigma$) | n/a | 1,883 (2,407) | 1,832 (2,227) |
| **2008** | | | |
| $N$ | 10,369,814 | 1,356,861 | 954,494 |
| Age ($\sigma$) | 72 (12) | 76 (8) | 76 (7) |
| Female | 0.58 | 0.63 | 0.65 |
| White | 0.83 | 0.95 | 0.96 |
| ESRD | 0.01 | 0.003 | 0.002 |
| Risk score ($\sigma$) | n/a | 0.91 (0.36) | 0.90 (0.34) |
| Annual drug spending ($\sigma$) | n/a | 1,907 (2,648) | 1,869 (2,479) |
| **2009** | | | |
| $N$ | 9,781,213 | 1,365,239 | 998,014 |
| Age ($\sigma$) | 71 (12) | 76 (8) | 76 (8) |
| Female | 0.55 | 0.63 | 0.65 |
| White | 0.83 | 0.95 | 0.96 |
| ESRD | 0.01 | 0.003 | 0.003 |
| Risk score ($\sigma$) | n/a | 0.92 (0.36) | 0.92 (0.35) |
| Annual drug spending ($\sigma$) | n/a | 1,950 (2,973) | 1,947 (2,934) |

* Risk scores are indices summarizing individual medical history from Medicare Parts A and B; 1 is average Medicare risk. The calculation of risk scores using Medicare A/B diagnostic records and Part D RxHCC software was generously provided by Amy Finkelstein and Ray Kluender.

Table 2: **Evidence of adverse selection**

- *Panel A: Positive correlation tests*

$$Y_{irt} = \alpha_r + \delta_t + \sum_{k=2}^{k=4} \beta_k \mathbf{1}\{ContractType_{it} = k\} + \epsilon_{irt}$$

| | (1) Annual drug spending | (2) Risk score | (3) Risk score projected to USD |
|---|---|---|---|
| Contracts of type 1 | reference category | | |
| Contracts of type 2 | -2.047 | 0.00927 | 24.98 |
| | (71.83) | (0.00730) | (20.89) |
| Contracts of type 3 | 1213.4*** | 0.146*** | 415.0*** |
| | (105.5) | (0.0107) | (30.46) |
| Contracts of type 4 | 3081.3*** | 0.260*** | 728.5*** |
| | (70.71) | (0.00521) | (14.57) |
| Year FE | Yes | Yes | Yes |
| Region FE | Yes | Yes | Yes |
| N | 3,892,280 | 3,892,280 | 3,892,280 |
| Mean Y | 1948.7 | 0.920 | 1948.7 |
| St. dev. Y | 2712.2 | 0.357 | 1018.7 |

Standard errors in parentheses clustered at the region level

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

- *Panel B: Dynamic changes in premiums and enrollment in Type 4 contracts*

| | Average Premiums, USD | | | Enrollment | | Expenditures, USD | |
|---|---|---|---|---|---|---|---|
| | **2006** | **2007** | **Change** | **2006** | **2007** | **2006** | **2007** |
| Type 1 | 171 | 216 | 27% | 26% | 22% | 1,428 | 1,489 |
| Type 2 | 264 | 331 | 25% | 63% | 63% | 1,596 | 1,724 |
| Type 3 | 472 | 610 | 29% | 6% | 14% | 2,483 | 2,802 |
| **Type 4** | **668** | **1,291** | **93%** | **5%** | **1%** | **3,852** | **4,377** |

*Panel A*. Baseline sample 2007-2009. Outcomes: (1) annual drug spending; (2) risk scores; (3) monetized risk score (constructed from $E[AnnualDrugSpending_i|.] = \alpha + \beta RiskScore_i$ run separately for each cross-section). *Panel B* records premiums, enrollment, and total drug expenditures in Plan Types 1-4, in years 2006-2007.

Table 3: **Model-free evidence of switching frictions: contract types**

| Cohorts of 65 year olds whose incumbent plans were not re-classified into a different type by the insurer | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 65 y.o. in 2006 | | | | 65 y.o. in 2007 | | | 65 y.o. in 2008 | | 65 y.o. in 2009 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| ***A.*** *Enrollment shares* | 2006 | 2007 | 2008 | 2009 | 2007 | 2008 | 2009 | 2008 | 2009 | 2009 |
| Contracts of type 1 | 22 % | 22 % | 19 % | 17 % | 17 % | 15 % | 14 % | 10 % | 11 % | 12 % |
| Contracts of type 2 | 73 % | 73 % | 77 % | 79 % | 72 % | 75 % | 77 % | 82 % | 82 % | 82 % |
| Contracts of type 3 | 4 % | 5 % | 5 % | 4 % | 11 % | 11 % | 10 % | 7 % | 7 % | 5 % |
| $N$ | 37,500 | 37,500 | 37,500 | 37,500 | 35,759 | 35,759 | 35,759 | 40,960 | 40,960 | 43, 520 |
| ***B.*** *Incremental premium* | **in year 2006** | | | | **in year 2007** | | | **in year 2008** | | **in year 2009** |
| Contracts of type 2 | $138 | | | | $125 | | | $54 | | $37 |
| Contracts of type 3 | $375 | | | | $360 | | | $410 | | $469 |

Panel A shows enrollment shares in each year across three types of plans for cohorts of 65 year olds entering the program in different years. The sample includes only individuals, whose incumbent plans were not re-classified in their type by the supply-side throughout the observed enrollment time. The choices of a given cohort are recorded for all subsequent years of available data. The calculation is based on the panel sub-sample data, as described in the data construction appendix. The table shows raw enrollment shares as observed in each year subject to the classification of contracts into the 4-type topology. The choices of cohorts show persistence over time and differ from the choices of newly entering cohorts within the same year. The difference is especially apparent between the first two and the last two years of the data. Panel B adds information about the development of the relative annual premiums over-time. The premiums are reported as increments relative to the *Type 1* contract with SDB deductible and no coverage in the gap. To reflect the market conditions, the premiums are constructed as averages weighted by enrollment of 65 year olds.

Table 4: **Parameter estimates of the baseline choice model**

| | | | **Interaction with demographics** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Plan characteristic** | Estimate | s.e. | Risk score | s.e. | White | s.e. | Female | s.e | Age | s.e | ESRD | s.e. |
| *in $100 or binary* | | | | | | | | | | | | |
| Annual premium | **-0.498** | 0.011 | - | - | - | - | - | - | - | - | - | - |
| Deductible - $\mu$ | **-1.238** | 0.127 | 0.038 | 0.037 | -0.023 | 0.057 | 0.013 | 0.025 | **0.008** | 0.002 | -0.039 | 0.258 |
| Deductible - $\sigma$ | **0.470** | 0.024 | - | - | - | - | - | - | - | - | - | - |
| Initial Coverage - $\mu$ | **-0.168** | 0.027 | **0.109** | 0.007 | **0.021** | 0.012 | 0.003 | 0.005 | 0.0003 | 0.0003 | **0.099** | 0.036 |
| Initial Coverage - $\sigma$ | **0.087** | 0.005 | - | - | - | - | - | - | - | - | - | - |
| Eligible for LIS, 1/0 | **-1.093** | 0.241 | **-0.125** | 0.072 | -0.089 | 0.106 | 0.017 | 0.047 | **0.026** | 0.003 | -0.683 | 0.563 |
| Fixed co-pays, 1/0 | **-1.538** | 0.394 | **0.589** | 0.119 | **-0.343** | 0.188 | 0.003 | 0.076 | **0.019** | 0.005 | **-1.521** | 0.789 |
| Partial gap, 1/0 - $\mu$ | **-2.153** | 0.334 | **1.126** | 0.089 | 0.105 | 0.150 | -0.059 | 0.065 | **0.016** | 0.004 | 0.055 | 0.540 |
| Partial gap, 1/0 - $\sigma$ | **1.238** | 0.054 | - | - | - | - | - | - | - | - | - | - |
| Default plan, 1/0 | **5.091** | 0.262 | **0.369** | 0.066 | **-0.815** | 0.125 | -0.067 | 0.046 | **0.015** | 0.003 | -0.394 | 0.381 |
| Log-likelihood | -59,224 | | | | | | | | | | | |
| N total observations | 2,435,171 | | | | | | | | | | | |
| N unique individuals | 12,769 | | | | | | | | | | | |
| N choice situations | 47,770 | | | | | | | | | | | |
| Max N alternatives | 66 | | | | | | | | | | | |
| N of regions | 34 | | | | | | | | | | | |

The table reports MSL estimates for the baseline specification of the model. To reduce the computational burden, the model was estimated on a sub-sample of individuals. This allows including panel observations for these individuals for all fours years of the data as well as to directly specifying each contract in the individuals' choice set in each year without any aggregation. To construct the estimation sample, I restricted the panel sub-sample to include individuals marked with CMS 5% sample flag (which amounts to taking a 25% draw of the panel sub-sample, since the original data represented a 20% population sample of Medicare). I then took a 5% random draw of that data in a way that preserves the original panel structure. The resulting dataset is not different from the original panel sub-sample in a statistically meaningful way. The table reports estimates of utility parameters and not marginal effects. Coefficients significant at 10% level are marked in bold font. Only key coefficients are reported; in particular, the model includes 10 fixed effects for insurer brands; the top 2 brands are interacted with obeservable demographics and risk scores.

Table 5: **Counterfactual simulations**

| **Year 2009 outcomes** | Type 1 contracts | Type 2 contracts | Type 3 contracts |
|---|---|---|---|
| **Average risk (monetized risk scores, USD)** | | | |
| A. Baseline prediction | \$1,823 | \$1,915 | \$2,386 |
| B. Remove inertia in 2007-2009; re-price | \$1,895 | \$1,904 | \$2,311 |
| C. Fill the gap; update prices, but keep inertia | \$1,838 | \$1,925 | \$2,180 |
| D. Fill the gap; update prices, remove inertia in 2009 | \$1,922 | \$1,951 | \$1,926 |
| **Average ex-post annual drug spending, USD** | | | |
| A. Baseline prediction | \$1,726 | \$1,876 | \$3,036 |
| B. Remove inertia in 2007-2009; re-price | \$1,915 | \$1,914 | \$2,330 |
| C. Fill the gap; update prices, but keep inertia | \$1,757 | \$1,873 | \$2,687 |
| D. Fill the gap; update prices, remove inertia in 2009 | \$1,947 | \$1,955 | \$1,961 |
| **Enrollment share** | | | |
| A. Baseline prediction | 20% | 71% | 9% |
| B. Remove inertia in 2007-2009; re-price | 26% | 64% | 9% |
| C. Fill the gap; update prices, but keep inertia | 19% | 68% | 13% |
| D. Fill the gap; update prices, remove inertia in 2009 | 25% | 61% | 13% |
| **Unweighted average annual premium** | | | |
| A. Baseline prediction | \$389 | \$513 | \$910 |
| B. Remove inertia in 2007-2009; re-price | \$368 | \$507 | \$866 |
| C. Fill the gap; update prices, keep inertia | \$1,466 | \$1,595 | \$1,855 |
| D. Fill the gap; update prices, remove inertia in 2009 | \$1,466 | \$1,595 | \$1,855 |

Summary of counterfactual simulations. (A) simulates the baseline, taking contract defaults in 2009 as given; it simulates premiums for 2009 using the data on observed 2008 enrollment and the pricing model. (B) shuts down the inertia channel in the utility function, and allows premiums to adjust to the new sorting of individuals when switching costs are not present. The premium adjustments are simulated from 2008 onwards - the pricing model is recursive and takes into account changes in 2007 risk-sorting. (C) and (D) implement the Affordable Care Act policy of filling the donut hole, i.e. set ICL to \$6,154 for all plans. Premiums are adjusted to reflect more generous coverage. In (C) I keep the switching friction parameter; in (D) I set it to zero. In (C) and (D) the distinction among contract types is based on the baseline simulation.

Table 6: **Welfare analysis**

| | **Quantifying the surplus effect of removing inertia** | | | |
|---|---|---|---|---|
| $\triangle E[CS]$, | Counterfactual 1 - remove inertia | | | Counterfactual 2 |
| except (2) and (3) | $\gamma = 0$ | Matching | Risk re-pricing | Fill the gap |
| (USD per capita) | (1) | (2) | (3) | (4) |
| Average | $621 | $455 | $10 | $576 |
| Standard deviation | $242 | $365 | $84 | $224 |
| % of mean spending | 31.9% | 23.4% | 0.5% | 29.6% |
| By risk percentile | | | | |
| <p5 | $623 | $472 | $12 | $588 |
| >p25 | $641 | $468 | $10 | $592 |
| >p50 | $660 | $484 | $12 | $607 |
| >p75 | $678 | $500 | $13 | $619 |
| >p95 | $711 | $548 | $15 | $648 |
| | **Quantifying the surplus effect of adverse selection** | | | |
| | Equal premiums | Remove Type 4 | Counterfactual 1 | Counterfactual 2 |
| $\triangle E[CS]$ | in 2009 | plan in 2006 | simulate forward | simulate forward |
| (in USD per capita) | (5) | (6) | (7) | (8) |
| Average | $26 | -$1 | $7 | $17 |
| Standard deviation | $40 | $4 | $12 | $15 |
| % of mean spending | 1.3% | -0.1% | 0.4% | 0.9% |
| By risk percentile | | | | |
| <p5 | $3 | $0 | $6 | $19 |
| >p25 | $31 | -$1 | $7 | $16 |
| >p50 | $39 | -$2 | $8 | $16 |
| >p75 | $52 | -$3 | $9 | $16 |
| >p95 | $82 | -$7 | $11 | $15 |
| >p99 | $115 | -$20 | $15 | $15 |

(1)-(3) report results for Counterfactual 1 that removes the switching friction. (1) Reports the average$\triangle E[CS]$ (Small and Rosen, 1981) when we treat the switching friction as a real cost. This change in surplus depends on the choice set of the consumer and not on the observed or simulated choices. (2) Reports the difference in WTP for the insurance plan that is selected with and without a switching friction. The WTP in both cases treats the switching friction itself as welfare-neutral. (3) Incremental change to (2) when premiums are adjusted. (4) Same as (1), but first fill the gap and adjust premiums; (5)-(8) do not include inertia. (5) Sets all premiums equal to 25% of the average risk. (6) Removes *Type 4* plans in year 2006 from the choice set; (7) and (8) forward simulate (1) and (4) to assess the effect of change in risk distribution.

# FOR ONLINE PUBLICATION APPENDIX

# 7 Appendix

## 7.1 Conceptual framework: interaction between adverse selection and switching cost in the presence of minimum standard regulation

A stylized model of insurance contract choice below highlights the key economic channels that are analyzed empirically in the paper. Consider a mass of beneficiaries, each described by a pair of characteristics - the individual's risk type $r$, as well as risk preferences and other demographic or idiosyncratic factors that may affect the individual's preference for insurance together denoted with $\phi$. For simplicity, assume that the individual faces a choice between two insurance contracts that differ only in their deductible. The more generous contract $H$ has a zero deductible and a premium $p_H$, while the less generous contract $L$ has a deductible $d > 0$ and a premium $p_L < p_H$.

Assuming the separability of prices in the indirect utility function and letting $v(d, \phi, r)$ denote the valuation of a contract with deductible $d$ by individual $(\phi, r)$, we arrive at a standard choice problem in a differentiated goods environment. Individual $(\phi, r)$ chooses contract $L$ if:

$$v(0, \phi, r) - v(d, \phi, r) < p_H - p_L$$

$$\triangle v(d, \phi, r) < p$$

where $p$ denotes the relative price. Suppose that for any given level of the deductible, the valuation of an insurance contract is increasing in risk $r$, i.e. $\frac{\partial v(d,\phi,r)}{\partial r} > 0$ and preferences such as risk aversion, i.e. $\frac{\partial v(d,\phi,r)}{\partial \phi} > 0$, while the valuation is decreasing in the deductible for a given $(\phi, r)$, i.e. $\frac{\partial v(d,\phi,r)}{\partial d} < 0$. Suppose further that the valuation and prices are such that the "market is covered" in the sense that all individuals find it optimal to buy one of the insurance contracts rather than to remain uninsured.[23] Then, there exists an individual of type $(\hat{\phi},\hat{r})$ who is indifferent between the two contracts, i.e. $\triangle v(d, \hat{\phi},\hat{r}) = p$. The average risk that contract $L$ expects to get after individuals choose between the two contracts is $E[r|\triangle v(d, \phi, r) < \triangle v(d, \hat{\phi},\hat{r})]$.

Now suppose we introduce an exogenous shock to the model that changes the features of the contract space. Consider, for instance, a one-dimensional minimum standard policy that only sets the maximum allowed deductible $\bar{d}$. Assume further that the less generous contract sets its deductible $d$ to always equal the maximum deductible set by the government: $d = \bar{d}$. The more generous contract, at the same time, always keeps zero deductible. This simplification implies that I am not modeling how insurers originally decide whether to offer the minimum standard or zero deductible, taking these decisions as given and stable from the policy perspective.

Now suppose the government changes its policy and increases the maximum allowed de-

[23] While this assumption is certainly restrictive and eliminates an important extensive margin on which the minimum standard may affect the market (Finkelstein, 2004), the empirical model in this paper focuses on the effects of the minimum standard on the intensive margin, across different levels of contract generosity, and thus I focus on this aspect of the question in this stylized model as well.

ductible from $d$ to $d' > d > 0$ and nothing else changes. In particular, suppose for a moment that relative prices remain the same $p$. Individuals that were choosing contract $L$ before, will switch to contract $H$ under the new policy if now:

$$\triangle v(d', \phi, r) > p$$

The risk pool of switchers from the less to the more generous contract under the new policy but without price adjustment is: $E[r|\triangle v(d, \phi, r) < p \text{ and } \triangle v(d', \phi, r) > p]$. Whether this re-sorting results in higher or lower risk in contract $L$ depends on whether the effect of risk on valuation grows faster at a higher deductible than the effect of non-risk preferences on valuation under a higher deductible. In other words, it depends on the relationship between $\frac{\partial^2 v(.)}{\partial r \partial d}$ and $\frac{\partial^2 v(.)}{\partial \phi \partial d}$.

Now suppose that individuals face a switching cost $\gamma$. This cost may be heterogeneous across individuals and correlate both with individual preferences $\phi$ and risk type $r$. Let $\gamma$ be a function of individual characteristics $\gamma(\phi, r)$. With the switching friction individuals that were choosing contract $L$ before the policy change, will switch to contract $H$ under the new policy if:

$$\triangle v(d', \phi, r) > p + \gamma(\phi, r)$$

The switching cost has the effect of diminishing and tilting the set of beneficiaries that are indifferent between switching to $H$ and staying in $L$. The first order effect is that the presence of the switching friction slows down the re-sorting process, as now fewer consumers react to the change in the contract space. The second-order tilting effect is that whether relatively higher or lower risks tend to stay in contract $L$ rather than change to $H$ in the presence of switching cost will depend on the partial and cross-partial derivatives of the switching cost with respect to risk $r$ and preferences $\phi$.

Allowing insurers to adjust prices to the new regulation and sorting patterns that are distorted by the switching costs produces theoretically ambiguous results that depend on the relationship between contract valuation and risk. For example, with a higher regulated deductible, the relative price will increase because a higher deductible mechanically reduces the liability of contract $L$. This, in turn tightens the switching constraint $\triangle v(d', \phi, r) > p' + \gamma(\phi, r) > p + \gamma(\phi, r)$, which can further decrease or increase the risk depending on the individual value function. Overall, the direction of change in sorting patterns induced by the change in the contract space are ambiguous if we allow for switching costs and allow insurers to adjust prices in response to changes in selection patterns. The effect that the regulation has on the allocation of risks across contracts will depend on the partial and cross-partial derivatives of the valuation and switching costs with respect to risk and preferences. The choice model in Section 4 estimates these inter-dependencies in Medicare Part D empirically and uses the estimates to simulate the role of switching costs in shaping the risk-sorting across contracts in response to market-driven and regulatory changes in contracts.

## 7.2 Construction of the empirical sample from Medicare's administrative data

I restrict the sample to individuals of age 65 and older residing within 34 Medicare Part D regions or 50 states (Medicare combines some states into the same PDP market), who did not die in the reference year and were originally entitled to Medicare because of old age rather than disability. In other words, I do not include individuals, who may become eligible for Medicare before they turn 65 as part of their SSDI benefit. I further drop observations on individuals that were dual eligible for Medicare and Medicaid in the reference year, since these individuals are assigned to plans by CMS rather than choosing plans on their own. This brings the sample down to 25.6 million beneficiary-year observations. I then eliminate individuals that did not enroll in Part D or were enrolled in Medicare Advantage (or another managed care) option that combines prescription drug coverage with healthcare insurance.

Most differences between the panel sub-sample and the baseline comes from the way CMS draws its 20% random sample of the Medicare population. These samples are only partially based on panel draws and thus not all individuals are observed in every year. For details on the CMS sampling procedures see the Chronic Condition Data Warehouse User Manual v.1.7. Some individuals in the panel sub-sample will be lost if they change from a PDP to a Medicare Advantage prescription drug plan simultaneously with switching from the "traditional" Medicare to the HMO system. Moreover, it is possible that some individuals leave the Part D program altogether; this option is likely to be very rare, however, since these beneficiaries would then face premium penalties if they decide to re-enter the program at a later date. Lastly, some observations will be lost in the panel sub-sample due to individuals dying in years 2007-2009.

Table A.1: **Construction of the baseline sample**

| | 2006 | | 2007 | | 2008 | | 2009 | |
|---|---|---|---|---|---|---|---|---|
| Full sample, $N$ | 9,999,538 | *100%* | 10,176,611 | *100%* | 10,369,814 | *100%* | 9,781,213 | *100%* |
| Keep age 65+ within 50 states | 8,385,276 | *84%* | 8,511,573 | *84%* | 8,658,693 | *83%* | 8,066,696 | *82%* |
| Drop if died in the reference year | 7,982,664 | *80%* | 8,111,023 | *80%* | 8,249,112 | *80%* | 7,714,002 | *79%* |
| Drop if dual eligible any month of year | 6,839,959 | *68%* | 6,952,339 | *68%* | 7,087,638 | *68%* | 6,637,418 | *68%* |
| Keep if Medicare b/c of old age | 6,412,259 | *64%* | 6,505,996 | *64%* | 6,619,029 | *64%* | 6,178,410 | *63%* |
| Keep PDP enrollees [a] | 1,797,409 | *18%* | 1,739,617 | *17%* | 1,800,364 | *17%* | 1,611,820 | *16%* |
| Drop recipients of premium subsidies | 1,551,253 | *16%* | 1,597,567 | *16%* | 1,668,923 | *16%* | 1,505,854 | *15%* |
| Drop RDS and missing risk scores | 1,221,252 | *12%* | 1,307,966 | *13%* | 1,356,861 | *13%* | 1,365,239 | *14%* |
| **Baseline sample** | **1,221,252** | *12%* | **1,307,966** | *13%* | **1,356,861** | *13%* | **1,365,239** | *14%* |
| Panel sub-sample | 871,818 | *9%* | 911,403 | *9%* | 954,494 | *9%* | 998,014 | *10%* |

The table shows the restrictions to the original sample of 20 % Medicare beneficiaries that were imposed to get to the baseline sample. The key restriction was to drop observations on individuals who didn't enroll in any Part D plan or enrolled in Part D through their managed care plan rather than through a stand-alone prescription drug plan (PDP). For years 2007-2009, I kept only individuals who were enrolled in a PDP for the whole year with the exception of the 65 year olds - this excludes those individuals who were allowed to join the plan outside of the open enrollment period because they e.g. changed their state of residence. In 2006, given the different special open enrollment period, many individuals were not enrolled for all 12 months and so I keep all individuals who initiated enrollment at some point during 2006 and didn't leave in subsequent months of 2006.

[a] Mainly drops those who did not enroll in Part D at all and those who enrolled in Medicare Advantage or other Part D coverage options.

### 7.3 Empirical model: specification checks and fit of the choice model; point estimates of the pricing regression

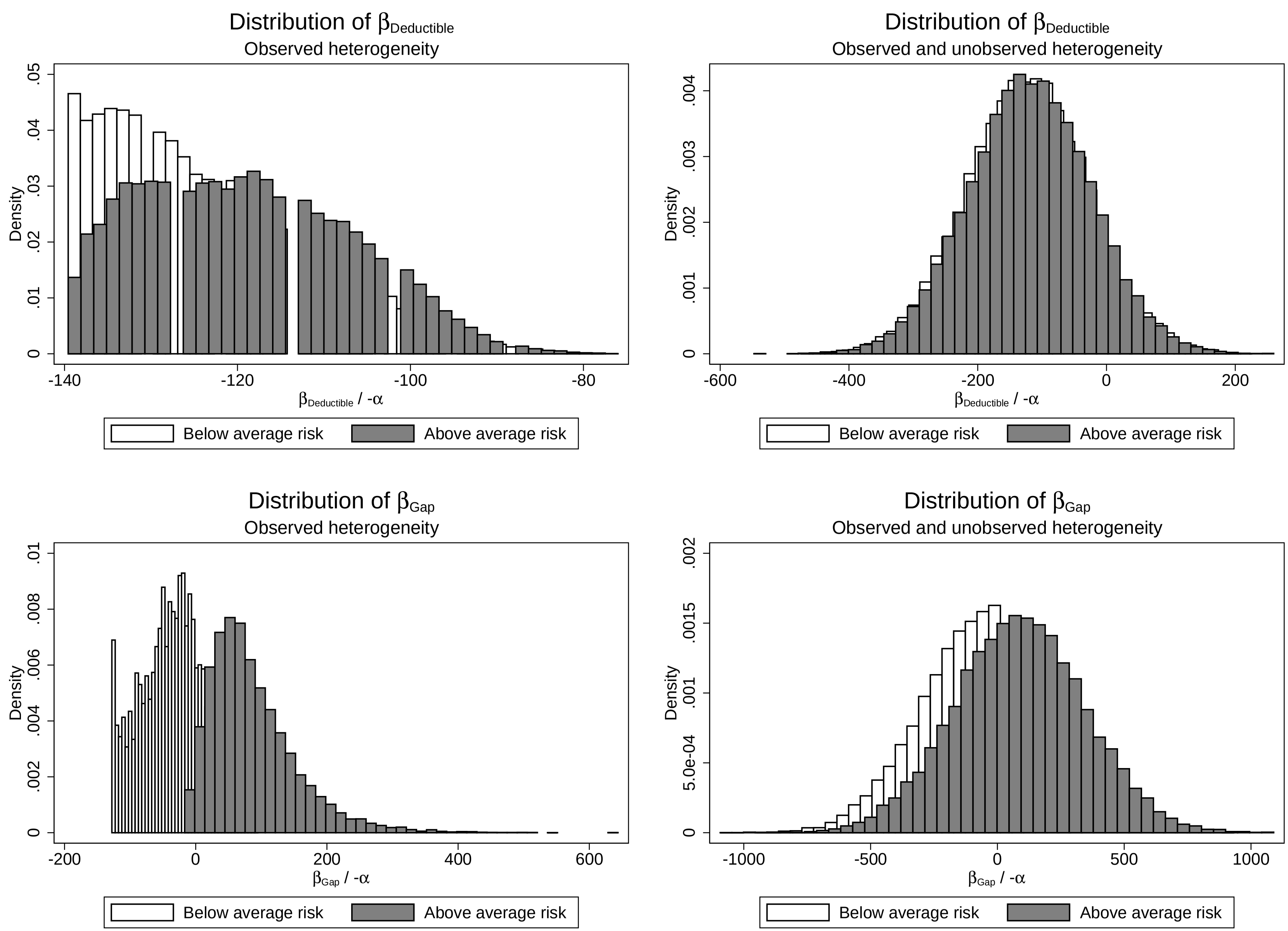
The following set of tables and figures report additional results related to the empirical model in Section 4. Table A.2 reports several alternative specifications of the choice model. The main differences across the specification are whether there is observed and unobserved heterogeneity, and if unobserved heterogeneity is present, how it is specified. I also report several specifications that do not use the control function IV to assess how the inclusion of the instrumental variable changes the results. Figure A.1 adds to the reported point estimates by plotting the estimated distributions of observed and unobserved heterogeneity for the baseline model specification. Tables A.3 and A.4 and Figures A.2 and A.3 report several metrics and simulations of model for for the baseline choice model. Table A.5 reports the results of the pricing regression.

Table A.2: **Contract choice model specifications**

| | Type of heterogeneity | | | | | |
|---|---|---|---|---|---|---|
| | No | No | Observed | Unobserved | Unobserved | LN Unobsered |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Annual premium, \$100 | -0.3727 | -0.4042 | -0.4299 | -0.4148 | -0.5272 | -0.5052 |
| | (0.0070) | (0.0096) | (0.0099) | (0.0078) | (0.0126) | (0.0123) |
| Deductible, \$100, $\mu$ | -0.4329 | -0.4405 | -0.8430 | -1.2329 | -1.1415 | 0.305 [-1.36] |
| | (0.0104) | (0.0105) | (0.1110) | (0.1266) | (0.1331) | (0.0906) |
| $\sigma$ | - | - | - | 0.4802 | 0.4112 | 0.304 [0.44] |
| | | | | (0.0232) | (0.0249) | (0.0303) |
| x Risk | - | - | 0.0715 | 0.0393 | 0.0529 | 0.0703 |
| | | | (0.0329) | (0.0367) | (0.0389) | (0.0362) |
| ICL, \$100 | 0.0305 | 0.0325 | -0.0839 | -0.1650 | -0.1631 | -1.648 [-0.19] |
| | (0.0017) | (0.0017) | (0.0179) | (0.0262) | (0.0276) | (0.1268) |
| $\sigma$ | - | - | - | 0.0815 | 0.0849 | 0.402 [0.087] |
| | | | | (0.0053) | (0.0055) | (0.0408) |
| x Risk | - | - | 0.0812 | 0.1052 | 0.1125 | 0.1061 |
| | | | (0.0045) | (0.0066) | (0.0071) | (0.0062) |
| Partial coverage in gap, 1/0 | 0.4102 | 0.4717 | -1.4298 | -2.0850 | -2.0823 | 0.55 [-1.734] |
| | (0.0297) | (0.0323) | (0.2754) | (0.3358) | (0.3634) | (0.1856) |
| $\sigma$ | - | - | - | 1.2640 | 1.3023 | 0.979 [3.55] |
| | | | | (0.0522) | (0.0589) | 0.1826 |
| x Risk | - | - | 0.8983 | 1.0954 | 1.1030 | 1.0729 |
| | | | (0.0762) | (0.0897) | (0.0976) | (0.0913) |
| Default plan, 1/0 | 5.7324 | 5.7330 | 5.1096 | 5.0675 | 4.8150 | 5.0449 |
| | (0.0213) | (0.0213) | (0.2413) | (0.2584) | (0.2917) | (0.2762) |
| x Risk | - | - | 0.2365 | 0.3589 | 0.4033 | 0.3788 |
| | | | (0.0612) | (0.0655) | (0.0741) | (0.0700) |
| Heterogeneity in preferences for specific insurers | No | Yes (observed) | Yes (observed) | Yes (obs on top 2) | Yes (obs+ unobs) | Yes (obs + unobs on top 2) |
| Control Function IV | No | Yes | Yes | No | Yes | Yes |
| Number of rand. coefficients | 0 | 0 | 3 | 3 | 13 | 5 |
| Observations | 2,435,171 | 2,435,171 | 2,435,171 | 2,435,171 | 2,435,171 | 2,435,171 |
| SL at convergence | -60,179 | -59,536 | -59,224 | -59,291 | -58,655 | -59,106 |
| **$\gamma$, 75 y.o. female, av.risk** | **5.73** | **5.70** | **5.80** | **5.78** | **5.44** | **5.63** |
| **$\frac{\gamma}{\alpha}$, 75 y.o. female, av.risk** | **\$1,538** | **\$1,326** | **\$1,164** | **\$1,392** | **\$1,032** | **\$1,114** |

The table reports estimates of utility parameters and not marginal effects. Reported are only the key estimates; the model also includes other contract parameters, demographic interactions and fixed effects as discussed in the main text. The IV specification uses the control function approach. In Column (6) the square brakets report the median or standard deviation of the random coefficients based on the point-estimates for the mean and s.d. of the natural logarithm of the coefficients.

Figure A.1: **Distributions of observed and unobserved heterogeneity in valuation of deductible and gap coverage**




Distributions are based on the estiamted mean and standard deviation of random coefficients from the baseline model specification in Table 5.1. The distributions are simulated in the estimation sample, and hence reflect the observed distribution of demographics.

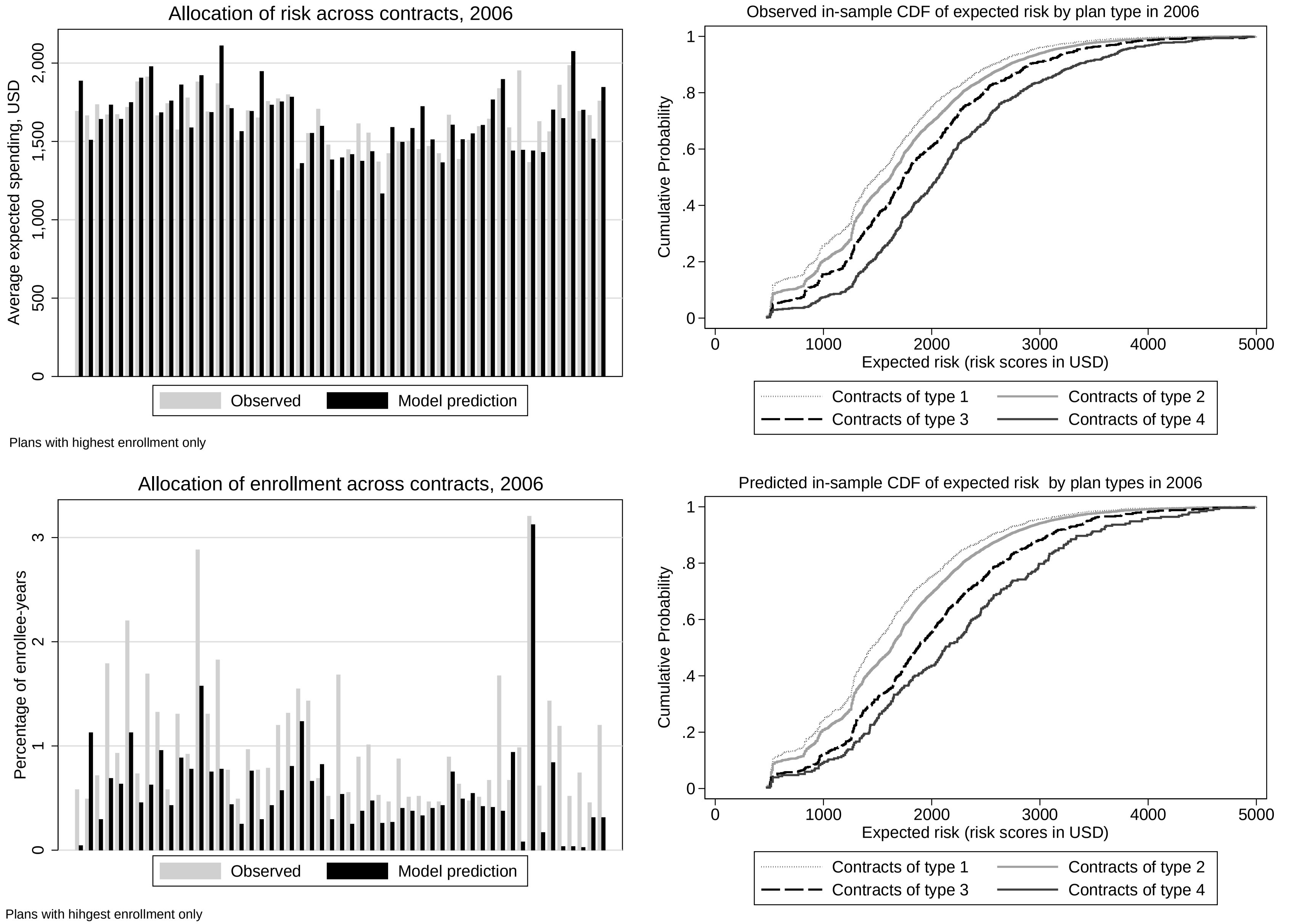
Table A.3: **Choice model fit: summary statistics by contract type and insurer for enrollment and risk distribution moments**

| | **Enrollment** | | |
|---|---|---|---|
| | Observed | Model simulation with observed defaults | Model simulation without observed defaults |
| Contracts of type 1 | 21.71% | 20.71% | 24.16% |
| Contracts of type 2 | 65.93% | 69.78% | 68.34% |
| Contracts of type 3 | 10.78% | 8.88% | 6.94% |
| Contracts of type 4 | 1.58% | 0.63% | 0.56% |
| Insurer A | 29.65% | 30.77% | 28.78% |
| Insurer B | 27.10% | 25.91% | 22.32% |

| | **Average risk score** | | |
|---|---|---|---|
| | Observed | Model simulation with observed defaults | Model simulation without observed defaults |
| Contracts of type 1 | 0.85 | 0.84 | 0.86 |
| Contracts of type 2 | 0.88 | 0.89 | 0.89 |
| Contracts of type 3 | 1.01 | 1.03 | 1.00 |
| Contracts of type 4 | 1.04 | 1.08 | 1.08 |
| Insurer A | 0.92 | 0.92 | 0.91 |
| Insurer B | 0.86 | 0.85 | 0.86 |

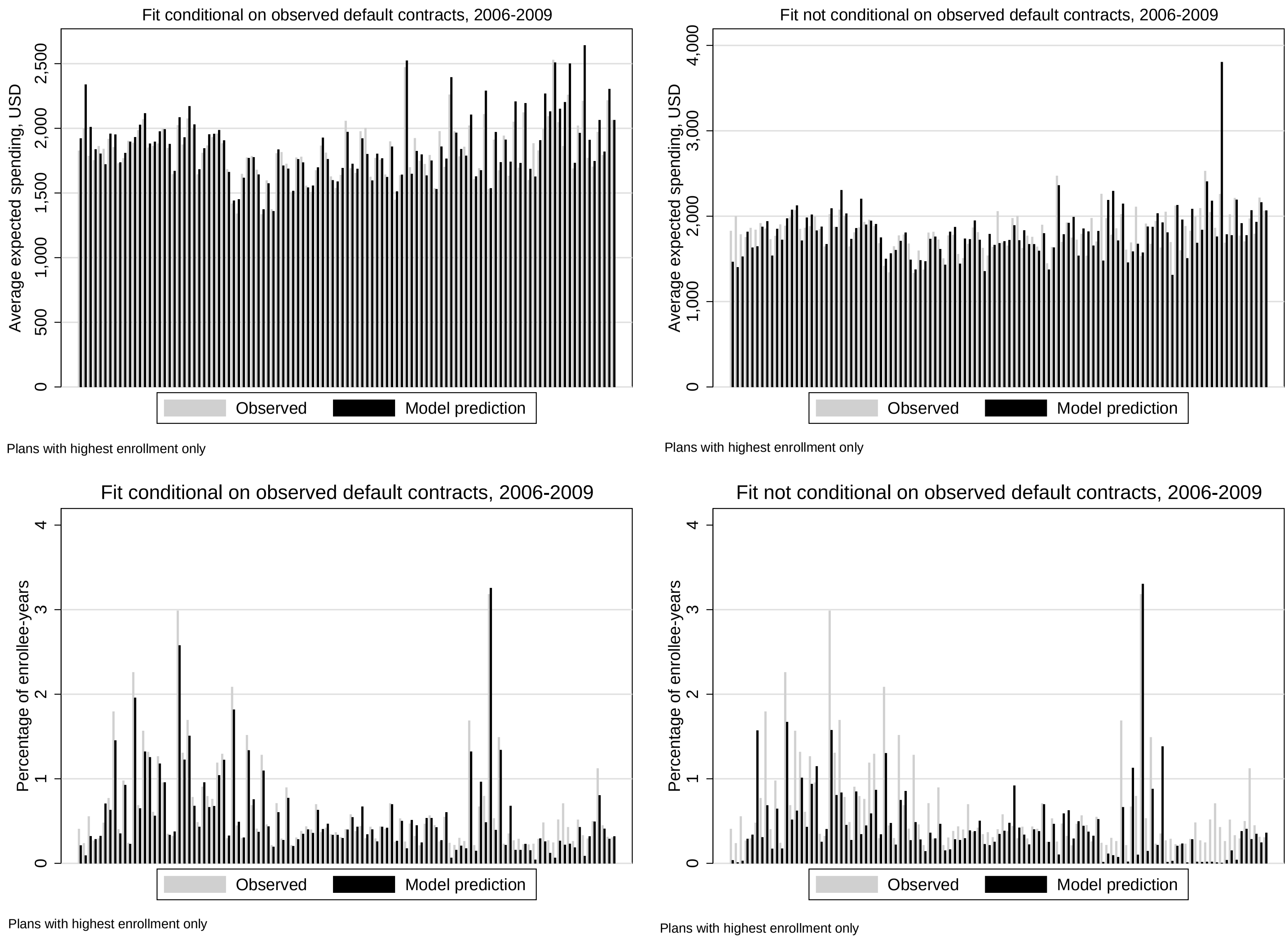
The table compares three within-sample predicted and observed moments in the data: 1) Enrollment shares in different types of plans and in different insurer brands; 2) Average drug spending in different types of plans and in different insurer brands and 3) Average risk scores in different types of plans and in different insurer brands. The data is pooled over time and regions. To simplify the contract space, the comparison is made at the 4-type plan aggregation and at brand-level aggregation for the top 2 insurers. A more disaggregated fit of the model is illustrated in Figure A.2. For the risk scores and drug bills, "predicted" measures refer to the sorting of the observed risks and expenditures as suggested by the simulation of the choice model.

Figure A.2: **In-sample fit of the choice model: fit in the first year of the program (2006)**




To construct simulated enrollment, the estimated coefficients of the choice model together with simulated random components were used to find the contract with the highest utility in each individual's choice set. The observed risk scores of the individuals predicted to enroll in different plans were used to compute the average predicted risk. Each pair of bars in the graph represents a different Medicare Part D plan ("plan" is region-specific). The graphs display only top 90 out of 2,357 contracts.

Figure A.3: **In-sample fit of the choice model: panel fit with and without observed defaults**




To construct simulated enrollment, the estimated coefficients of the choice model together with simulated random components were used to find the contract with the highest utility in each individual's choice set. The observed risk scores of the individuals predicted to enroll in different plans were used to compute the average predicted risk. Each pair of bars in the graph represents a different Medicare Part D plan ("plan" is region-specific). The graphs display only top 90 out of 2,357 contracts.

Table A.4: **Basic descriptive evidence generated in the model: share of enrollees choosing the "default" option**

| | 2007 | 2008 | 2009 |
|---|---|---|---|
| **1. Share observed in the baseline sample** | | | |
| Probability of choosing default plan for 66+ y.o. enrollees | 89.9 % | 88.7 % | 89.1 % |
| $N$ | 1,089,978 | 1,162,545 | 1,194,036 |
| **2. Share observed in the estimation sample** | | | |
| Probability of choosing default plan for 66+ y.o. enrollees | 89.9% | 89.5% | 89.6% |
| $N$ | 11,170 | 11,640 | 12,197 |
| **3. Share predicted in the estimation sample (conditional on observed defaults)** | | | |
| Probability of choosing default plan for 66+ y.o. enrollees | 86.3% | 84.8% | 86.3% |
| $N$ | 11,170 | 11,640 | 12,197 |
| **4. Share predicted in the estimation sample (not conditional on observed defaults)** | | | |
| Probability of choosing default plan for 66+ y.o. enrollees | 89.4% | 88.5% | 88.9% |
| $N$ | 11,170 | 11,640 | 12,197 |

This tables reports the simulation of the baseline descriptive evidence on the switching rates in the contract choice model.

Table A.5: **Pricing model used for the simulation of premiums in the counterfactual scenarios**

$$E[Y_{jbt}|.] = \alpha_b + \delta_r + M'_{jbt-1}\beta + \gamma_1 Ded_{jbt} + \gamma_2 ICL_{jbt} + \gamma_3 1\{PartialGap\}_{jbt}$$

where $j$ indexes region-specific plans, $b$ indexes insurers (brands), $r$ indexes 34 Part D regions, $t$ indexes years

| | (1)<br>Annual premium, USD |
|---|---|
| Lagged mean spending | 0.132*** |
| | (0.00992) |
| Deductible amount, USD | -0.489*** |
| | (0.0262) |
| ICL amount, USD | 0.312*** |
| | (0.0198) |
| Partial coverage in the gap, 1/0 | 293.9*** |
| | (11.89) |
| Insurer FE | Yes |
| Region FE | Yes |
| N | 2566 |
| Mean Y | 540.2 |
| St. dev. Y | 253.3 |
| R-squared | 0.802 |

Clustered standard errors in parentheses

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

The pricing regression is estimated on a dataset that records, for all prescription drug plans, the annual premium, the mean, the standard deviation and other moments of the lagged drug spending distribution in the plan (by plan enrollees in the baseline sample). The data also records the key financial characteristics of the plans - the deductible, the ICL and the gap coverage indicator of each plan in the program for years 2007-2009. For the cases where plans changed their ID over time due to mergers, I use Medicare plan cross-walk to match plans. The regression output doesn't report the coefficients on the set of fixed effects, as well as on the standard deviation, the kurtosis, the inter-quartile range, the 95th and 5th percentiles of the lagged distribution of realized expenditures, but these variables are included in the regression.

### 7.4 Model-free evidence from Section 3 for all years in the data

Figure A.4: **Positive correlation tests: years 2006-2009**

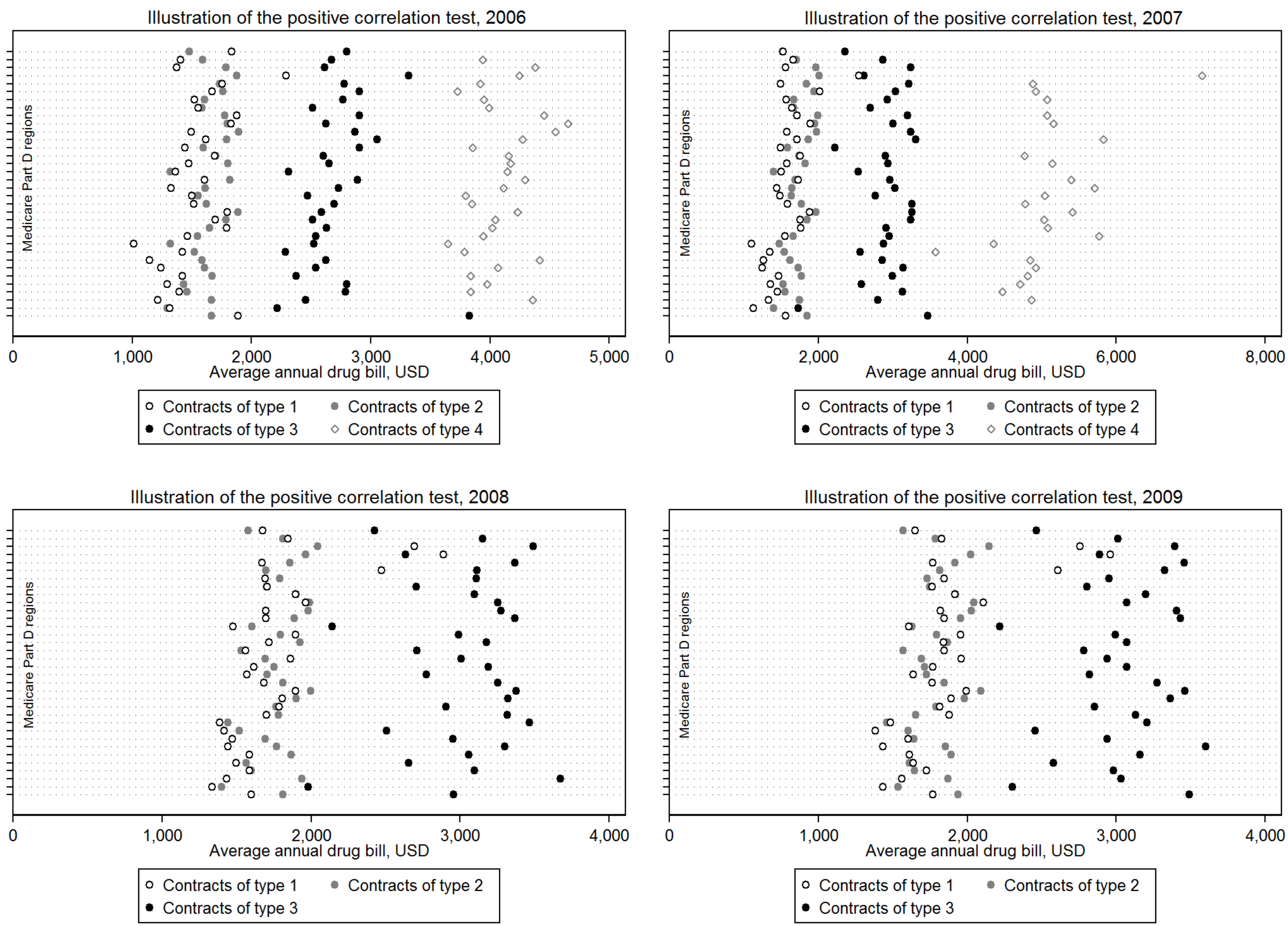


Figure A.5: **Distribution of risks by type of plan: years 2006-2009**

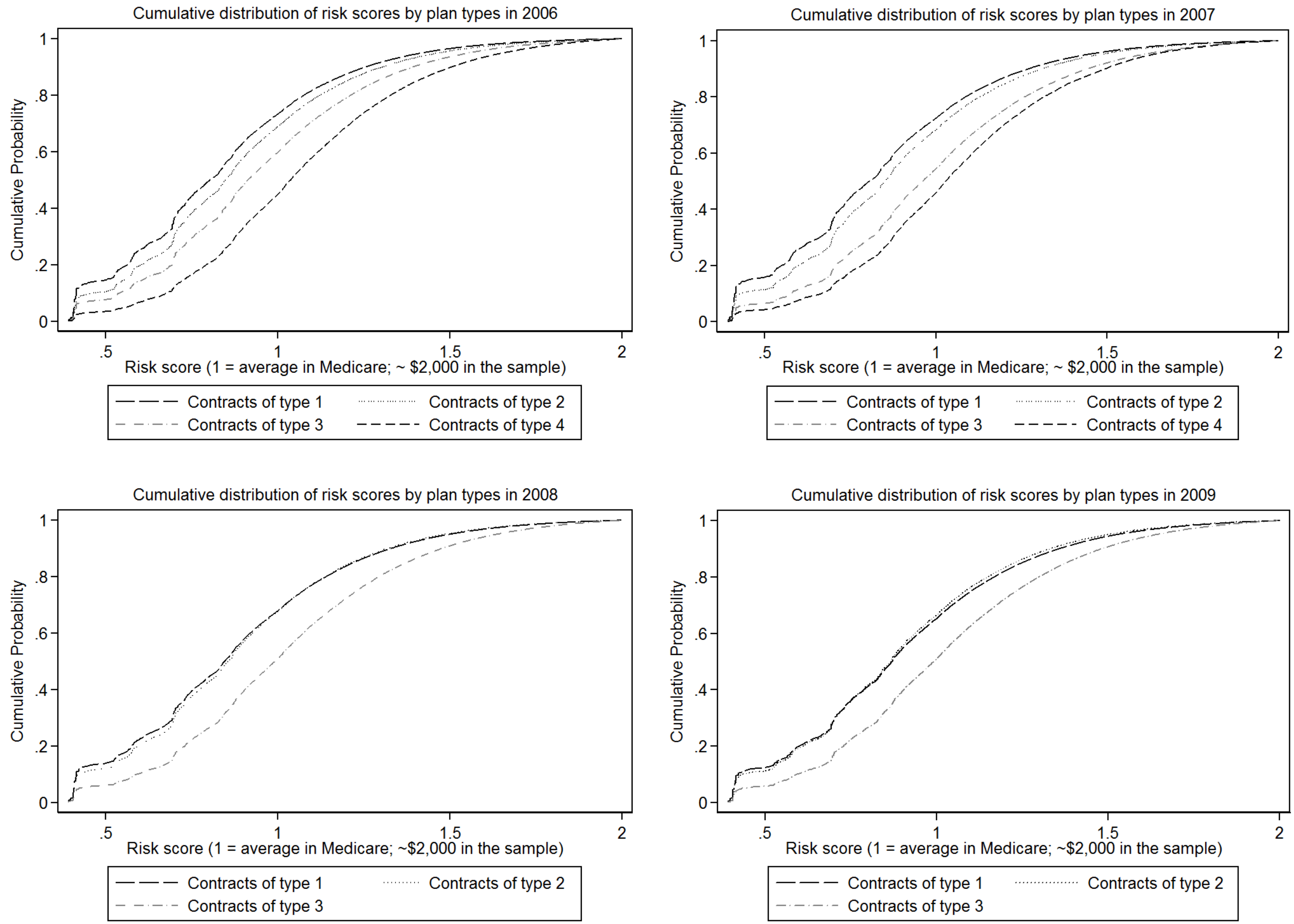
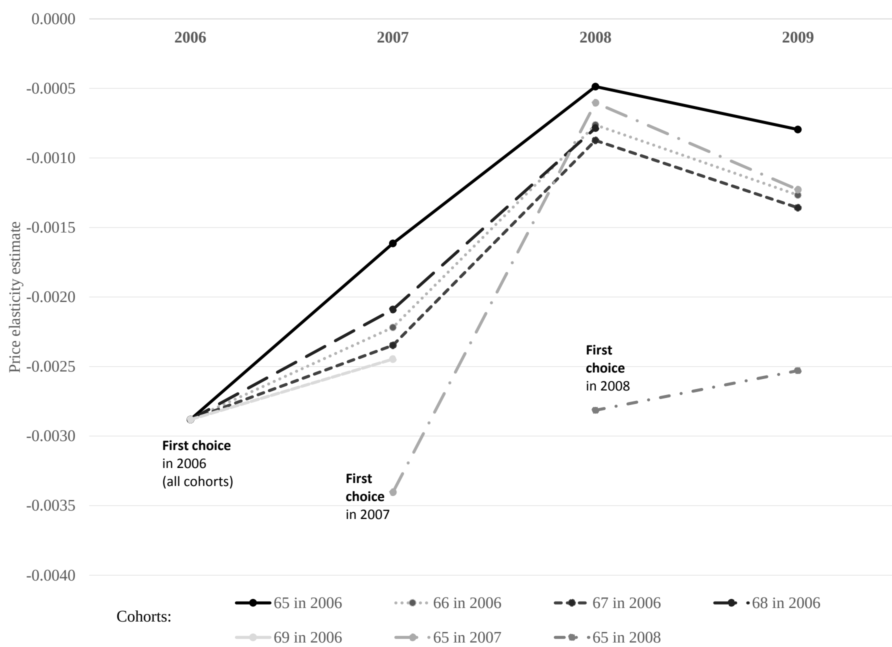


Figure A.6: **Evidence of switching costs: price sensitivity estimates by cohorts over time**




I use a simple conditional logit regression to test whether there are statistically significant differences in the price sensitivity of the cohorts of new and continuing enrollees. Under the null hypothesis of no switching costs, we would expect the coefficients on plan premiums for new (65 y.o.) and existing enrollees of similar age (66 - 70 y.o.) in the same year to be very close to each other. The estimates allow me to reject this null. I find that price sensitivity is significantly higher in magnitude for 65 year olds than for all cohorts of 66-70 year olds in years 2007-2009. This does not hold in 2006 when beneficiaries of all ages are entering the program anew. Furthermore, the estimates of the price coefficient are virtually identical for each age group among 66-70 year olds, suggesting that the difference between the estimated price sensitivity for the new and continuing cohorts is not driven by age differences per se, but instead are related to the lack of switching costs for the 65 year old beneficiaries. The price coefficients are estimated using the following random utility specification:

$$
\begin{aligned}
u_{ij} &= -\alpha_{65}p_{ij} + \alpha_{66}p_{ij}\mathbf{1}\{Age = 66\} + \alpha_{67}p_{ij}\mathbf{1}\{Age = 67\} + \\
&+ \alpha_{68}p_{ij}\mathbf{1}\{Age = 68\} + \alpha_{69}p_{ij}\mathbf{1}\{Age = 69\} + \alpha_{70}p_{ij}\mathbf{1}\{Age = 70\} + brand_j + \epsilon_{ij}
\end{aligned}
$$

$\epsilon_{ij} \sim iid$ Type 1 EV. The specification includes fixed effects for eight largest insurers. The estimates use separate cross-sectional parts of the data sample that is used later to estimate the full choice model. The sample is restricted to only include individuals that are 65-70 years old. The graph plots the (sum of) coefficients on premiums in the utility function and not marginal effects.