Health Insurance Plan Design and Risk Protection: Evidence on Variation in Cost-sharing for ACA Exchange Plans

1. Introduction

In many health insurance markets in the U.S., consumers can choose between several plans that differ in many dimensions. A robust literature explores employer-sponsored health insurance and Medical Part D, and establishes how consumers select plans from the many options they face and are often confused and appear to choose sub-optimally (Abaluck and Gruber 2011, Loewenstein et al. 2013, Handel and Kolstad 2015, Bhargava et al 2017). There has been much less research, however, into the supply-side of these markets to understand better which options are available to consumers.

This paper addresses the variation in health insurance options in the context of private-market insurance plans available through the Affordable Care Act (ACA) Exchanges. The individual market established following the ACA since 2014 provides a unique opportunity to study the issue. Unlike employer-sponsored health insurance where employers help choose and design a limited number of plan options, consumers in the Exchanges usually face a larger choice set requiring their own active decision. In 2017, an average consumer in the Exchanges faced more than 30 plan options (Avery et al., 2016). Across the country, there were more than 3,000 unique plan designs in the market. Recent regulatory changes that eliminate the "meaningful difference requirement" allow insurers to launch multiple plans with only small differences in design, which may further proliferate supply-side innovations in plan designs. Rich information about plan designs in the private market is made publicly available, providing a valuable chance to examine the landscape of types of plans available and their economic implications.

This paper evaluates the Exchange plan options by their financial values and explores how much risk protections are provided by different plans. Current regulation places limits on the actuarial value (AV) of plans, which is the share of medical spending the plan is expected to cover for an average population. Plans are available in four metal tiers (Bronze, Silver, Gold, and Platinum) with actuarial values of 60%, 70%, 80%, and 90%. So a Silver plan is expected to cover 70% of the total medical expenditure. Current policy mainly emphasizes the differences in financial protection in terms of AV. If one wants to get more coverage, purchase the Platinum plan; if one wants to get less coverage, choose the Bronze plan. However, the point of insurance is not only to get expected coverage, but also to "reduce the variance" of spending. Holding the AV fixed, insurers are free to offer plan designs with different combinations of deductibles, co-pays, co-insurance, and max-out-of-pocket limits (MOOP). I establish that these features of plan designs can, in theory, have very large effects on the risk protection afforded by a plan. The empirical question of interest for this study is whether there is economically meaningful variation in risk protection across plans available in the market. Further, to the extent that plans offer substantially different risk profiles, can we understand why this variation exists?

The primary finding of this study is that there is substantial and economically meaningful variation in the risk protection offered by plans available in the market within the same actuarial-value tier. Using data on financial attributes of all plans launched in the federal Marketplace, I calculate the out-of-pocket spending distribution using a representative ex-ante medical expenditure distribution. I then construct two measures on the financial protection provided by each plan within the same AV level: the standard deviation of out-of-pocket spending and the risk premium relative to full insurance. A typical consumer comparing Exchange plans in the Silver tier, for example, will have 10 plan options. The standard deviation of the uninsured risks varies by more than $1,000 across available plans. With moderate levels of risk aversion, a consumer who chose a silver plan with the best possible risk protection could get an increase in value equivalent to being given the average Gold plan. These differences are economically significant because plans with different risk premiums are equally likely to attract enrollees.

After documenting these new stylized facts on the variation in risk protection available across Exchange plans in the same coverage tier, I turn to the issue of trying to understand what explains why this variation exists. Why do plans with higher out-of-pocket risk exist in the market alongside lower-risk plans? More generally, why has the market not converged on a common plan design within an AV level? One possibility is that the observed variation is driven by temporary plan-design experimentation, but I find that the variation in risk protection is not converging over time. A second possibility is that market-level factors, either on the supply side or demand side, create cross-geographic variation in plan designs. However, I find substantial variation in risk protection across available plans *within* almost all geographic regions in the Exchanges. I also find that most insurers operating in the Exchanges offer multiple plans within the same metal tier with significant variation in risk protection. Together these findings suggest that the variation in plan design is not driven by forces that vary at the market level.

I next explore the possibility that firms are offering these different designs to cater to heterogeneity in consumer preferences. Note, however, that the work on confusion in health insurance significantly calls into question the ability of consumers to sort effectively across plan variations. As such, this analysis should be thought of as exploring whether consumer heterogeneity could rationalize plan variations and not necessarily whether the observed plan variation represents an optimal market response to consumer heterogeneity. I simulate plan values under a range of moral hazard responses and illustrate that the variation in terms of moral hazard is also unlikely to rationalize having substantial plan variation in the market. I then examine the heterogeneity in ex-ante medical spending risk. I show that even though a large proportion of plan designs cannot be rationalized by heterogeneity in ex-ante risk types, a small group of higher risk premium plans may be attractive to healthy types. Claim data suggests that the average allowed expenditure is significantly higher for straight deductible plans, consistent with consumers sorting by ex-ante risk types.

These results leave open the possibility, however, that other forms of consumer heterogeneity could potentially rationalize the variation in plan designs. I discuss specifically the possible impacts of consumer heterogeneity in liquidity constraints. The results may also be consistent with behavioral biases, such as confusion that causes a lack of competitive pressure that would otherwise push the market to lower-risk plan designs.

This paper presents the first systematic analyses of the range of plan designs available in the ACA Exchanges. A robust literature documents evidence that some options in the employer-sponsored health plan financially dominate others (Handel 2013, Bhargava et al. 2017). This paper contributes to the literature by providing evidence that the variation in financial values is prevalent in the individual market. More broadly, the paper contributes to the literature on the design of health insurance financial attributes and consumer behavior. Previous research explores plan standardization (Ericson and Starc, 2015), the externality of stand-alone drug plans (Starc and Town, 2015), and cream skimming of drug formulary (Geruso et al., 2016). This paper also contributes to the growing literature on individual health insurance markets. The existing literature explores topics such as narrow networks (Ghili 2017), the effects of premium subsidy (Tebaldi 2018), market competition (Dafny et al. 2015), and labor market search (Aizawa and Fang 2015). This paper is the first to study the variation in financial protection across Exchange plans.

The paper is organized as follows. In Section 2, I describe the dataset; in Section 3, I demonstrate the magnitude of the differences in risk protection among plans and hence in their financial values; in Section 4, I explore the reasons driving the variation; and finally, in Section 5 I conclude and discusses the implication for future research.

2. Data

I use two sources of data. The first is the information on plans launched through the federal marketplace. The second is data on total medical spending distributions from both the Center for Medicare and Medicaid Services (CMS) Actuarial Value Calculator and the Medical Expenditure Panel Survey (MEPS).

2.1 ACA plan information

Exchange plan information comes from the Center for Medicare and Medicaid Services Health Insurance Exchange Public Use Files.[1] The data include all plans offered in the individual marketplace for the years 2014 to 2017 in the 40 states using the federal exchange system (e.g., Healthcare.gov).[2] The dataset contains information on full plan designs, including deductibles, MOOP, and cost-sharing rules for various services like physician visits, inpatient stays, emergency room, and outpatient surgery. Other than financial attributes, the dataset also contains information on the counties within which the plan is launched, the carrier of the plan (the plan-state enrollment number is available for the year 2015 to 2016). To supplement the analyses, I also combine these data with information on the number of enrollees to the exchange plans in each county from the Open Enrollment Period County Level Public Use File,[3] and insurer information, such as total member months, from their 2016 Medical Loss Ratio Filings, the most recent year available. [4]

---

[1] See https://www.cms.gov/CCIIO/Resources/Data-Resources/marketplace-puf.html

[2] Appendix table 1 shows the states in the analyses sample. Information about plans offered through the state-based exchange, and plans offered off the exchange, are less complete and are not uniformly reported, especially for 2014 and 2015. As a result I did not include them in the sample.

[3] See https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Marketplace-Products/Plan_Selection_ZIP.html

[4] See https://www.cms.gov/CCIIO/Resources/Data-Resources/mlr.html

There are a large number of plans available in the market, allowing exploration of the variation in plan designs. I define each observation based on the plan id, which represents a unique combination of financial schedule, network, service area, and drug formulary. Each Silver plan has three cost-sharing variations available only to certain low-income households. To maintain homogeneity, I exclude these variations and refer to the remaining sample as standard plans. I use the 2017 plan-year in the primary analyses as this is the most recent sample with complete information.[5] In 2017, there are 3,106 standard plans available across 2,724 counties in the 39 states covered by the data set.

I ignore several dimensions of the plan design due to insignificance or data limitations. For financial features, I restrict my attention to individual coverage. Family coverage is similar but more difficult to compare efficiently. When a plan has tiered medical coverage, I focus on the cost-sharing for the in-network tier-one coverage, which is mostly used by enrollees. Next, I account for cost-sharing rules in different drug tiers (generics, specialty, preferred brand, non-preferred brand) but ignore the drug formulary. Different plans may put the same drug into different tiers, but such information is not publicly available. I also cannot account for the variation in the negotiated price for each service in each plan.[6] It is important to note that the purpose of the analyses is not to measure all aspects of a plan for every possible spending scenario. Instead, the goal is to provide a benchmark for the average, most commonly used services. The analyses are valid because such omission represents small variation in a plan's financial value.

2.2 Medical Distribution

To evaluate the plans' financial values, I need to start with the same ex-ante medical expenditure distribution that is applied to all plans, so the comparison is fair. For this analysis, I use the medical expenditure distribution from the 2017 Actuarial Value Calculator Continuance table (referred to as the AV Calculator Distribution hereafter). This distribution is constructed based on a large commercial claim database as well as the claim data of Exchange enrollees and represents the national average distribution for individuals who enroll in the marketplace plans. This distribution is used to measure the actuarial value of plans for the purposes of meeting

[5] In later sections, I discuss how the result 2017 compare with the other years.

[6] This might be relevant if the coverage is in the deductible range, or the individual pays a coinsurance rate. Note that if price negotiations are at the insurer level (rather than plan level) this likely will not be an issue since most insurers are offering multiple plan designs (as I document below).

regulatory requirements around actuarial value. As such, it provides a useful benchmark because insurers need to hit the actuarial value target based on this distribution when designing a plan.[7] Of course, any given consumer might face a different ex-ante distribution of medical risk, which is an issue I discuss after presenting the main results using this benchmark distribution.

The data I observe is a distribution discretized into 84 buckets based on total spending levels. The first bucket is spending at $0. The second bucket is spending between $100 and $200, the third bucket is spending between $200 and $300, etc. In Appendix Figure 1, I show the probability distribution function by plotting the probability of falling into each bucket against the average spending in each bucket. Other than the probability of each bucket and the average total spending, I also observe the breakdown of the spending and frequency of utilization of twenty subcategories of benefits, including primary care visits, outpatient, inpatient, emergency room, drug, etc. These benefits are required to be covered by all plans (Essential Health Benefits). This information allows me to exploit the variation in the service-specific cost-sharing rule for each plan.[8]

For each plan, I determine the out-of-pocket spending by mapping the discretized total spending level based on the financial attributes. For a plan with a deductible d, MOOP m, and coinsurance rate c, the out-of-pocket spending oop given a total spending level t is:

$$oop = \min(t * (t < d) + [d + c * (t - d)] * (t \geq d), m)$$

In some cases, there are service-specific cost-sharing rules. For example, some plans may cover primary office visits before hitting the deductible level. They may also have a copayment for a specific service. In these cases, I break down the total spending level into each subcategory and use the information on service utilization frequency and spending from the AV Calculator Continuance table to get the out-of-spending level. This allows me to determine the out-of-pocket spending distribution for all plans based on the AV Calculator Distribution.

[7] The AV Calculator has four distributions which applies to plans in the Bronze, Silver, Gold and Platinum tiers respectively. I used the distributions associated with the Gold tier. The results are similar using other metal distribution. But the point is to apply the same distribution to all plan schedules. In section 4.3 I will discuss how the ranking of plans will shuffle when using other total spending distributions.
[8] For plans in my main sample, the average percent of the plan's total premium contributing to the Essential Health Benefits coverage is 99.6%.

3 Variation in Risk Protection Provided by Exchange Plans

Before delving into details of actual plan design, it is worth discussing the potential room for variation in risk protections theoretically. Holding fixed the AV level, how different the risk protection can potentially be created by various plan designs? To illustrate, I consider three hypothetical plan design types:

(1) A deductible-only plan that covers all spending after a deductible;

(2) A coinsurance plan that pays a fixed percentage of all bills; and

(3) A benefit-cap plan that pays all bills up to a total coverage limit.

These contract types are motivated by many actual health insurance plans in the market.[9] For each AV level, there is a unique plan for each type. For example, if we want to design a plan with 70% AV, then contract type (2) will simply have a coinsurance rate of 30%. Similarly, there is a unique deductible level for contract type (1), and a unique benefit-cap for contract type (3), that will generate a 70% AV. After finding the plan of each type, I calculate the out-of-pocket spending using the AV Calculator Distribution. Figure 1 shows the standard deviation of the out-of-pocket spending of each plan type when AV ranges between 0 and 1.

The first fact is that with full insurance (AV=1) or no insurance (AV=0), all plans types have the same risk protection. This is natural as all plan types degenerate to the same plan at the two endpoints. What's interesting is that the deductible-only plan always has the lowest risks among the three when AV is between 0 and 1. In fact, as shown in Arrow (1965), the deductible-only plan has the lowest risks among *any* plan designs, holding fixed the AV level. So from the perspective of risk protection, the deductible-only plan is the optimal choice. The second striking fact is that the differences in the standard deviation of uncovered expenses can be substantial. At 70% AV, the standard deviation of uncovered expenses is $1,460 for the deductible-only plan, $6,200 for the coinsurance plan, and $17,400 for the benefit-cap plan.

ACA regulations on actuarial value and maximum-out-of-pocket limits greatly reduce the possible contract space. The contracting space consistent with ACA regulations is shown in the

[9] Empirical example of the coinsurance plan is the Medicare Part B coverage, where the proportion is set to 20%. Empirical example of the benefit cap plan is the Medicare Part A coverage, where there has no coverage when the inpatient days are beyond certain level. Before ACA, many employer-sponsored health insurance plans also have benefit caps (Mitchell 2010).

area between the dashed and solid blue line in Figure 1. However, even within the reduced contract space of ACA regulation, there is still scope for plans that generate standard deviation increases in spending beyond those in the deductible-only plan of more than $1,000. It is then an empirical question that how actual plan designs lie on this graph.

In this section, I am going to show that the actual Exchange plans expand the whole risk protection variation scope that is allowed by the regulation, and such variation is economically significant, as consumers are enrolled in all types of plans.

3.1 Variation in Deductible and MOOP Level

The first stylized fact is that even within sets of plans offered in the same metal tier, which by regulation have a similar actuarial value, there is substantial variation in the deductible and maximum-out-of-pocket limit level. In Figure 1 I show the distribution of deductible and MOOP level for the 2017 Bronze (with 60% actuarial value), Silver (with 70% actuarial value) and Gold plans (with 80% actuarial value). For example, the deductible level for Silver plans ranges between $1000 and $7150; the MOOP limit for Gold plans ranges between $2000 and $7150. As MOOP measures the worst-case risks of a plan, the variation in MOOP provides model-free evidence that plans with similar actuarial value expose consumers to a wide range of risks. Since the comparison is within a metal tier, a lower deductible often is combined with a higher MOOP. Counter to what many might initially think, it is plans with higher deductibles that expose people to less overall risk, since they have lower MOOP usually.

The variation in risk protection provided by plans is large when I consider the full schedule of the Exchange plan designs. Take the Silver plans as an example. There are actually plans that are deductible-only plans, whose standard deviation of the out-of-pocket costs is around $1,500, while the highest risk plans have a standard deviation of more than $2,500. This means that if zooming in the dashed area of Figure 1, the actual plan designs spread across the entire contract space allowed by the ACA regulation.

3.2 Variation in Risk Premium

The economic importance of this variation can be quantified using the concept of risk premium. The risk premium for a plan is defined relative to a full-insurance benchmark. It represents an amount of money that a person would need to receive to be indifferent between enrolling in that plan versus a full-insurance plan when both plans are priced at their fair actuarial value. Under the expected utility framework, risk premium R is the value satisfying the indifference relationship:

$$EU(w - a) = U\ (w - E(a) - R),$$

where w represents the wealth level, a represents the stochastic out-of-pocket spending, and $U(\cdot)$ is the utility function. The risk premium is zero for a risk-neutral enrollee and is positive for risk-averse individuals, and rises with the level of uninsured risk in the plan. Following the literature in measuring the financial value of health insurance (for example, Handel 2013), I use the constant-absolute-risk-averse utility model, so wealth level is irrelevant. I set the benchmark risk-averse coefficient at 0.0004, which is mean and median of risk aversion level estimated in Handel (2013).[10]

The variation in the risk premium of plan options within a metal tier is sizable. In Figure 3, I plot the risk premium and the expected covered spending for all plans in the standard metal tiers for 2017. The expected covered spending is a scaled function of AV, and the target metal tier is represented by the vertical line. Not all plans lie perfectly along the targeted AV line, partially because of the 2% error margin allowed by regulation and partially because of measurement error. However, the vertical differentiation exists for a range of AV levels. Consider, for example, the standard Silver-tier plans. For these plans, the smallest risk premium, relative to full insurance is around \$500. Consistent with Arrow (1965), these are the plan that set deductible equal to the MOOP limit, and they are represented by the black solid “Arrow frontier” in the graph. In contrast, the largest risk premium for Silver plans is nearly \$1,000 larger, originating from plans that typically have lower deductibles but MOOP limits closer to the maximum allowed by regulation.

The sizable variation in risk premium could be better benchmarked by comparing the financial values of plans across the metal tiers. On average, a Silver plan (70% AV) has an expected covered expenditure of \$4,100 (the purple vertical line in Figure 1), while the Gold plan (80% AV) has an expected covered expenditure of \$4,700 (the green vertical line in Figure 1). The difference is around \$600, much smaller than the risk premium difference between the best and worst silver

[10] Someone with this level of risk averse will be indifferent about accepting a 50-50 gamble where one wins \$100 and loss \$96.15 versus if nothing happened. Risk premium of plans increases monotonically when the level of risk aversion increases. Appendix Figure 2 shows the main results with a wider range of risk aversion coefficient level.

plans. In fact, the variation in risk premium within a metal tier imply potential “financial dominance” across metal tiers. Figure 4 combines risk premium and expected covered expenditure together and plot the probability density distribution of the certainty equivalent of plans by metal tier. The fact that there are crossings between metal tiers suggest that plans from lower metal tier could induce higher utility than plans in the higher metal tier. Given that plans in the lower metal tiers are also cheaper, this implies a clear financial dominance.

In the following analyses, I use the concept of “relative risk premium” to adjust for the AV variation within a metal tier. I define “relative risk premium” as the distance between the risk premium and the Arrow frontier. For example, plans on the Arrow frontier have a relative risk premium of 0. All other plans have a positive relative risk premium. Since the frontier is almost linear within a metal tier, this measure adjusts for the small variation mechanically driven by the difference in the smallest risk premium at different AV levels.

Such variation in the risk protection among available plans is economically meaningful because plans with high risk premium attract a sizable number of consumers. Using the plan-state level enrollment number in 2015 and 2016,[11] I group plans in the Bronze and Gold tier into quartiles based on the relative risk premium. I did not include Silver plans because the enrollment number represents enrollment to both the standard Silver plans as well as the cost-sharing variations. In Figure 5, I plot the mean and 95% confidence interval of the log enrollment number for each the risk premium group, each metal tier and year. I took logarithm to mitigate the effect of extreme values. In 2015, the average Bronze log enrollment is 5.5 for the 1st, 2nd and last relative risk premium quartile. For Gold plans, the first and last quartile has smaller enrollment number. In 2016, the pattern is similar. Overall, higher risk premium plans have similar enrollment number as low risk premium plans.

This stable enrollment in plans with different risk premiums transfers the large variation in risk premium *offered* to consumers to the large difference in risk premium *chosen* by consumers. Table 1 shows for the year 2016 the average risk premium chosen by consumers by mental tier. I weight each plan’s relative risk premium by its enrollment. On average, consumers are enrolled in plans with $180 more risk premium than the Arrow plan. Consumers in standard Silver, Bronze, Gold, and 73% cost-sharing variation tiers are in plans with a sizable relative risk premium. There

[11] Enrollment data is only available for year 2015 and 2016. Here I classify plans based their 2015 and 2016 relative risk premium respectively. In section 4.1 I will show that the risk premium distribution is stable over time.

is also a large difference in the relative risk premium consumer chosen: for all plans on the market, the difference in the relative risk premium between the 90th and 10th consumers are over $600. As a result, it is important to understand the reason and implication of such variation.

3.3 Correlation with Premium

The above analyses illustrate that the risk protection provided by the Exchange plans has a large variation within a metal tier. It abstracts away from the premium difference among plans within the same metal tier, partially because premium also captures differences in non-financial attributes like network, brand name, service area etc. So the analyses could be interpreted as a thought experiment, where I compare plans financial values holding fixed the premium level.

A natural question is then whether premiums are systematically correlated with the risk protection provided by a plan. In theory, if there is no selection, then the difference in risk premium should not be reflected in the premium, because insurers should only care about the expected spending level, which is already captured by metal tier (AV level). But if individuals with different risk types select into plans with different level of risk premium *within* a metal tier, and insurers can price partially for that, then there might be a correlation between premium and risk premium within a metal tier.

Lacking claim data, I would not be able to directly test the existence of selection among plans within the same metal tier. I could test, however, that the correlation between the risk premium and premium level. Since the premium of a plan differs by rating area, I restructured the dataset so that each observation is a plan-rating area combination. I then run an OLS regression of premium,[12] on plan's AV and risk premium. AV is included as there is some variation in AV within a metal tier. I control for metal tier fixed effects as the exercise is intended to compare the correlation between premium and risk premium within a metal tier. I also control for the issuer, network type, rating area fixed effects, as a way to single out confounding factors. Table 2 shows that the coefficient of the risk premium is almost zero, indicating that there is no correlation between risk premium and premium for Exchange plans in the same metal tier.

[12] The premium is measured as the monthly individual rate for a 21-year-old non-tobacco user. Premium for other age groups are mechanically determined by the "3-1" ratio. Insurers are not allowed to price separately for other demographics like gender. The premium variation for family coverage is similar as in the individual plans, though more complicated depending on the number of households etc. So I view the individual premium as a good proxy of the overall premium variation faced by enrollees.

4. Reasons for Plan Design Variation

This section explores several reasons that might drive the wide variation in the financial protection of plans. First, the variation is stable over time, suggesting it is not driven by temporary plan experimentation. Second, the variation in risk premium is mainly driven by within county difference than across county difference. The variation also not seems to be driven by aggregated demand or supply-side factors. Third, I explore the possibility that insurers might offer plans catering to heterogeneity in risk types. I find that over 85% plan designs have poor value for all individuals facing a range of ex-ante medical spending distribution. Finally, I consider how moral hazard response to the plan might change the value of plans. Again, for a range of moral hazard responses, plans' rankings are stable, suggesting moral hazard cannot easily rationale the variation in plan designs.

4.1 Risk-Premium Variation is Stable over Time

In the early years of the Exchanges, insurers might be uncertain about the selection patterns participants would have across plans, and also uncertain about any possible differences in the spending patterns induced by different cost-sharing rules. As insurers learned the market dynamics over time, plan designs should converge. Such convergence, however, does not happen. For all Silver plans launched between 2014 and 2017, I calculate the relative risk premium using the 2017 AV Calculator Distribution. Figure 6 shows that although there is a slight increase in the median of the relative risk premium (indicated by the red line in the middle of the box), the interquartile range of relative risk premium is stable at around \$350 (indicated by the blue border of the box), and the range is around \$1000 for 2014 to 2017.[13] The results using plans in other metal tiers show a similarly stable pattern of substantial variation and limited convergence in plan designs.

4.2 Market level variation

The non-convergence over time may reflect time-invariant local market differences. In the Exchange, the relevant market is defined as a county, as the smallest level at which the plan menu could vary is at the county level.[14] There is indeed some variation across counties: in Figure 7 I plot the mean relative risk premium of Silver plans in each county. Plans in part of New Mexico,

[13] The increase in median risk premium over time might be driven by the fact that medical expenditure is increasing over years, but I evaluate plans using the same medical expenditure distributions. The CMS updates the AV Calculator distribution over time though.

[14] Plan's premium varies at rating area level, which is usually a collection of counties. However, insurers can offer plans only in part of the rating area, and it is always a collection of counties Fang and Ko (2018).

Alabama and North Carolina seems to have a higher average risk premium. There is also a larger difference within a state than between states, which is consistent with the fact that insurers' county entry decision, as well as plan offering decision, is highly correlated within the state.

To quantify the size of the cross-markets variation, I calculate the mean and standard deviation of the Silver plan relative risk premium for each county. I then rank counties by their mean relative risk premium and plot both statistics in Figure 8. Among the 2600 counties, over half of them have the average relative risk premium between \$400 and \$600. The difference between the 90th and 10th percentile county is \$380. On the other hand, each county has a sizable within-county variation: the standard deviation of the risk premium, is over \$100 for all most all counties. This pattern is not unique for Silver plans: the difference is relative risk premium between the 10th and 90th percentile county is \$196 and \$317 for Bronze and Gold plans respectively.

Similar variation in risk premium is also prevalent in counties with different supply and aggregate demand side characteristics. Using County Open Enrollment Period Profile, I classify each county by total enrollment quartiles, which measures the overall market size of a county. In Panel A of Figure 9, I show that even though the median of risk premium increases slightly with regard to market size, the interquartile range and range are similar for plans in counties with different market size. I also group counties based on the number of insurers, which measures the competitive environment. Since less than 5% of counties have more than 4 insurers, I group these counties together. Panel B of Figure 9 shows that both the interquartile range and range of relative risk premium is similar for all groups. The results are robust using other metal tiers.

The next stylized fact is that the risk premium variation is not driven by a specific type of insurer. First, many insurers seem to offer a wide range of plan designs. I group plans by the insurer and calculate the range of relative risk premium of 2017 Silver plans within an insurer. Figure 10 Panel A illustrates that half of all insurers offer plans with more than a \$500 difference in risk premiums. One may wonder whether larger, or not-for-profit insures are more likely to offer plans converging to the Arrow frontier. To test this, I link the main dataset to the Medical Loss Ratio (MLR) filing of each insurer and group insurers by its overall member months in the individual market.[15] Figure 10 Panel B and C shows the distribution of 2017 Silver plan relative risk premium

[15] Not all insurers with plan information are matched with the MLR filing. In total, I have 115 out of 167 number of insurers in my sample matched. The member month measure for each insurer is calculated based on both on and off exchange enrollment. Insurers may offer different plans off the federal exchange platform, but I don't have measures on the size of that.

by two insurer classifications: whether an insurer is not-for-profit, and whether the insurers' individual member month is larger than the median. Again, the distribution of risk premiums is similar for both groups. The analyses do not rule out the possibility that there is heterogeneity in the ways insurers design plans' financial attributes. Instead, it shows that such heterogeneity does not seem to correlate with insurer type and size.

In summary, the variation in risk premium is mainly driven by within-market variation than by cross-market variation. The difference in risk premium is hard to explain by county difference in aggregate demand or supply factors.

4.3 Heterogeneity in Moral Hazard

In the above analyses, I have examined the risk protection provided by the Exchange plans in the case where medical spending needs are not affected by the plan structure. There is, however, a robust literature documenting the phenomenon of "moral hazard" in which people adjust their medical spending in response to the level of cost-sharing they face in their plan (Aron-Dine et al. 2015, Einav et al. 2013, and Brot-Goldberg 2017). When medical spending responds to cost-sharing structures, it can change the optimal insurance design (Drèze and Schokkeart, 2013). In the classic case of moral hazard, people overuse care when they have generous coverage (Pauly, 1968) and hence might prefer a contract with higher cost-sharing so they are not paying premiums to insure low-value care. On the other hand, Baicker et al. (2015) highlight that sometimes people may anticipate a form of "behavioral hazard" where they might underuse valuable care when faced with cost-sharing and might then prefer plans with low cost-sharing.

Although moral hazard can affect the optimal *level* of cost-sharing, it is unclear whether it could rationalize the variation in plan design I observe. The literature documenting consumer's behavioral response to cost-sharing typically leverage the variation on the margin of actuarial value, that is, the change in medical utilization when facing a larger or smaller expected level of coverage (Einav et al. 2013). It leaves open the question of how individuals respond dynamically to the non-linearity of the plan schedule, holding fixed the AV level. Second, to rationalize the variation in plan design, we need not only the existence of such a phenomenon but also heterogeneity among the population along the dimension of moral/behavioral hazard. If people are uniformly likely to reduce valuable care, and they should obtain the same level of valuable care, then we should expect to see a convergence of plan designs. There is limited research on the prevalence and heterogeneity of people's moral/behavioral hazard response. There also is no definite characterization of the optimal level of care people should get.

To shed light on this issue, I follow the model developed by Einav et al. (2013) and simulate the financial values of plans if individuals have a range of moral hazard responses. The main change is that individuals now will decide how many medical services to use after observing a negative health draw. Let λ denote the negative health shock, measured as the medical expenditure when individuals face no insurance. Their utility under plan j when they spend m in health care given the negative health draw λ, moral hazard level ω is:

$$\mathrm{u}(m;\lambda,\omega,\mathrm{j}) = \left[(m-\lambda) - \frac{1}{2\omega}(m-\lambda)^2\right] + \left[\mathrm{y} - \mathrm{p_j} - \mathrm{c_j}(m)\right],$$

where the item in the first bracket represents the utility from extra health (referred to as the health component henceforth) and the second bracket indicates the monetary utility (referred to as cost component henceforth). The moral hazard level, ω, is the extra medical expenditure one will consume if one goes from no insurance to full insurance. The larger the ω, the higher the moral hazard response. In the extreme case where ω is zero, the second term in the health component goes to negative infinity, so individuals will set $m = \lambda$, under which case the model reduces to the no-moral hazard version. The cost component consists of income (y), the premium of plan j ($\mathrm{p_j}$), and the out-of-pocket costs. ($\mathrm{c_j}(m)$ represents the cost-sharing rule of plan j. It specifies when the medical expenditure is m, how much the out-of-cost will be.) In this model, the value of a plan comes from both the health and cost component.

Individuals choose m to maximize (1). The FOC implies:

$$1 - \frac{\mathrm{m}^* - \lambda}{\mathrm{w}} = \frac{\mathrm{dc_j}\,(\mathrm{m}^*)}{\mathrm{dm}}$$

The solution depends on the specification of c (m). To simplify things, I transform each health plan in my sample into an AV-equivalent simplified three-arm design, with a deductible d, a MOOP x, and a coinsurance rate of c:

$$\mathrm{c(m)} = \begin{cases} m, if\ m \le d \\ d + c(m-d), if\ m > d\ and\ d + c(m-d) \le \mathrm{x} \\ x, otherwise \end{cases}$$

In this case, there are four candidates of m*: $0, \lambda, \lambda + (1 - c)\omega$, and $\lambda + \omega$. I verify which solution satisfies the c (m) schedule and pick the one gives the highest utility. After getting the utility of each expenditure level, I then calculate the risk-adjusted expected utility and its certainty equivalent using the CARA utility framework as before. This comparison ignores plan differences in premiums: theoretically, different plans might induce different overutilization and hence have different premiums. On the other hand, results in section 3.3 suggest that there is no correlation between risk premium and premium, suggesting that at least empirically, the premium difference does not matter in this context.

A prediction from the model is that consumers will always have the same or higher evaluation of a plan when they have higher moral hazard - they can always not over utilize to realize the same utility as before. Even though overutilization exposes them to higher monetary risks, they also get extra utility from the increase in health component, which offsets the monetary costs. The value of plan then depends on the relative role of these two off-setting forces. Naturally, plans with higher generosity will induce more overutilization and have a higher value to consumers. But what is unclear is the comparison for plans within the same metal tier. Theoretically, plans with a higher deductible and lower MOOP will induce more moral hazard when expenditure is high, and the low-deductible-high-MOOP will induce more expenditure in the deductible and coinsurance range. The overall effect of moral hazard depends on the relevant importance of the "marginal cost" at each arm. It is not clear beforehand that which plan design will induce more overutilization, and which one will have a higher value.[16]

The model in Einav et al. (2013) is trackable given the dataset I have: I assume λ follows the AV Calculator distribution I used in the benchmark analysis. As before, I consider a case where all plans have the same premium within a metal tier, so y and p_j can be abstract away as before. To get a reasonable range of variation in moral hazard responses, I used the estimates from Einav et al. (2013), which is from a large employer-sponsored health insurance menu. Their 10th, 25th, 50th, 75th and 90th percentile of ω is \$0, \$86, \$410, \$1,126 and \$3,236 respectively. Note that the expected medical expenses for the AV Calculator distribution are around \$5,600 in 2017. So a

[16] Drèze and Schokkeart (2013) shows that if the elasticity of medical expenditure with regard to insurance coverage is constant across states, the Arrow deductible-only plan will still be the best. Their conclusion does not apply here, because in my model the elasticity is not constant. (Though empirically the deductible-only plan is still the best for the range of moral hazard level I considered.)

$3,236 moral hazard response is equivalent to a 60% overutilization, which I interpret as a high level.

It turns out with the range of moral hazard responses considered, the rankings of plan values are quite stable, especially for smaller moral hazard response. Figure 11 shows the certainty equivalent of each plan with regard to the ω level. Plans are sorted by the certainty equivalent when there is no moral hazard. Under all moral hazard response levels, the straight-deductible plan is the plan with the best value. The variation in plan values increases when the moral hazard responses get higher. When moral hazard level is as high as $3,236, the range of plan values is over $2,000. This suggests that incorporating moral hazard might exaggerate the variation in plan values.

4.4 Heterogeneity in Consumer Risk Types

Within the classic expected utility framework, the risk premium measure is a function of the ex-ante medical spending distribution. Theoretically, consumers with different ex-ante risks may prefer different plan designs.[17] If low-risk consumers prefer a type of plan design, while high-risk consumers prefer another plan type, then the existence of a wide range of plans might be rationalized by heterogeneity in consumer types. It is also unclear how the range of plan values will change when consumers get healthier or unhealthier. It is possible that the economic consequences of large plan variation could be more substantial for certain populations.

To get a wider range of medical distribution, I use the 2014-2015 wave of Medical Expenditure Panel Survey (MEPS) data. MEPS is a nationally representative survey of medical spending and demographic information. I use the years 2014 and 2015, the most recent years after the implementation of the ACA, and inflated the spending number to the 2017 equivalent using the Personal Health Care Price Indices.[18] Since the sample size of individuals purchasing through the federal exchange is small, I restrict my sample instead to any individual who purchased private insurance (through the Federal and State Exchanges, or employer-sponsored plans) for the whole

[17] Arrow (1965) illustrates that straight-deductible is always the best plan within an AV level. However, the AV of a plan is calculated based on the Benchmark distribution, so when consumers have a different ex-ante risk, both the AV and the risk premium will change. This means plans in the same metal tier no longer have the same actuarial value, and hence non-straight-deductible plan might be preferred.
[18] This is the price index commonly used in the literature to adjust for health care spending over time. See https://meps.ahrq.gov/about_meps/Price_Index.shtml

period and who is under age 65. I then run an OLS regression of total medical expenditure for the year on a cubic polynomial in age, gender, and indicators for self-reported health status (5-categories), and the interactions of all of those variables.[19] I then classify individuals into eight equal-sized groups based on their predicted total medical spending. For each of these eight groups, I calculate the distribution of realized medical spending. That is, I generate eight representative individuals with different ex-ante medical spending distributions (which match the distribution of ex-post realizations for the MEPS respondents in that group).[20] After getting the alternative distributions, I then calculate the AV and risk premium for each plan using each total expenditure distribution.

Considering a broader range of ex-ante health risk introduces two sources of variation that could, in theory, cause people with different risk levels to prefer different plan designs. First, while plans in the same tier are designed to have the same actuarial value for the representative population in the AV Calculator Continuation table, they will not all present the same expected coverage to people with different risk profiles. For example, people with low expected risk will get fairly little coverage in most Exchange plans because a lot of their care will come under the deductible. It is possible that the expected coverage they receive will be higher in plan designs with lower deductibles (and hence higher maximum out of pocket limits). The reverse dynamics would hold for very high-risk individuals. The second source of variation is that plan designs interact with the distribution of medical spending needs and it is possible that some plan designs result in lower risk for the same expected coverage for some ex-ante risk types but not others. To summarize both variations, in this section I look at variation in certainty equivalent, which represents the sure amount of transfer equivalent as having that health insurance plan. The closer the certainty equivalent is to zero, the more financial protection a plan provides.

It turns out that most plan designs could not be rationalized with a range of distributions. One way to quantify "rationalize" is whether the certainty equivalent of a plan is within 1 standard deviation away from the best plan. For each plan, I count the number of MEPS distributions under which that plan could be rationalized. Figure 12 Panel A shows that about 90% of plans cannot be

[19] MEPS respondents evaluate their health status according to 5 categories: excellent, very good, good fair, and poor.
[20] For this analyses I use the sample weights provided in the MEPS in the regression predicting total spending. I also use the weights to calculate discrete probabilities of different total spending levels (84 atoms), consistent with the atoms in the CMS continuance tables used in the primary analyses.

rationalized by any distribution. So even if we consider a wide range of distributions, many plan designs still have poor value.

Different risk types also face different consequences of the variation in risk protection. Figure 12 Panel B plots the certainty equivalent of plans under each distribution. Plans are ranked by the certainty equivalent under the Benchmark distribution, and the ranking is shown on the x-axis, with the worst plan on the left and the best plan on the right. The y-axis shows the certainty equivalent under each distribution. Two interesting facts emerge from the graph. First, the variation in the financial protection provided by plans is much larger for the riskier types than the less risky types. For example, the difference in certainty equivalent for the best and worst plans when evaluated using the lowest risk MEPS distribution, is around $450, while the difference is around $2000 for the highest risk MEPS distribution. This suggests that the variation in plan designs could create welfare distortion among different segments of populations. Second, for individuals with above the median expenditure risks, for whom the variation matters most, the ranking of plans does not change much. The correlation of the certainty equivalent evaluated using the median risk and the highest risk is 89.57%.

This raises an interesting aspect of risk premium: theoretically, insurers could use higher risk premium plans as a tool to attract healthy consumers, because the risk premium difference is much smaller for healthy individuals than for unhealthy individuals. If consumers respond to the plan design optimally, we should expect to see that healthy consumers are enrolled in plans with higher risk premium, while the unhealthy consumers are enrolled in plans with lower risk premiums. To examine this, I combined information from the 2014 – 2016 Uniform Rate Review Data. About 60% Exchange plans report the average allowed claim costs, which is the amount paid by insurer and consumers to the hospitals, per member per month. I group plans by the quartile of their average claim costs. Group 0 represents plans whose claim costs data is missing. The average allowed claim costs per member month in each quartile are $69, $207, $306 and $637 for the Bronze plans, and $232, $544, $771 and $1535 for the Gold plans. I then calculate the average relative risk premium for plans in each metal tier and claim costs quartile. Figure 13 shows the results. For both metal tiers, consumers in the higher claim costs quartile are enrolled in plans with lower relative risk premium. The difference between the 1st and last quartile is about $50 for the Bronze plans, and $80 for the Gold plans. This suggests a potential selection within metal tier by plan design.

5. Discussion

In this paper, I systematically document the wide range of financial protection in the insurance plans offered on the ACA Exchanges. For plans in the same metal tier, large variation exists in the implied out-of-pocket spending distribution variance, and this transfers to a range of risk premium of more than $1000 for Silver plans. The economic implications for such variation are important, not only because of the size of variation but also because individuals enroll in all types of plans.

Several possible explanations for the variation in the financial protection of health insurance plans are tested and shows to be inconsistent with the data. First, I investigate whether temporary market experimentation explains the variation and I find that it is not, as the variation is stable over time. Second, I study geographic variation in healthcare markets and find that it also does not explain the variation. Third, I analyze and find that heterogeneity in moral hazard responses can not easily rationalize the variation in plan values. The heterogeneity in risk types may explain partially the variation in plan types. I find evidence that the average allowed expenditure is significantly higher for straight deductible plans, consistent with consumers sorting by ex-ante risk types. The sorting within metal tier may have implication for risk transfer programs. Future research on the interaction between AV regulation and risk adjustment is needed to better understand the issue.

An open question is whether the variation in plans' financial protection is consistent with other heterogeneity in consumer preferences that are not accounted for in the classic expected utility framework, namely liquidity constraints, and consumer confusion. Liquidity constraints might be a motivation why people may prefer high-risk premium plans (Ericson and Sydnor 2018). The intuition is that when people are liquidity constrained, they may avoid choosing a plan with lumpy deductible shocks, i.e. the deductible plan. But under the classic framework, the low deductible plan will have a higher risk premium. While theoretically possible, the relevance of liquidity constraints as an explanation for the variety of Exchange plan designs is unclear. Even though a large share of enrollees are in the lower income population, they are highly subsidized through premium credits and cost-sharing reductions, which would likely counteract some of the importance of liquidity constraints. On the other hand, there is much less discussion on the optimal design of the cost-sharing variation plans with regard to liquidity constraints. This is an area that requires future research.

The final, more subtle point is that even though the above-mentioned preferences exist, how effectively people sort into the plans that are best for them is in question. The literature documents that people are unable to sort into the “right plan” for them in various health insurance markets (Abaluck and Gruber 2011: Medicare Part D. Bhargava et al 2017: employer-sponsored plan.) The individual market is no less likely to generate consumer confusion, given the large choice set and the huge variation in plan designs that an individual usually faces. The design of the shopping website, where it sorts plan by either premium or “estimated expected total costs,” may also emphasize certain aspects of a plan, like expected spending, while ignoring the other attributes like the variance of spending, moral or behavioral hazard, etc. Distinguishing between heterogeneity in preferences and consumer confusion has important welfare implications: if preference heterogeneity drives the pattern, then the existence of the plan design variation might be optimal. On the other hand, if people largely sort into a random plan, and insurers respond by randomly offering plan designs, or even offer plans that could exploit consumer confusion, regulation like plan standardization and, the educational program might be helpful.

References

Abaluck, Jason, and Jonathan Gruber. “Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program.” *The American Economic Review* 101, no. 4 (June 2011): 1180–1210.

Aizawa, Naoki, and Hanming Fang. “Equilibrium Labor Market Search and Health Insurance Reform.” Working Paper. National Bureau of Economic Research, January 2013.

Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen. “Moral Hazard in Health Insurance: Do Dynamic Incentives Matter?” *Review of Economics and Statistics* 97, no. 4 (October 2015): 725–41.

Arrow, Kenneth J. “Uncertainty and the Welfare Economics of Medical Care.” *The American Economic Review* 53, no. 5 (1963): 941–73.

Avery, Kelsey, Mathias Gardner, Emily Gee, Elena Marchetti-Bowick, Audrey McDowell, and Aditi Sen. “Health Plan Choice and Premiums in the 2017 Health Insurance Marketplace” *ASPE Research Brief* (2016).

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. “Behavioral Hazard in Health Insurance.” *The Quarterly Journal of Economics* 130, no. 4 (November 1, 2015): 1623–67.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. “Choose to Lose: Health Plan Choices from a Menu with Dominated Option.” *The Quarterly Journal of Economics* 132, no. 3 (August 1, 2017): 1319–72.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. “What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics.” *The Quarterly Journal of Economics*.

Dafny, Leemore, Jonathan Gruber, and Christopher Ody. “More Insurers Lower Premiums: Evidence from Initial Pricing in the Health Insurance Marketplaces.” *American Journal of Health Economics*, 2015.

Drèze, Jacques H., and Erik Schokkaert. “Arrow’s Theorem of the Deductible: Moral Hazard and Stop-Loss in Health Insurance.” *Journal of Risk and Uncertainty* 47, no. 2 (October 1, 2013): 147–63.

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen. “Selection on Moral Hazard in Health Insurance.” *The American Economic Review* 103, no. 1 (February 2013): 178–219.

Ericson, Keith M., and Amanda Starc. “How Product Standardization Affects Choice: Evidence from the Massachusetts Health Insurance Exchange.” *Journal of Health Economics* 50 (December 2016): 71–85.

Ericson, Keith M. and Justin Sydnor “Liquidity Constraints and Insurance Demand” Working Paper 2018

Geruso, Michael, Timothy J. Layton, and Daniel Prinz. “Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges.” NBER Working Paper 22832, November 2016.

Ghili, Soheil. “Network Formation and Bargaining in Vertical Markets: The Case of Narrow Networks in Health Insurance,” Working Paper, 2016.

Handel, Benjamin R. “Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts.” *American Economic Review* 103, no. 7 (December 2013): 2643–82.

Handel, Benjamin R., and Jonathan T. Kolstad. “Health Insurance for ‘Humans’: Information Frictions, Plan Choice, and Consumer Welfare.” *American Economic Review* 105, no. 8 (August 2015): 2449–2500.

Loewenstein, George, Joelle Y. Friedman, Barbara McGill, Sarah Ahmad, Suzanne Linck, Stacey Sinkula, John Beshears, et al. “Consumers’ Misunderstanding of Health Insurance.” *Journal of Health Economics* 32, no. 5 (2013): 850–862.

Mitchell, Dan. “McDonald’s health insurance plans: Not lovin' it.” *Fortune.* October 5, 2010. Accessed on May 10, 2018. http://archive.fortune.com/2010/10/05/news/companies/mini_meds_insurance_mcdonalds.fortune/index.htm.

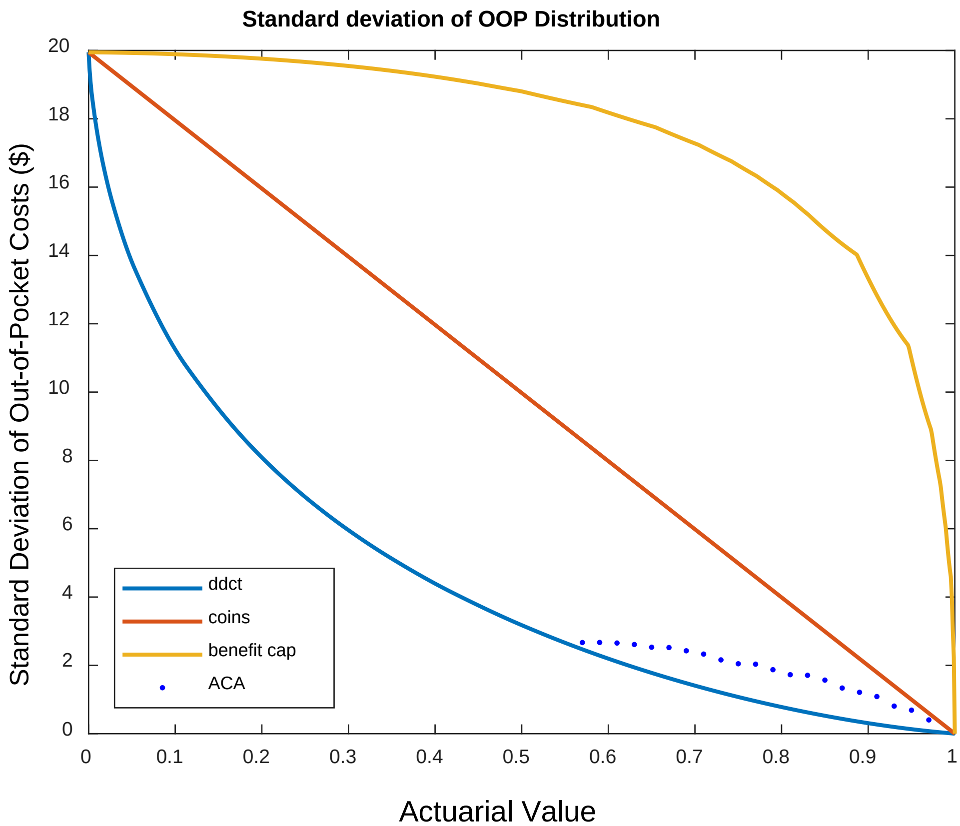
Pauly, Mark V. “The Economics of Moral Hazard: Comment.” *The American Economic Review* 58, no. 3 (1968): 531–537.

Starc, Amanda, and Robert J. Town. “Internalizing Behavioral Externalities: Benefit Integration in Health Insurance.” NBER Working Paper 21783, 2015.

Tebaldi, Pietro. “Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA.” Available at SSRN 2688894, 2015.

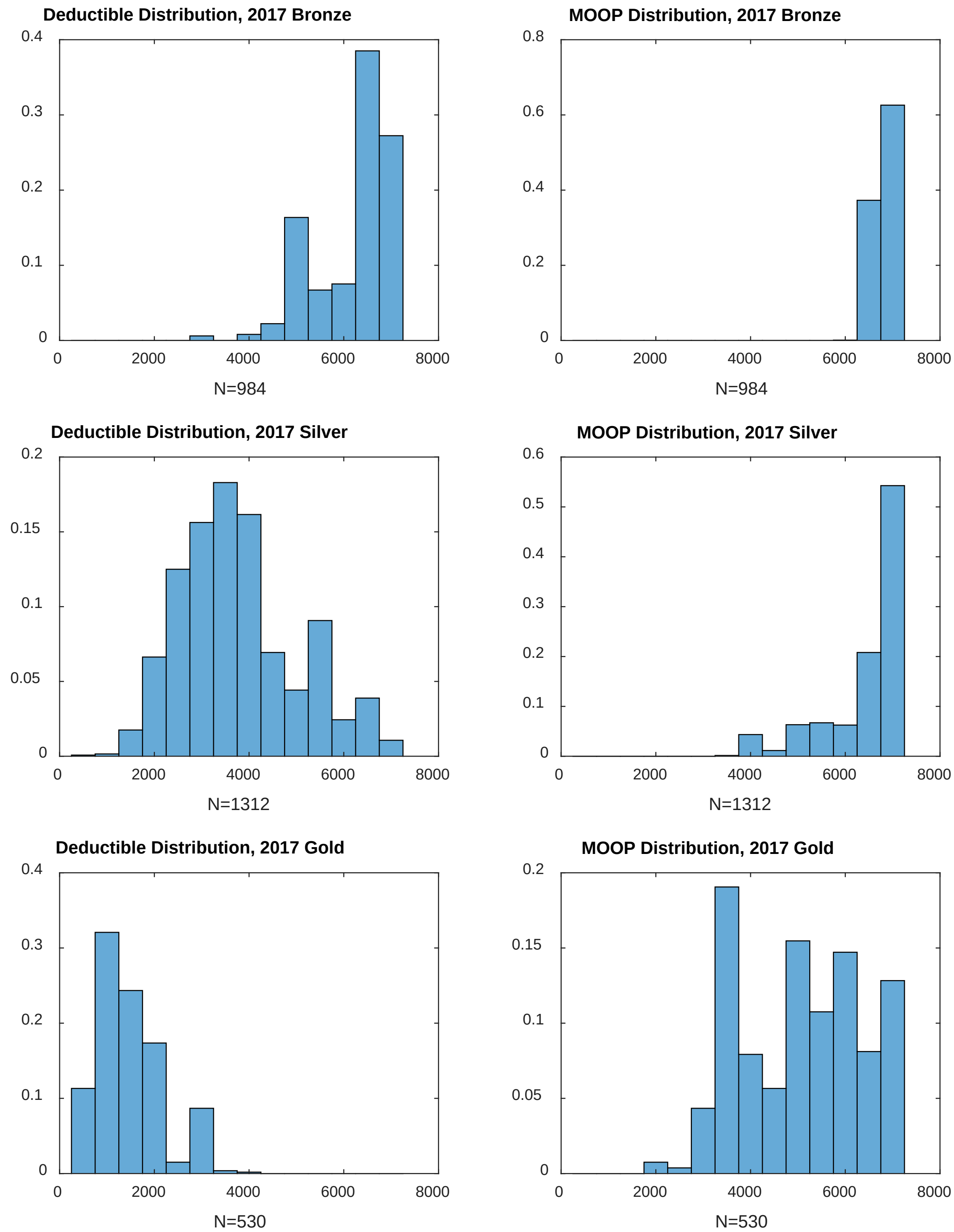
Figures and Tables

Figure 1. Financial values of simulated contract types



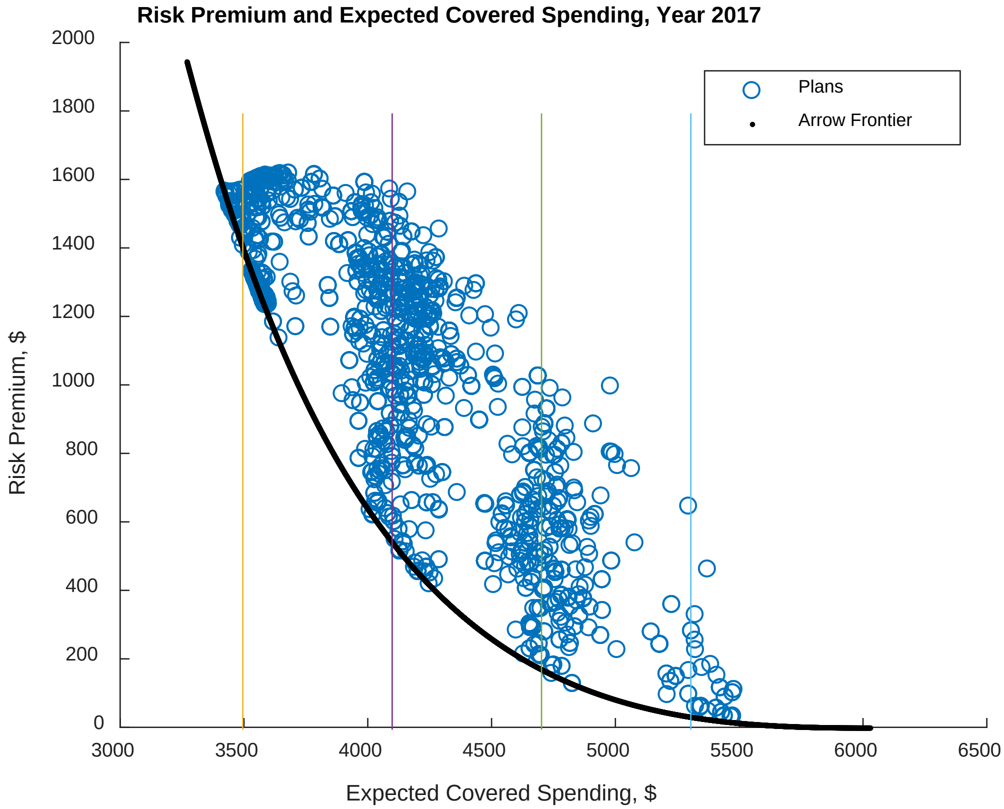
Note: The yellow line represents plan with full coverage before the benefit cap, and no coverage after the cap. Red line represents plan with fixed coinsurance rate. The solid blue line represents plan with no coverage before the deductible, full coverage after the deductible. For each contract type, I calculate the standard deviation of the out-of-pocket spending at each actuarial value, using 2017 CMS AV Calculator Gold Distribution. The area between the dashed blue line and the solid blue line are the contract space consistent with ACA regulation.

Figure 2. Distribution of deductible and MOOP by metal tier for 2017 Plans



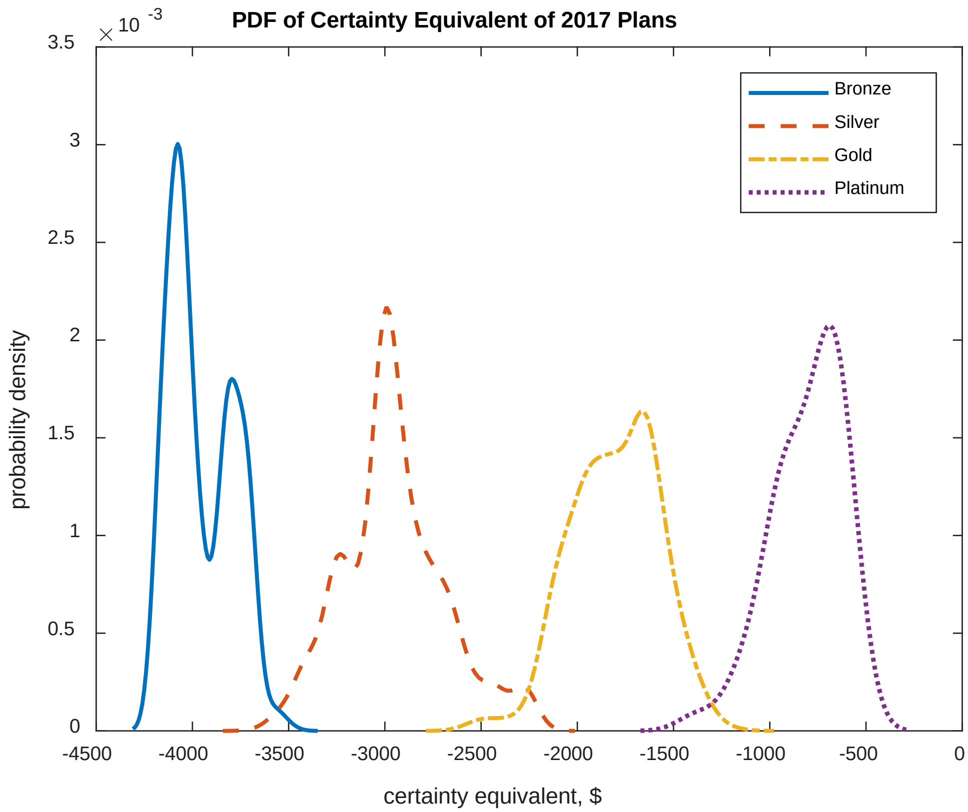
Note: Data from 2017 CMS Health Insurance Exchange Public Use Files. The sample includes all exchange qualified health plans offered to individuals through the Health Insurance Exchange. For Silver tier, sample excludes cost-sharing variations.

Figure 3. Risk premium and expected covered spending for 2017 plans



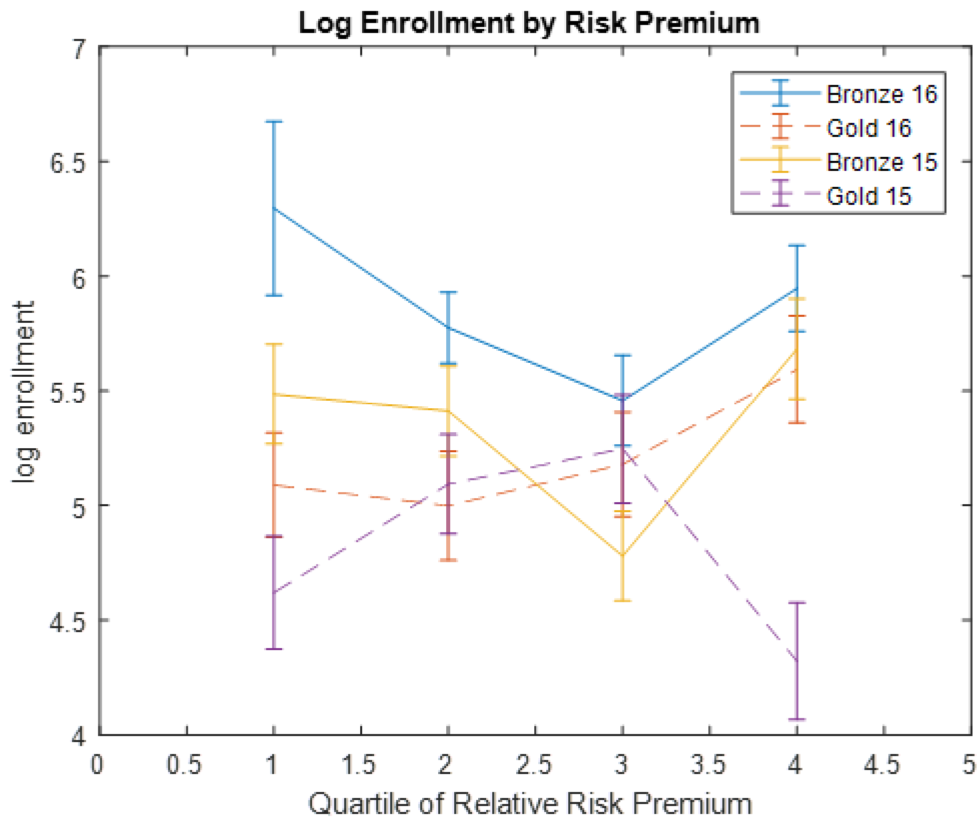
Note: Data from 2017 CMS Health Insurance Exchange Public Use Files. The sample includes all exchange qualified health plans offered to individuals through Health Insurance Exchange. For Silver tier, sample excludes cost-sharing variations. Risk premium and expected covered spending is calculated using 2017 CMS AV Calculator Distribution. Each dot represents one unique plan. The darker the area the more plans with that specific combination of expected covered spending and risk premium. Arrow frontier shows the lowest possible risk premium conditional on expected covered spending level, and are not actual plans. The vertical lines show the targeted actuarial value. Note that not all plans line up with the vertical lines perfectly, partially because the regulator allows for a 2 percentage point error margin, and partially because of measurement error in my calculation.

Figure 4 Probability density function of certainty equivalent of 2017 plans



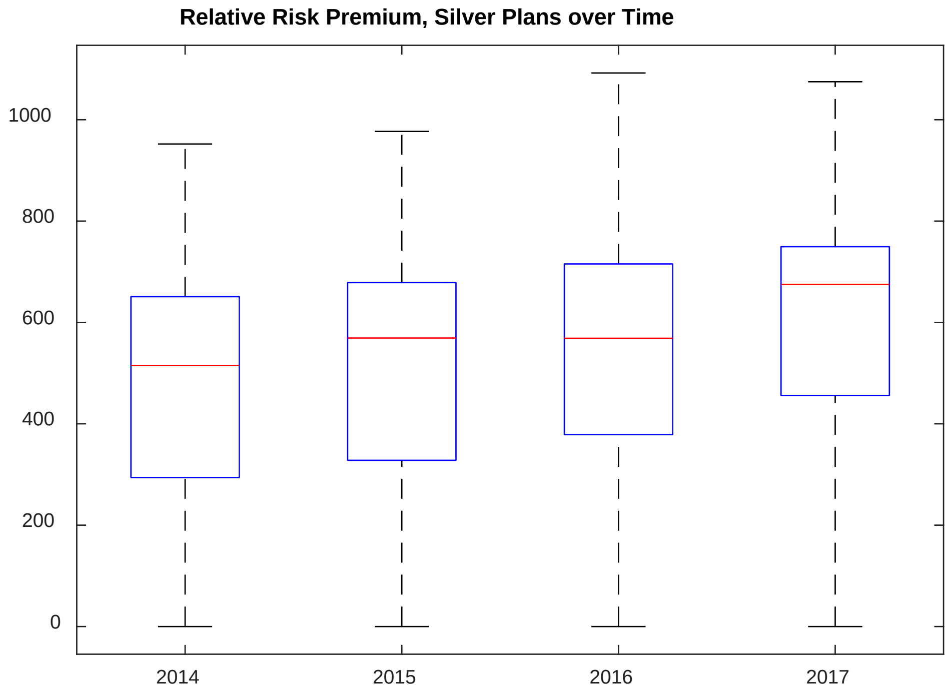
Note: Data from 2017 CMS Health Insurance Exchange Public Use Files. The sample includes all exchange qualified health plans offered to individuals through Health Insurance Exchange. For Silver tier, sample excludes cost-sharing variations. Certainty equivalent is calculated using 2017 CMS AV Calculator Distribution. PDF estimated using the Kernel smooth function.

Figure 5 Enrollment in Plans with Various Risk Premium Level



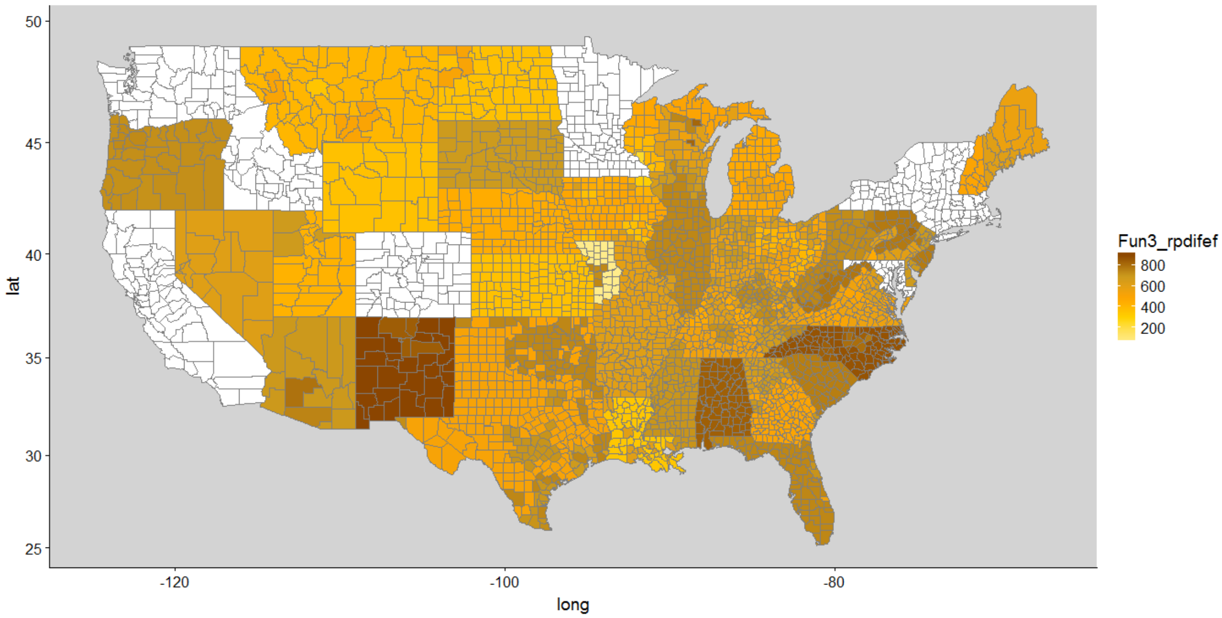
Note: For Gold and Bronze Plans in 2015 and 2016, I calculate the relative risk premium as the difference between each plan's risk premium and the smallest risk premium at that expected covered spending level, using the 2017 AV Calculator Distribution. I then group each year's plans into quartiles based on the relative risk premium. The point represents the mean of log enrollment number for all plans in that group, while the error bar represents the 95% confidence interval of the mean.

Figure 6 Distribution of risk premium over time



Note: Relative risk premium is the difference between each plan's risk premium and the smallest risk premium at that expected covered spending level. The box plot shows the minimum, maximum (black lines), the 25th and 75th percentile (the blue lines) and the median (the red line) of the relative risk premium for each year's Silver plans. There are 923 standard Silver plans in 2014, 1645 in 2015, 1441 in 2016 and 1312 in 2017.

Figure 7 Map of Mean Relative risk premium of 2017 Silver plans in each county



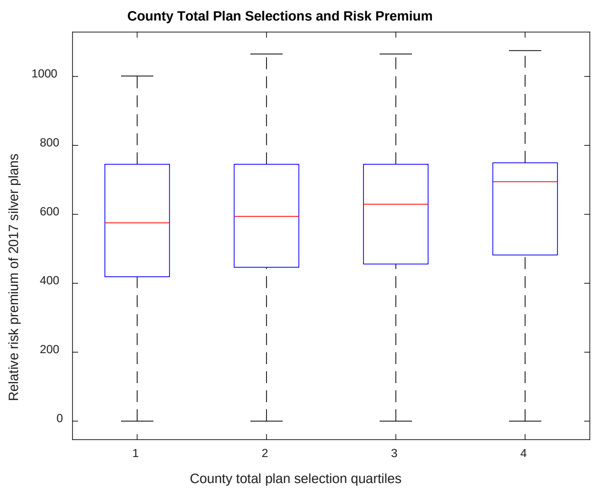
Note: Relative risk premium is the difference between each plan's risk premium and the smallest risk premium at that expected covered spending level. Blank states are not in the dataset. Borderlines represent county borders.

Figure 8 Median and standard deviation of relative risk premium of 2016 Silver plans

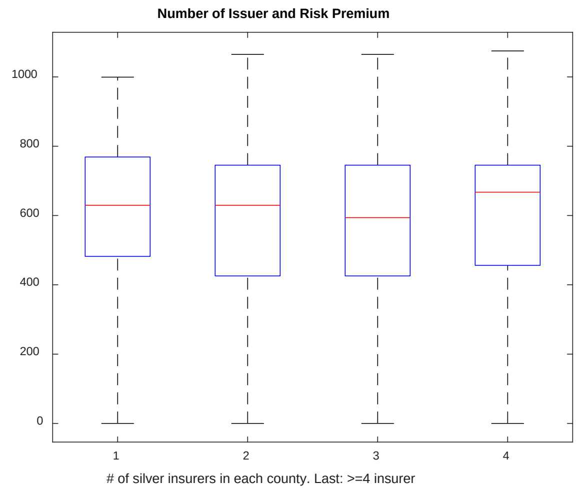
Note: For each rating area, I calculated the mean and standard deviation of relative risk premium, which is the difference between each plan's risk premium and the smallest risk premium at that expected covered spending level. I then rank each county by mean of the relative risk premium, plot the ranking on the x-axis, and plot both the mean and standard deviation on the y-axis.

Figure 9 Risk premium and county characteristics

Panel A: County Total Plan Selections



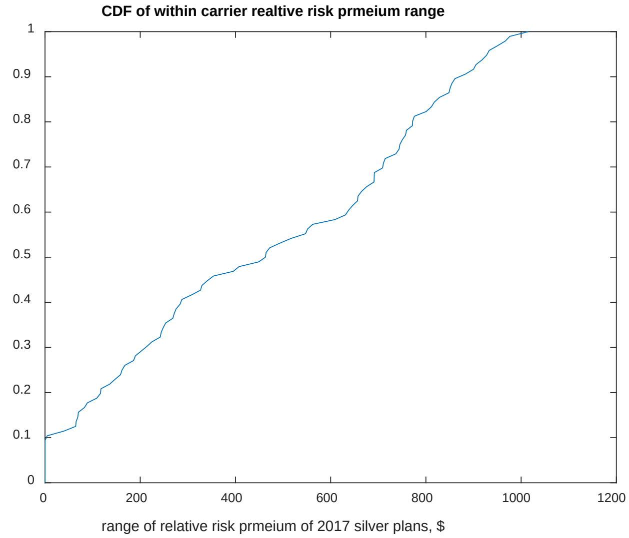
Panel B: Number of Insurers



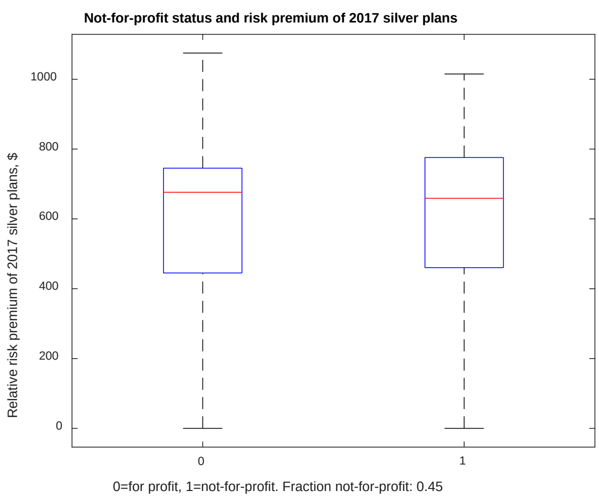
Notes: For all plans in the 2017 Silver tier, I construct a new dataset where each observation represents one county-plan. I then classify plans based on their counties' market size, measured as total plan selections, and competition intensity, measured by the number of insurers. I plot the relative risk premium by quartiles of county market size, and by the number of insurers. Since less than 5% of counties have more than 4 insurers, I group these counties together.

Figure 10 Risk Premium Variation among Insurers

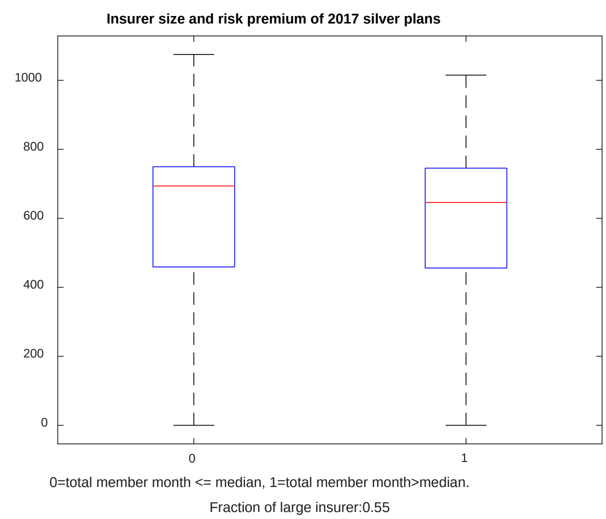
Panel A: Distribution of within insurer relative risk premium range of 2017 Silver plans



Panel B: Not-for-profit status

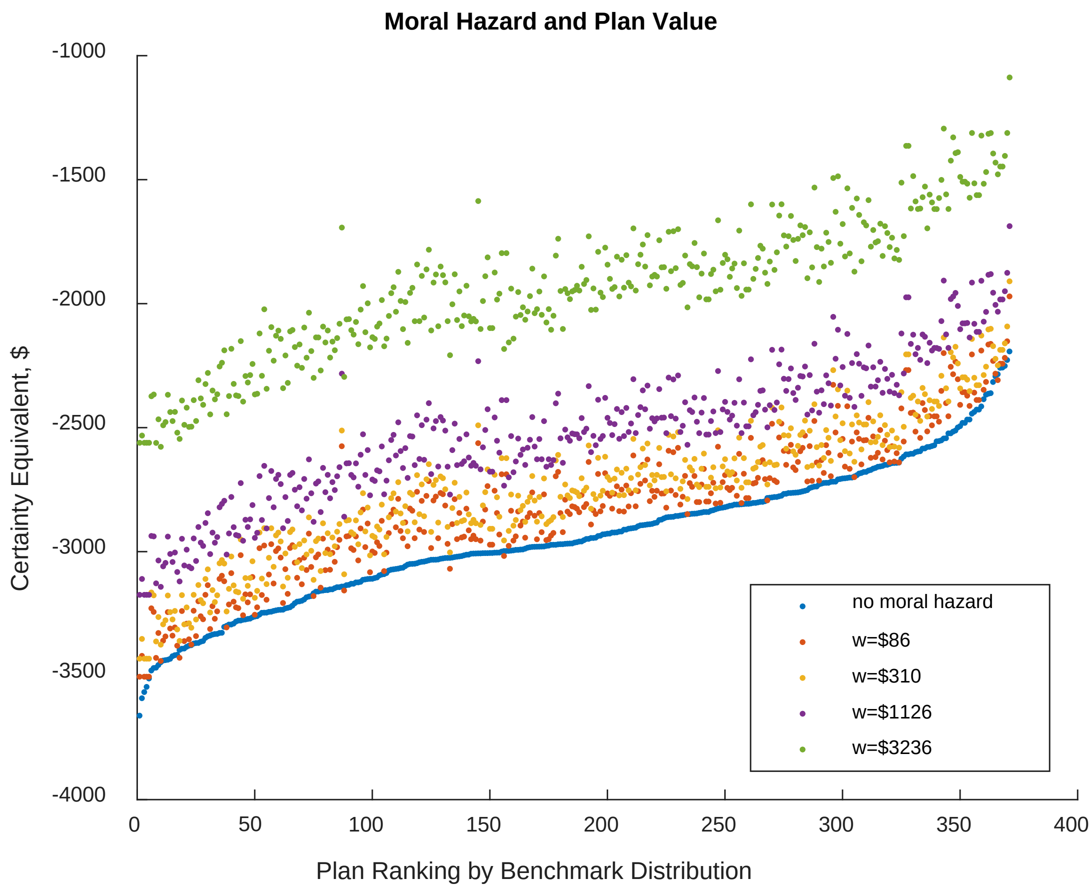


Panel C: insurer size measured by member-months



Notes: the boxplot classifies plans based on its insurer's status. Not-for-profit status classifies insurers based on whether they are for-profit (0) or not-for-profit (1). Insurer size is based on the total member month underwritten in the 2015 individual market, where 1 represents insurers with the member months larger than the median. Data from the 2015 Medical Loss Ratio filing.

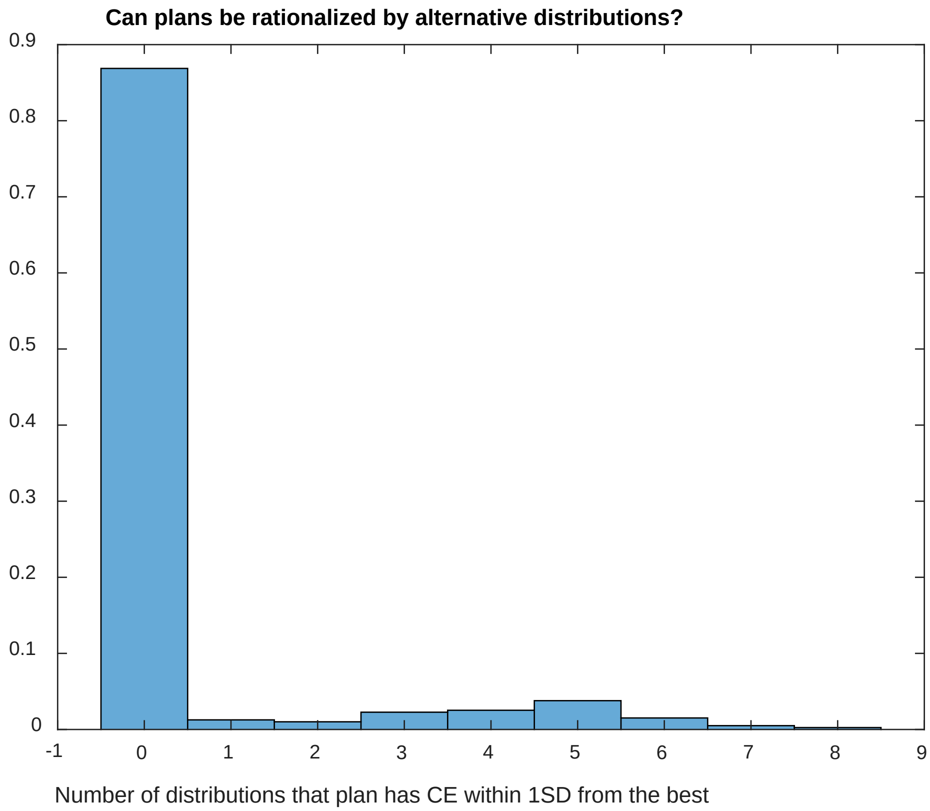
Figure 11 Plan Values under moral hazard responses



Notes: Plans are sorted by its certainty equivalent under the benchmark distribution. The x-axis shows the ranking, where 1 represents the worst plan. Each dotted line corresponds to the 25th, 50th, 75th and the 90th percentile of the moral hazard responses level, w, estimated in Einav et al. (2013). The moral hazard level, w, is the extra medical expenditure if one goes from no insurance to full insurance.

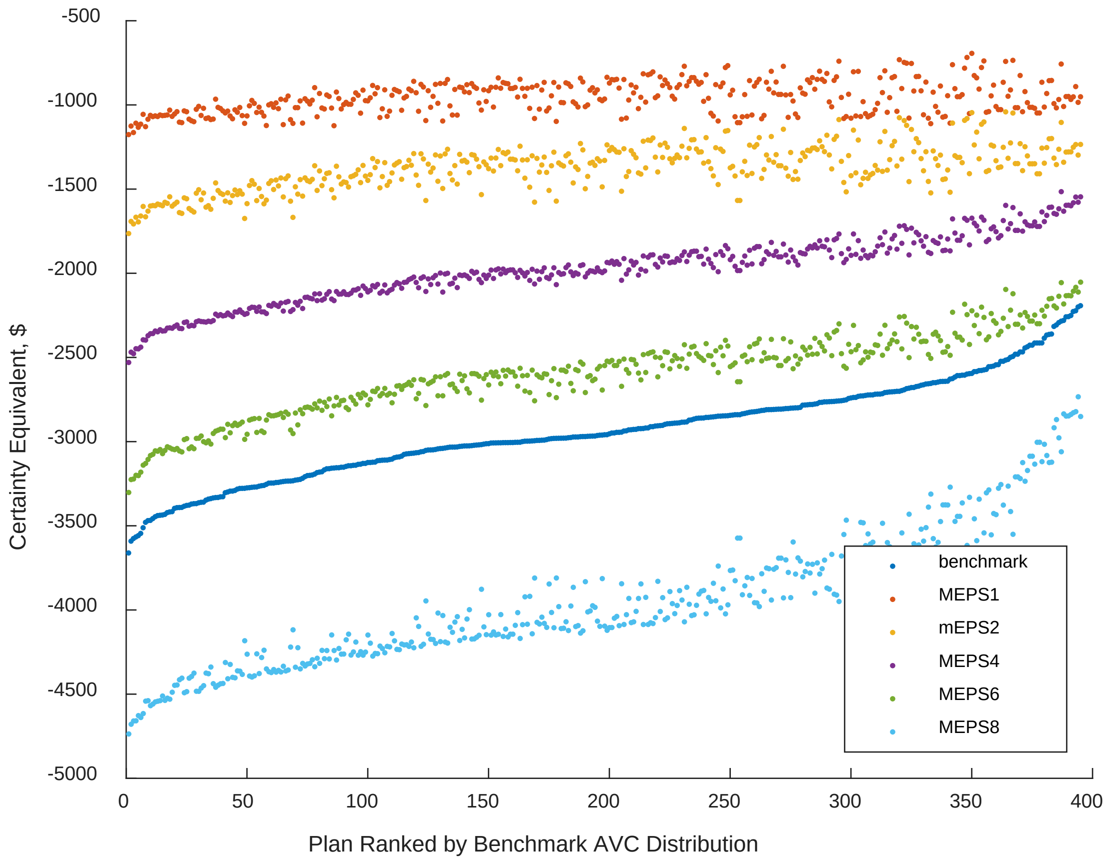
Figure 12 Plan values under alternative medical spending distributions

Panel A



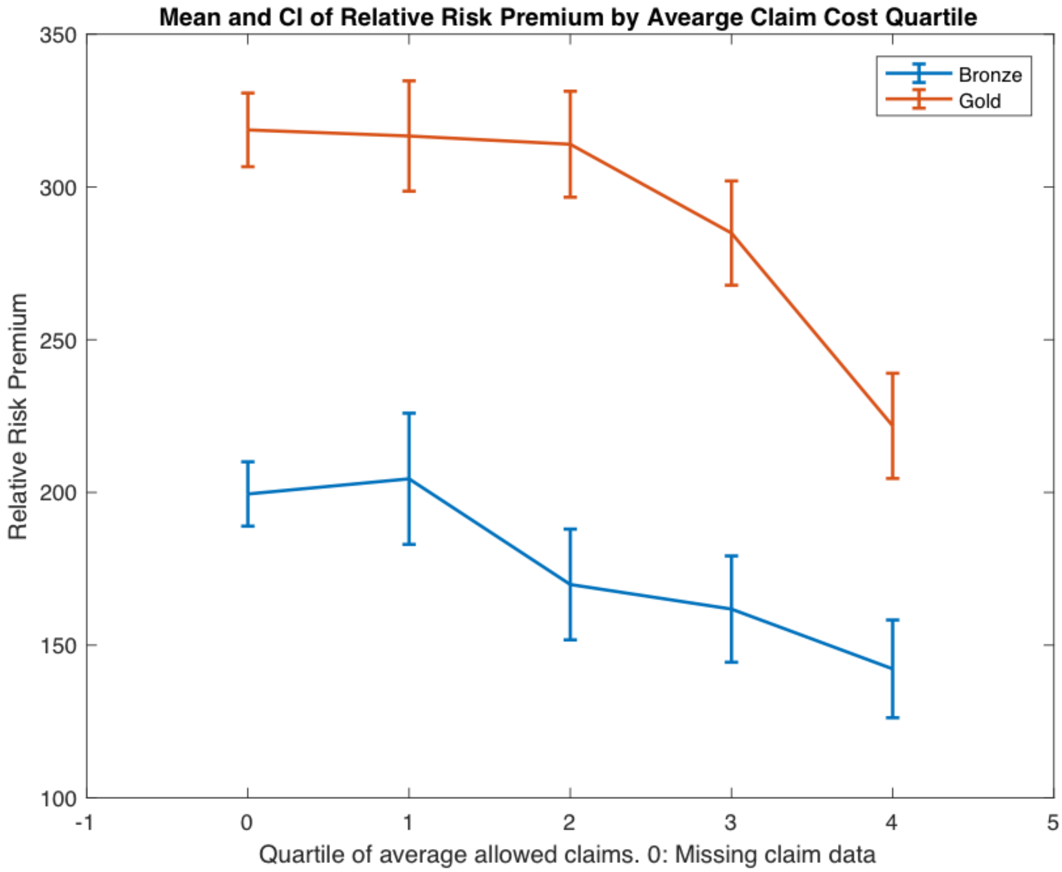
Notes: For each plan, I calculate under how many MEPS distributions the plan could be rationalized. Being rationalized is defined as having a certainty equivalent within 1 standard deviation of the best plan. The graph plots the distribution of all 2017 silver plans.

Panel B



Notes: Plans are sorted by its certainty equivalent under the benchmark distribution. The x-axis shows the ranking, where 1 represents the worst plan. Each line represents one MEPS distribution, and the y-axis shows the certainty equivalent value.

Figure 13 Relative Risk Premium and Average Claim Costs



Notes: The sample is all Bronze and Gold plans in year 2014-2016. Average allowed claims per member month is the sum of the claim payment from both consumers and insurers to the provider. Group 0 represents plans missing the claim information.

Table 1 Enrollment-weighted relative risk premium by metal tier

Metal Tier	Enrollment Share	Relative Risk Premium		
		Mean	Median	90th – 10th percentile
Catastrophic	0.82%	256.91	300.71	311.21
Bronze	20.73%	173.83	96.77	481.74
Silver - standard	12.49%	597.39	664.97	828.10
Silver - 73%	8.42%	366.22	383.24	367.80
Silver - 87%	18.34%	55.87	56.52	79.84
Silver - 94%	32.30%	18.36	8.70	69.68
Gold	6.16%	378.14	320.15	568.03
Platinum	0.74%	99.49	73.35	215.37
Total	100%	183.79	58.87	600.94

Notes: The table shows the mean, median and the difference between the 90th and 10th percentile of the relative risk premium of 2016 plans by metal tier. Numbers are weighted by plan's enrollment share to reflect the value chosen by consumers.