Should There Be Vertical Choice in Health Insurance Markets?*

Victoria R. Marone§ Adrienne Sabety†

July 2020

Click here for most recent version

Abstract

Choice over coverage levels—"vertical choice"—is widely available in U.S. health insurance markets, but there is limited evidence of its effect on welfare. For a given consumer, the socially efficient level of coverage trades off the value of risk protection and the social cost from moral hazard. Providing choice does not necessarily lead consumers to select their efficient coverage level. We show that in regulated health insurance markets, vertical choice should be offered only if consumers with higher willingness to pay for insurance have a higher efficient level of coverage. We test for this condition empirically using administrative data from a large employer. Our estimates imply substantial heterogeneity in efficient coverage level, but we do not find that households with higher efficient coverage levels have higher willingness to pay. It is therefore optimal to offer only a single level of coverage. Relative to a status quo with vertical choice, mandating the optimal single level of coverage increases welfare by $330 per household per year.

Keywords: risk protection, moral hazard, health insurance
JEL Codes: D82, G22, H2, I13

*We are grateful to Vivek Bhattacharya, David Cutler, Leemore Dafny, David Dranove, Liran Einav, Sebastian Fleitas, Tal Gross, Igal Hendel, Gaston Illanes, Matthew Leisten, Matt Notowidigdo, Chris Ody, Rob Porter, Elena Prager, Mar Reguant, Bill Rogerson, Amanda Starc, and Gabriel Ziegler for excellent advice and suggestions that benefited this research. We also thank seminar participants at the 8th Conference of the American Society of Health Economists, Northwestern, BI Norwegian Business School, University of Chicago, Princeton, MIT, Washington University, Yale School of Management, University of Rochester Simon, New York University, MIT Sloan, University of Texas at Austin, Chicago Booth, and Wisconsin for helpful comments. Finally, we thank Jason Abaluck and Jon Gruber for access to the data and for their support of this research project. §Department of Economics, Northwestern University. E-mail: marone@u.northwestern.edu. †Interfaculty Initiative in Health Policy, Harvard University. E-mail: asabety@g.harvard.edu.

I Introduction

Choice over the financial extent of coverage—which we term "vertical choice"—is widely available in U.S. health insurance markets.[1] A leading example is the tiered plans (e.g., Bronze, Silver, Gold) offered on Affordable Care Act exchanges. In contrast, national health insurance schemes typically offer a single level of coverage. Regulation plays a central role in determining the extent of vertical choice in health insurance markets, but the economics literature provides limited guidance to regulators on this topic. In this paper, we develop a theoretical and empirical framework for evaluating the welfare effects of vertical choice.

The basic argument in favor of vertical choice is the standard argument in favor of product variety: With more options, consumers can more closely match with their socially efficient product by revealed preference (Dixit and Stiglitz, 1977). This argument relies critically, however, on the condition that privately optimal choices align with socially optimal choices. In competitive markets in which costs are independent of consumers' private valuations, this alignment is standard. But in markets with selection, this alignment may not be possible. Health insurance markets are classic examples of selection markets. Costs are inextricably related to private valuations, and asymmetric information prevents prices from reflecting marginal costs (Akerlof, 1970; Rothschild and Stiglitz, 1976). We show that even if selection markets are competitive, regulated, and populated by informed consumers, whether choice can lead to a more efficient allocation is theoretically ambiguous.

Our welfare metric derives from the seminal literature on optimal insurance, which holds that the efficient level of coverage equates the marginal benefit of risk protection and the marginal social cost of utilization induced by insurance (Arrow, 1965; Pauly, 1968, 1974; Zeckhauser, 1970). We focus attention on the fact that this central tradeoff between the "value of risk protection" and the "social cost of moral hazard" plays out on a consumer-by-consumer basis, meaning that the efficient level of coverage likely varies across consumers. Socially optimal regulation aims to design plan menus such that consumers self-select into their efficient level of coverage. Private incentives are such that consumers with higher willingness to pay for insurance self-select into (weakly) higher levels of coverage. The problem is that consumers with higher willingness to pay do not necessarily also have a higher efficient level of coverage. It is precisely this statement that captures the theoretical

[1] Coverage level is determined by plan features such as deductibles and out-of-pocket maximums. Though currently widespread, vertical choice is a key point of differentiation among federal healthcare policy proposals. The "Medicare for all" proposal (endorsed by Bernie Sanders and Elizabeth Warren) would not feature vertical choice, while the plan to introduce a public option to existing exchanges (endorsed by Joe Biden) and the "American Health Care Act" (endorsed by Donald Trump) would continue to do so.

ambiguity of whether vertical choice is efficient.

We ask whether vertical choice should be offered from the perspective of a market regulator that can offer vertically differentiated plans and set premiums.[2] The regulator's objective is to maximize allocational efficiency of consumers to plans. As is standard in national health insurance schemes and employer-sponsored insurance, consumer premiums need not equal plan average cost.[3] If the regulator sets premiums such that more than one plan is demanded, we say it has offered vertical choice. Extending the widely used graphical framework of Einav, Finkelstein and Cullen (2010), we show that the key condition determining whether vertical choice should be offered is whether consumers with higher willingness to pay have a higher efficient level of coverage. The principal empirical focus of this paper is to determine whether this is likely to be true.

We begin by presenting a model of consumer demand for health insurance, building on Cardon and Hendel (2001) and Einav et al. (2013). The model has two stages. In the first, consumers make a discrete choice over plans under uncertainty about their health. In the second, upon realizing their health, consumers make a continuous choice of healthcare utilization. We use the model to show that willingness to pay for insurance can be partitioned into two parts: one that is both privately *and* socially relevant (the value of risk protection), and one that is only privately relevant (expected reduction in out-of-pocket spending). Because a portion of a consumer's private valuation of insurance is a transfer, it is not necessarily the case that higher willingness to pay implies higher social surplus. For example, a sick but risk-neutral person obtains a large private benefit from higher coverage, but generates no social benefit; the burden of her expected healthcare spending is simply shifted to others. If she consumes more healthcare than she values in response to higher coverage, the regulator would prefer she had lower coverage.

We estimate the model using data from the population of public school employees in Oregon. The data contain health insurance plan menus, plan choices, and the subsequent healthcare utilization of nearly 45,000 households between 2008 and 2013. Crucially for identification, we observe plausibly exogenous variation in the plan premiums and plan

[2]By market regulator, we mean the entity that administers a particular health insurance market. In employer-sponsored insurance, this is the employer; in Medicare, this is the Centers for Medicare and Medicaid Services; in Norway, this is the Norwegian government. The regulator can set premiums in a competitive market by strategically taxing or subsidizing plans, or it can just supply all plans itself.

[3]We depart from the standard competitive equilibria studied in health insurance markets (e.g., Rothschild and Stiglitz, 1976; Handel, Hendel and Whinston, 2015; Azevedo and Gottlieb, 2017) by removing price as an equilibrium object in order to render a larger set of allocations feasible. We find this desirable because it reflects realistic regulatory powers and focuses attention on the economic constraint of unobserved types.

options offered to employees. The variation is driven by the fact that plan menus are set independently by each of the 187 school districts in the state, where districts select plans from a common superset determined at the state level. In addition, employees are offered several different coverage levels by the same insurer with the same provider network, providing isolated variation along our focal dimension.

Our empirical model incorporates both observed and unobserved heterogeneity along three key dimensions of household type: health status, moral hazard, and risk aversion. We use the model to recover the joint distribution of household types in the population. We then construct each household's willingness to pay for different levels of coverage and the social surplus generated by allocating each household to different levels of coverage. We construct these objects for a set of coverage levels that span the range offered in our empirical setting and on Affordable Care Act exchanges. Each coverage level, or contract, is characterized by a deductible, a coinsurance rate, and an out-of-pocket maximum. The least generous coverage we consider is a "Catastrophic" contract with a deductible and out-of-pocket maximum of $10,000. The most generous coverage is full insurance.

We do not find that households with higher willingness to pay have a higher efficient level of coverage. High willingness to pay is primarily driven by high expected reduction in out-of-pocket spending, rather than by a high value of risk protection. While high willingness-to-pay households do tend to be more risk averse, they are so likely to hit any out-of-pocket maximum that they face little uncertainty. Households with low willingness to pay are more prone to moral hazard and are less risk averse, but also face more uncertainty over out-of-pocket costs. This negative correlation between willingness to pay and *risk* is a central driver of our results. Ultimately, because prices cannot efficiently screen consumers, we find that optimal regulation is to offer a single contract. This contract has an actuarial value of 84 percent.[4] Introducing any other contract leads to over- or under-insurance (on average) among households that would choose the alternative.

We benchmark welfare outcomes against the allocation of all households to the Catastrophic contract. The first-best allocation increases social surplus by $1,802 per household per year relative to this benchmark. Because households with the same willingness to pay can have different efficient levels of coverage, the first-best allocation cannot be sup-

[4] Actuarial value (AV) measures the percentage of a population's total healthcare spending that would be insured under a particular contract. Full insurance implies an AV of 100 percent. Our Catastrophic contract has an AV of 53 percent. We find that households' efficient coverage levels range between 65 percent AV and full insurance. For reference, 70 percent is the AV provided by Silver plans on Affordable Care Act exchanges.

ported unless premiums can vary by household type.[5] Under optimal regulation (the single contract), 30 percent of households are not allocated to their efficient coverage level. Nevertheless, we find that optimal regulation generates 96 percent of social surplus at the first-best allocation. Three factors contribute: (i) the value of risk protection is increasing in coverage level, but at a decreasing rate; (ii) the social cost of moral hazard is increasing in coverage level, and at an increasing rate; and (iii) at the optimal single contract, the magnitude of the value of risk protection is roughly six times as large as the social cost of moral hazard. As a result, in the neighborhood of the optimal single contract, the welfare stakes of misallocation are small.

We compare outcomes under several alternative policies, including competitive pricing and full vertical choice. Under competitive pricing, all contracts must break even. We find that in our population, the market unravels to the lowest level of coverage (Catastrophic) due to adverse selection. Though choice is permitted, the market cannot deliver it. Under full vertical choice, we implement subsidies to support an allocation in which all contracts are traded. Using subsidies designed to mimic enrollment shares observed on Affordable Care Act exchanges, this policy generates 78 percent of the surplus generated by the first-best allocation. Social surplus is $330 higher per household per year under optimal regulation than under full vertical choice, but these gains are not shared evenly. The highest willingness-to-pay households fare best under optimal regulation, while the lowest willingness-to-pay households fare best under full vertical choice. Even so, we find that all households prefer full vertical choice to competitive pricing, and that 82 percent of households prefer optimal regulation to full vertical choice.

Beyond the work noted above, our theoretical approach is most closely related to Azevedo and Gottlieb (2017), who also model demand for health insurance in a setting with vertically differentiated contracts and multiple dimensions of consumer heterogeneity. While their focus is on competitive equilibria, their numerical simulations also consider optimal pricing. They document that under certain distributions of consumer types, offering choice is optimal, while under others it is not.[6] Our paper focuses directly on why this is the case. We are (to our knowledge) the first to characterize the conditions under which it is optimal to offer vertical choice.[7] We also bring to bear a rich empirical approach that

[5] While our baseline is community rating, we also investigate allocations that can be supported by type-specific pricing based on age and whether a household has children.

[6] Their simulated population of consumers is characterized by lognormal distributions of types with moments set to match those estimated empirically by Einav et al. (2013).

[7] Ericson and Sydnor (2017) also consider the question of whether vertical choice is welfare-improving. They focus on consumer confusion as a source of inefficiency, while we focus on a setting with informed

permits substantially more flexibility in the distribution of consumer types.

Our paper also closely relates to the literature on health insurance menu design. Bundorf, Levin and Mahoney (2012) investigate the socially optimal allocation of consumers to insurers and find that optimal allocations cannot be achieved under uniform pricing. Einav, Finkelstein and Levin (2010) discuss, and Geruso (2017) studies empirically, the idea that difficulties in optimal screening can arise when observably different consumers have the same willingness to pay for insurance; this is a central issue in our setting. In concurrent work, Ho and Lee (2019) use a closely related framework to study the optimal choice of coverage level from the perspective of an employer offering a single coverage option. More generally, our paper relates to a growing empirical literature on allocational efficiency in health insurance markets (Cutler and Reber, 1998; Lustig, 2008; Carlin and Town, 2008; Dafny, Ho and Varela, 2013; Kowalski, 2015; Tilipman, 2018). Our work adds to this literature by focusing on the financial dimension of insurance and by permitting a rich space of potential contracts that need not themselves be observed in data.

Finally, we view our work as complementary to the large literature documenting the fact that consumers have difficultly optimizing over health insurance plans (Abaluck and Gruber, 2011, 2016; Ketcham et al., 2012; Handel and Kolstad, 2015; Bhargava, Loewenstein and Sydnor, 2017), which has recently also focused on ways in which consumers can be nudged into doing so (Abaluck and Gruber, 2016, 2017; Gruber, 2019; Bundorf et al., 2019). Importantly, if privately and socially optimal allocations do not align, more diligent consumers may just as well lead to less efficient outcomes (as is found by Handel, 2013). Our aim is to inform the design of health insurance markets so that better-informed consumers will always lead to better allocations.

The paper proceeds as follows. Section 2 presents our theoretical model and derives the objects needed to determine whether vertical choice should be offered. Section 3 describes our data and the variation it provides. Section 4 presents the empirical implementation of our model. Section 5 presents the model estimates and constructs willingness to pay and social surplus. Section 6 evaluates welfare and distributional outcomes under alternative pricing policies. Section 7 concludes.

consumers.

II Theoretical Framework

II.A Model

We consider a model of a health insurance market where consumers are heterogeneous along multiple dimensions and the set of traded contracts is endogenous. Since consumer health is not verifiable, contracts specify claims contingent only on healthcare utilization. We assume that premiums cannot vary by consumer characteristics, and we assert that each consumer will select a single contract.[8]

We denote a set of potential contracts by $X = \{x_0, x_1, ..., x_n\}$, where x_0 is a null contract that provides no insurance. Within X, contracts are vertically differentiated by the financial level of insurance coverage provided. Consumers are characterized by type $\theta : \{F, \psi, \omega\}$, where F is a distribution over potential health states, $\psi \in \mathbb{R}_{++}$ is a risk-aversion parameter, and ω is a parameter that describes consumer preferences over healthcare utilization (capturing the degree of moral hazard). We define a population by a distribution $G(\theta)$.

Demand for Health Insurance and Healthcare Utilization. Consumers are subject to a stochastic health state l, drawn from their distribution F. After their health state is realized, consumers decide the dollar amount $m \in \mathbb{R}_+$ of healthcare utilization ("spending") to consume. Contracts are characterized by the out-of-pocket cost $c(m, x)$ a consumer would need to pay in order to utilize healthcare.

Consumers value healthcare spending m and residual income y: $u(m, y) = u_\psi(y + b(m, l, \omega))$, where u_ψ is strictly increasing and concave, and b is the money-metric valuation of healthcare. Upon realizing their health state, consumers choose an amount of healthcare spending by trading off the benefit and the out-of-pocket cost: $m^*(l, \omega, x) = \text{argmax}_m \left(b(m, l, \omega) - c(m, x)\right)$.[9] Privately optimal healthcare utilization implies indirect benefit $b^*(l, \omega, x) = b(m^*(l, \omega, x), l, \omega)$ and indirect out-of-pocket cost $c^*_x(l, \omega) = c_x(m^*(l, \omega, x))$. Before the health state is realized, expected utility is given by

$$U(x, p, \theta) = \mathbb{E}\left[\, u_\psi\big(\hat{y} - p - c^*(l, \omega, x) + b^*(l, \omega, x)\big) \,|\, l \sim F\,\right], \tag{1}$$

[8]It may not be possible to condition premiums on consumer attributes if consumers have private information (see Cardon and Hendel (2001)). A regulator may not want to do so to prevent exposing consumers to costly reclassification risk (see Handel, Hendel and Whinston (2015)). Otherwise, the market could be partitioned according to observable characteristics, and each submarket could be considered separately.

[9]For convenience, we assume m^* is unique.

where p is the contract premium and $\hat{y}$ is initial income.

Private vs. Social Incentives. Absent insurance, consumers pay the full cost of healthcare, m. Socially optimal healthcare utilization is therefore the same as privately optimal utilization absent insurance.[10] In order to reach an expression for the social cost of insurance, it is useful to keep track of any difference between privately optimal spending $m^*(l,\omega,x)$ and socially optimal spending $m^*(l,\omega,x_0)$. Since insurance reduces the price consumers pay for healthcare, $m^*(l,\omega,x)$ typically exceeds $m^*(l,\omega,x_0)$. We refer to this induced utilization as "moral hazard spending."[11] A consumer's payoff from moral hazard spending is given by

$$v(l,\omega,x) = \underbrace{b^*(l,\omega,x) - b_0(l,\omega)}_{\substack{\textit{Benefit of moral}\\ \textit{hazard spending}}} - \underbrace{\big(c^*(l,\omega,x) - c_0(l,\omega,x)\big)}_{\substack{\textit{Out-of-pocket cost of}\\ \textit{moral hazard spending}}},$$

where $b_0(l,\omega) = b^*(l,\omega,x_0)$ is the indirect benefit of uninsured behavior, and $c_0(l,\omega,x) = c(m^*(l,\omega,x_0),x)$ is the indirect out-of-pocket cost of uninsured behavior at insured prices. Assuming insurance makes out-of-pocket costs weakly lower, $v(l,\omega,x)$ is weakly positive.

Calculations in Appendix A.1 show that if u_ψ features constant absolute risk aversion, willingness to pay for contract x relative to the null contract x_0 can be expressed as[12]

$$WTP(x,\theta) = \underbrace{\mathbb{E}_l[c_0(l,\omega,x_0) - c_0(l,\omega,x)]}_{\substack{\textit{Mean reduced out-of-pocket}\\ \textit{holding behavior fixed}}} + \underbrace{\mathbb{E}_l[v(l,\omega,x)]}_{\substack{\textit{Mean payoff from}\\ \textit{moral hazard spending}}} + \underbrace{\Psi(x,\theta)}_{\substack{\textit{Value of risk}\\ \textit{protection}}} . \tag{2}$$

Willingness to pay is composed of three terms: mean reduced out-of-pocket cost holding behavior fixed (at uninsured behavior), mean payoff from moral hazard spending, and the value of risk protection.[13] The first term captures the transfer of healthcare cost liability from the consumer to the insurer, which occurs even absent moral hazard. It is a financial

[10]Importantly, this is true only if m represents the true cost of healthcare provision, as we assume here.

[11]Following convention, we use the term "moral hazard" to describe elastic demand for the insured good when the state is not contractible. Note that this is not a problem of hidden action, but rather of hidden information. A fuller discussion of this (ab)use of terminology in the health insurance literature can be found in Section I.B of Einav et al. (2013), as well as in Pauly (1968) and Arrow (1968).

[12]Contracts represent a gamble over income and utility from healthcare utilization. WTP represents a certainty equivalent, equal to expected value plus risk premium. The role of constant absolute risk aversion is to ensure that the risk premium does not depend on the contract premium.

[13]Azevedo and Gottlieb (2017) also discuss how willingness to pay in this setting is composed of these three terms. Our formulation generalizes the decomposition in that it does not depend on particular functional forms for b, c, or F.

expected value that appears as an equal and opposite cost to the insurer. In contrast, the second and third terms depend on consumer preferences and are relevant to social welfare. Consumers may value the ability to consume more healthcare when they have better coverage, as well as the ability to smooth consumption across health states. Our accounting of social welfare takes both into consideration.

Insurer costs are given by $k(m, x)$, where $m = k(m, x) + c(m, x)$. A reduction in out-of-pocket cost is an increase in insurer cost, so $c_0(l, \omega, x_0) - c_0(l, \omega, x) = k(m^*(l, \omega, x_0), x) = k_0(l, \omega, x)$.[14] The social surplus generated by allocating a consumer to contract x (relative to allocating the same consumer to the null contract) is the difference between $WTP(x, \theta)$ and expected insured cost $\mathbb{E}_l[k^*(l, \omega, x)]$, which can be written:

$$SS(x,\theta) = \underbrace{\Psi(x,\theta)}_{\substack{\text{Value of risk} \\ \text{protection}}} - \underbrace{\mathbb{E}_l[\ k^*(l,\omega,x) - k_0(l,\omega,x) - v(l,\omega,x)\ ]}_{\substack{\text{Social cost} \\ \text{of moral hazard}}}. \tag{3}$$

Because the insurer is risk neutral, it bears no extra cost from uncertain payoffs. If there is moral hazard, the consumer's value of her expected healthcare spending falls below its cost, generating a welfare loss from insurance.[15]

The efficient contract for each type of consumer optimally trades off risk protection and the social cost of moral hazard: $x^{eff}(\theta) = \text{argmax}_{x \in X} SS(x, \theta)$. Given premium vector $\boldsymbol{p} = \{p_x\}_{x \in X}$, consumers choose the privately optimal contract that optimally trades off private utility and premium: $x^*(\theta, \boldsymbol{p}) = \text{argmax}_{x \in X}(WTP(x, \theta) - p_x)$.

Supply and Regulation. Contracts are supplied by a regulator, which can observe the distribution of consumer types and set premiums. The regulator need not break even on any given contract or in aggregate.[16] It could remove a contract from the set of contracts on offer by setting a premium of infinity. This model of supply is equivalent to perfect competition with a market regulator that has the power to tax and subsidize contracts.[17]

[14]To see this, note that $c_0(l, \omega, x_0) = m^*(l, \omega, x_0)$ and $c_0(l, \omega, x) = c(m^*(l, \omega, x_0), x)$.

[15]We can bound the social cost of moral hazard using revealed preference. $v(l, x, \omega)$ must be weakly positive (or else the consumer would not have changed their behavior in response to insurance), and it must be weakly lower than the insured cost of moral hazard spending (or else the consumer would have chosen that level of spending even absent insurance): $0 \leq v(l, x, \omega) \leq k^*(l, \omega, x) - k_0(l, \omega, x)$. The social cost of moral hazard can therefore be at most the expected insurer cost of moral hazard spending $\mathbb{E}_l[k^*(l, \omega, x) - k_0(l, \omega, x)]$.

[16]We assume any aggregate deficit can be funded by taxing consumer incomes. Since we assume constant absolute risk aversion, this is not different from increasing premiums on all contracts and calling it a tax.

[17]Precisely such a model is formalized in Section 6 of Azevedo and Gottlieb (2017)

The regulator sets premiums in order to align privately optimal $x^*(\theta, \boldsymbol{p})$ and socially optimal $x^{eff}(\theta)$ allocations as closely as possible. Equilibrium social welfare is given by

$$W(\boldsymbol{p}) = \int SS\big(x^*(\theta, \boldsymbol{p}), \theta\big)\ dG(\theta).$$

Our question is whether, or when, the regulator's solution will involve vertical choice. In other words, we ask whether there will be enrollment in more than one contract at the optimal allocation.[18]

II.B Graphical Analysis

We characterize the answer graphically for the case of a market with two potential contracts. This case conveys the basic intuition and can be depicted easily using the graphical framework introduced by Einav, Finkelstein and Cullen (2010).

First, it is useful to recognize that moral hazard and consumer heterogeneity are necessary conditions for the regulator to want to offer vertical choice. If there were no moral hazard, higher coverage would weakly increase social welfare for every consumer. The maximum possible coverage level would be the socially optimal contract for everyone, and to achieve this allocation the regulator could set the premiums of all other contracts sufficiently high that they are not chosen. If there were no consumer heterogeneity, all consumers would again have the same socially optimal contract, and the regulator would again set the premium of all other contracts sufficiently high that they are not chosen. In the following, we explore the more interesting (and more realistic) cases in which consumers do not all have the same socially optimal contract.

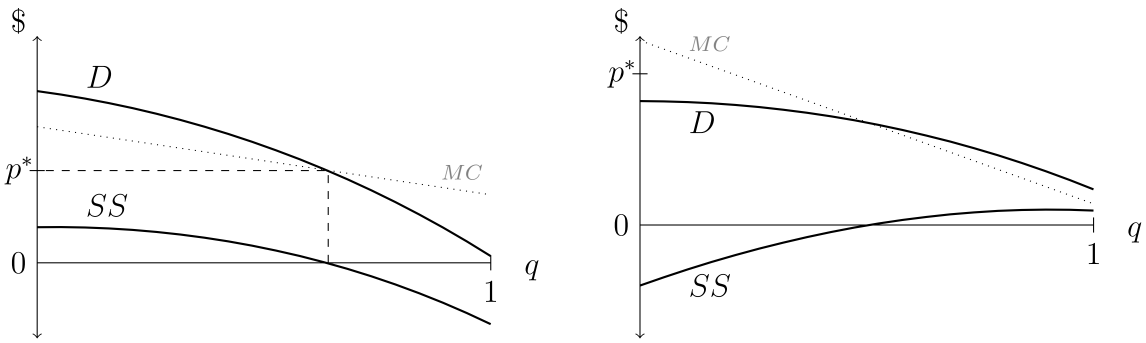
We consider an example with two potential contracts, x_H and x_L, where x_H provides more coverage than x_L. Figure 1 depicts two possible markets, corresponding to two populations $G^A(\theta)$ and $G^B(\theta)$. If a consumer does not choose x_H, they receive x_L. Since contracts are vertically differentiated, $WTP(x_H, \theta) \geq WTP(x_L, \theta)$ for all consumers. Each panel shows the demand curve D for contract x_H, representing marginal willingness to pay for x_H relative to x_L. The vertical axis plots the marginal price $p = p_H - p_L$ at which the contracts are offered. The horizontal axis plots the fraction q of consumers that choose x_H.

Each panel also shows the marginal cost curve MC and the marginal social surplus curve

[18] If the regulator sets premiums such that all consumers choose the same plan, we say that it has not offered vertical choice. This is to avoid discussion of, for example, whether an option of a plan with a premium of infinity is in fact an option at all.

Figure 1. Markets in which There (a) *Should* and (b) *Should Not* Be Vertical Choice

(a) Population $G^A(\theta)$ (b) Population $G^B(\theta)$



Notes: The figure shows two health insurance markets in which there are two contracts available: x_H and x_L, where x_H provides more coverage than x_L. Each panel shows the demand curve D, the marginal cost curve MC, and the social surplus curve SS for contract x_H relative to contract x_L. In the left panel, the regulator optimally offers vertical choice, and there is enrollment in both contracts. In the right panel, the regulator optimally does not offer vertical choice, and all consumers choose x_L.

SS. The marginal cost curve measures the expected marginal cost of insuring consumers under x_H relative to x_L. Because consumers with the same willingness to pay can have different costs, MC represents the average marginal cost among all consumers at a particular point on the horizontal axis (a particular willingness to pay). The social social surplus curve SS plots the vertical difference between D and MC. A particular point on the social surplus curve represents the average marginal social surplus $SS(x_H, \theta) - SS(x_L, \theta)$ among all consumers at that level of willingness to pay.

While D and MC must be weakly positive, the presence of moral hazard means that SS need not be. It is possible for a consumer to be over-insured. Moreover, our precondition that all consumers do not have the same optimal contract guarantees that in both populations, marginal social surplus will be positive for some consumers and negative for others. Given that SS represents an average, this condition does not guarantee that SS will itself cross zero. Since it is necessary for SS to cross zero for vertical choice to be optimal, we focus both graphical examples on cases in which that occurs.[19]

The key difference between populations $G^A(\theta)$ and $G^B(\theta)$ is whether consumers with high or low willingness to pay have a higher efficient level of coverage. In Figure 1a, marginal social surplus is increasing in marginal willingness to pay. The optimal marginal premium

[19] If SS does not cross zero, a single plan is on-average optimal at every level of willingness to pay. While allocating all consumers to that plan does not achieve the first best, vertical choice cannot offer something better. For example if SS lies everywhere above zero, the regulator will optimally offer only x_H. Note that this result of a single plan being on-average optimal across the distribution of willingness to pay corresponds to what we find empirically (cf. Figure 7).

p^* can sort consumers with on-average positive SS into x_H, and on-average negative SS into x_L. Because private and social incentives are aligned, it is possible to get consumers to self-select efficiently. In Figure 1b, marginal social surplus is decreasing in consumer willingness to pay, and efficient screening is no longer possible.

In population $G^B(\theta)$, any marginal premium between the minimum and the maximum value of D will result in some avoidable amount of "backward sorting." Consequently, any allocation with enrollment in both plans will be dominated by an allocation with enrollment in only one plan. No sorting dominates backward sorting because declaring no sorting means it is always possible to prevent "one side" of the backward sort.[20] In the example shown, the integral of SS is negative, meaning that the population would on average be over-insured in contract x_H. p^* is therefore anything high enough to induce all consumers to choose contract x_L.

Considering all cases, if the social surplus curve crosses zero at most once, vertical choice should be offered if and only if it crosses from above. More generally, the key characteristic of a population that determines whether vertical choice should be offered is whether consumers with higher willingness to pay have a higher efficient coverage level. This condition itself is complex. It is both theoretically ambiguous and, by our own assessment, not obvious whether we should expect it to be true. If healthy consumers change their behavior more in response to insurance, as is suggested by findings in Brot-Goldberg et al. (2017), this would tend toward positively aligning willingness to pay and efficient coverage level. If healthy consumers are more risk averse, as is suggested by findings in Finkelstein and McGarry (2006), this would tend toward negatively aligning them.

There is a question of what characteristics drive variation in willingness to pay, and in turn how those characteristics are correlated with the efficient level of coverage. The net result depends on the joint distribution of expected health spending, uncertainty in health spending, risk aversion, and moral hazard in the population. Moreover, it depends on how these primitives map into marginal willingness to pay and marginal insurer cost across nonlinear insurance contracts, as are common in U.S. health insurance markets and present in the empirical setting we study. Ultimately, whether high willingness to pay consumers should have higher coverage than low willingness to pay consumers is an open empirical question.

[20] To see this, consider the (worst possible) allocation $\tilde{q}$ at the point where SS intersects zero: A slightly higher allocation q' strictly dominates, as more consumers with positive marginal social surplus now enroll in contract x_H. The same logic applies to the left of $\tilde{q}$. The only allocations that cannot easily be ruled out as suboptimal are the endpoints, at which all consumers enroll in the same contract.

III Empirical Setting

III.A Data

Our data are derived from the employer-sponsored health insurance market for public school employees in Oregon between 2008 and 2013. The market is operated by the Oregon Educators Benefit Board (OEBB), which administers benefits for the employees of Oregon's 187 school districts. Each year, OEBB contracts with insurers to create a state-level "master list" of plans and associated premiums that school districts can offer to their employees. During our time period, OEBB contracted with three insurers, each of which offered a selection of plans. School districts then independently select a subset of plans from the state-level menu and set an "employer contribution" toward plan premiums, leading to variation in the subsidized premiums and set of plans available to different employees.[21]

The data contain employees' plan choice sets, realized plan choices, plan characteristics, and medical and pharmaceutical claims for all insured individuals. We observe detailed demographic information about employees and their families, including age, gender, zip code, health risk score, family type, and employee occupation type.[22,23] An employee's *plan menu* consists of the plan choice set and plan prices. Plan prices consist of the subsidized premium, potential contributions to a Health Reimbursement Arrangement (HRA) or a Health Savings Account (HSA), and potential contributions toward a vision or dental insurance plan.[24]

The decentralized determination of plan menus provides a plausibly exogenous source of

[21] Between 2008 and 2010, school districts could offer at most four plans; after 2010, there was no restriction on the number of plans a district could offer, but many still offered only a subset.

[22] Individual risk scores are calculated based on prior-year medical diagnoses and demographics using Johns Hopkins ACG Case-Mix software. This software uses diagnostic information contained in past claims data as well as demographic information to predict future healthcare spending. See, for example, Brot-Goldberg et al. (2017); Carlin and Town (2008); or Handel and Kolstad (2015) for more in-depth explanation of the software and examples of its use in economic research.

[23] Possible employee occupation types are licensed administrator, non-licensed administrator, classified, community college non-instructional, community college faculty, confidential, licensed, substitute, and superintendent. Within each type, an employee can be either full-time or part-time. Possible family types are employee only; employee and spouse; employee and child(ren); and employee, spouse, and child(ren).

[24] Decisions about HSA/HRA and vision/dental contributions are also made independently by school districts. An HRA is a notional account that employers can use to reimburse employees' uninsured medical expenses on a pre-tax basis; balances expire at the end of the year or when the employee leaves the employer. An HSA is a financial account maintained by an external broker to which employers or employees can make pre-tax contributions. Data on employer premium contributions and savings account contributions were hand-collected via surveys of each school district. Additional details on the data collection process can be found in Abaluck and Gruber (2016).

variation in both prices and choice sets. While all plan menus we observe are quite generous in that the plans are highly subsidized, there is substantial variation across districts in the range of coverage levels offered and in the exact nature of the subsidies.[25] Moreover, school districts can adjust plan menus by family type and occupation type, resulting in variation both within and across districts. Plan menu decisions are made by benefits committees consisting of district administrators and employees, and subsidy designs are influenced by bargaining agreements with local teachers' unions. Between 2008 and 2013, we observe 13,661 unique combinations of year, school district, family type, and occupation type, resulting in 7,835 unique plan menus.

Household Characteristics. We restrict our analysis sample to households in which the oldest member is not older than 65, the employee is not retired, and all members are enrolled in the same plan for the entire year. Further, because a prior year of claims data is required to estimate an individual's prospective health risk score, we require that households have one year of data prior to inclusion. This means that our sample begins in 2009. In total, our sample consists of 44,562 unique households, representing 117,949 individuals.[26]

Table 1 provides summary statistics on our panel of households. The average employee is age 47.4, and the average enrollee (employees and their covered dependents) is age 39.8. Enrollees are 54 percent female, and 72 percent of households are "families" (purchased insurance to cover more than the employee alone). Households have 2.54 enrollees on average.

Employees receive large subsidies toward the purchase of health insurance. The average household paid only $880 per year for their chosen plan; the median household paid nothing. Meanwhile, the average full premium paid to insurers was $11,500, meaning that the average household received an employer contribution of $10,620. Households had average out-of-pocket spending of $1,694 and average total healthcare spending of $10,754.

Households were highly likely to remain in the same plan and with the same insurer they chose the previous year. However, OEBB can adjust the state-level master list of available plans, and school districts can adjust choice sets over time. Because their prior choice was no longer available, such adjustments forced 19.6 percent of household-years to switch

[25] The majority of school districts used either a fixed dollar contribution or a percentage contribution, but the levels of the contribution varied widely. Other districts used a fixed employee contribution. In addition, the districts' policies for how "excess" contributions were treated varied; in some cases, contribution amounts in excess of the full plan premium could be "banked" by the employee in a HSA or HRA, or else put toward the purchase of a vision or dental insurance plan.

[26] Table A.1 provides additional details on sample construction.

Table 1. Household Summary Statistics

Sample demographics	2009	2010	2011	2012	2013
Number of households	31,074	29,538	29,279	27,897	24,283
Number of enrollees	78,932	75,129	75,601	72,311	63,264
Enrollee age, mean (med.)	39.7	39.8	39.8	40.1	40.0
	(38.0)	(38.0)	(37.7)	(38.0)	(37.8)
Premiums					
Employee premium ($), mean (med.)	885	1,023	523	1,079	905
	(0)	(0)	(0)	(0)	(0)
Full premium ($), mean (med.)	11,170	11,785	10,433	12,253	12,000
	(11,665)	(11,801)	(11,021)	(12,278)	(12,362)
Household healthcare spending					
Total spending ($), mean (med.)	10,563	10,405	10,911	10,984	10,967
	(4,753)	(4,589)	(4,595)	(4,569)	(4,559)
Out-of-pocket ($), mean (med.)	1,152	1,634	1,884	1,897	1,998
	(743)	(1,089)	(1,306)	(1,292)	(1,234)
Switching (percentage of households)					
Forced to switch plan	0.06	0.34	0.12	0.05	0.46
insurer	0.01	0.02	0.02	0.02	0.00
Unforced, switched plan	0.13	0.23	0.22	0.22	0.04
insurer	0.06	0.05	0.03	0.01	0.02

Notes: Enrollees are employees plus their covered dependents. Statistics for premiums are for households' chosen plans, as opposed to for all possible plans. Sample medians are shown in parentheses.

plans and 1.4 percent to switch insurers. Among household-years where the prior choice was available, 17.2 percent voluntarily switched plans and 3.4 percent voluntarily switched insurers. The presence of both forced and unforced switching is particularly important in our empirical model for identifying the extent of "inertia" in households' plan and insurer choices.

To allow for geographic variation in tastes for each insurer, we divide the state into three regions, based on groups of adjacent Hospital Referral Regions (HRRs): the Portland and Salem HRRs in northwest Oregon (containing 64 percent of households); the Eugene and Medford HRRs in southwest Oregon (26 percent of households); and the Bend, Spokane, and Boise HRRs in eastern Oregon (10 percent of households).[27]

Plan Characteristics. During our sample period, OEBB contracted with three insurers: Kaiser, Providence, and Moda. Kaiser offered HMO plans that require enrollees to use only Kaiser healthcare providers and obtain referrals for specialist care. Moda and Providence offered PPO plans with broad provider networks. Kaiser and Providence each offered two or three plans per year at high coverage levels. Moda offered between seven and nine plans per

[27]Because HRRs do not respect state boundaries, some HRRs in our regions have names of cities outside Oregon, but nonetheless contain parts of Oregon. For more information and HRR maps, see http://www.dartmouthatlas.org/data/region.

year, with wide variation in coverage level. Within each insurer, plans were differentiated only by coverage level.

Table 2 summarizes the state-level master list of plans made available by OEBB in 2009. The average employee premium reflects the average annual premium employees would have had to pay for each plan. The full premium reflects the per-employee premium negotiated by OEBB and the insurer.[28] The difference between the employee premium and the full premium is the contribution by the school district. Plan cost-sharing features vary by whether the household is an individual (the employee alone) or a family (anything else). The deductible and out-of-pocket maximum shown are for a family household.[29]

Table 2. Plan Characteristics, 2009

Plan	Actuarial Value	Avg. Employee Premium ($)	Full Premium ($)	Deductible ($)	OOP Max. ($)	Market Share
Kaiser - 1	0.97	688	10,971	0	1,200	0.07
Kaiser - 2	0.96	554	10,485	0	2,000	0.11
Kaiser - 3	0.95	473	10,163	0	3,000	0.00
Moda - 1	0.92	1,594	12,421	300	500	0.27
Moda - 2	0.89	1,223	11,839	300	1,000	0.05
Moda - 3	0.88	809	11,174	600	1,000	0.11
Moda - 4	0.86	621	10,702	900	1,500	0.10
Moda - 5	0.82	428	9,912	1,500	2,000	0.13
Moda - 6	0.78	271	8,959	3,000	3,000	0.04
Moda - 7	0.68	92	6,841	3,000	10,000	0.01
Providence - 1	0.96	2,264	13,217	900	1,200	0.07
Providence - 2	0.95	1,995	12,895	900	2,000	0.02
Providence - 3	0.94	1,825	12,683	900	3,000	0.01

Notes: Actuarial value is calculated as the ratio of the sum across all households of insured spending to that of total spending. The average employee premium is taken across all employees, even those who did not choose a particular plan. The full premium reflects the premium negotiated by OEBB and the insurer; the one shown is for an employee plus spouse. The deductible and in-network out-of-pocket maximum shown are for in-network services for a family household.

One way to summarize and compare plan coverage levels is by using actuarial value. This measure reflects the share of total population spending that would be insured under a given plan. Less generous plans have lower actuarial values. To calculate actuarial value, we simulate the out-of-pocket spending that *all* households would have had in every potential plan, and then compute average insured spending divided by average total spending across all households for each plan.[30]

[28] This full premium varies formulaically by family type; the premium shown is for an employee plus spouse.

[29] Many other cost-sharing details determine plan coverage level. For the purposes of our empirical model, we estimate a deductible, coinsurance rate, and out-of-pocket maximum that best fit the relationship between out-of-pocket spending and total spending observed in the claims data; this procedure is described in Appendix A.2.

[30] We calculate counterfactual out-of-pocket spending using the "claims calculator" developed for this setting

Plan offerings in later years look qualitatively similar.[31] The notable exception is that Providence was no longer available in 2012 and 2013. Moda maintained a roughly 75 percent market share throughout 2009 to 2013; Kaiser and Providence initially split the remaining share, but Kaiser steadily gained share thereafter.

III.B Variation in Coverage Levels and Plan Menus

For the purposes of this research, the two most important features of our setting are the isolated variation along the dimension of coverage level and the plausibly exogenous variation in plan menus. Variation in coverage level exists primarily among the plans offered by Moda. Variation in plan menus stems from the decentralized determination of employee health benefits. Both are central to identification of our empirical model.

To provide a sense of this variation, Figure 2 shows the relationship between healthcare spending and plan actuarial value (AV) for households that chose Moda in 2009. In the left panel, households are grouped by their *chosen* plan. The plot shows average spending among households in each of the seven Moda plans, weighting each plan by enrollment. Unsurprisingly, households enrolled in more generous plans had higher spending, reflecting adverse selection, moral hazard, or both.

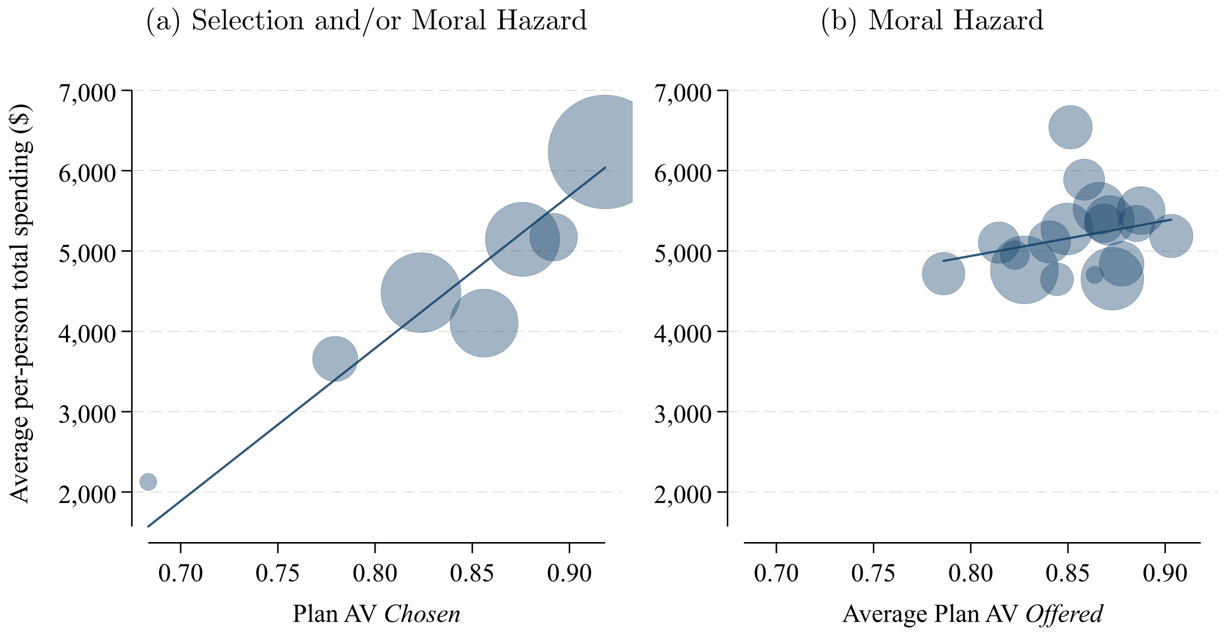
The right panel groups households by their plan menu. It plots the average actuarial value of plans that were *offered* in a given menu against the average spending of households presented with that menu. There is one observation for each unique value of average actuarial value offered. Households that were offered higher coverage—and which therefore were presumably more likely to choose higher coverage—had higher spending. The patterns depicted in both panels persist when we control for observables, suggesting the presence of adverse selection on unobservables and of moral hazard.

Identification of our structural model will proceed in much the same way as the above arguments. A key identifying assumption is that plan menus are independent of household unobservables, conditional on household observables. An important threat to identification is that school districts chose plan menu generosity in response to unobservable information about employees that would also drive healthcare spending. To the extent that districts with unobservably sicker households offered better coverage, this would lead us to overstate the extent of moral hazard.

by Abaluck and Gruber (2016).

[31] Corresponding tables for the plans offered between 2010 and 2013 are shown in Table A.2.

Figure 2. Average Spending by Coverage Level *Chosen* and *Offered*



Notes: The figure shows the relationship between average per-person total spending and plan actuarial value (AV) for households that selected Moda in 2009. In the left panel, each dot represents one of the seven Moda plans. In the right panel, households are grouped by their plan menu, and each dot represents a unique value of average AV offered. The size of each dot indicates the number of households represented. Lines of best fit are weighted accordingly.

We investigate this possibility by attempting to explain plan menu generosity with observable household characteristics, in particular health. We argue that if plan menus were not responding to *observable* information about household health, it is unlikely that they were responding to *unobservable* information. We find this argument compelling because we almost certainly have better information on household health (through health risk scores) than school districts did at the time they made plan menu decisions. Table A.4 presents this exercise. Conditional on family type, we find no correlation between plan menu generosity and household risk score. Appendix A.3 describes these results in greater detail and discusses how we measure plan menu generosity (accommodating plan prices in addition to plan choice sets). It also presents additional tests for what *does* explain variation in plan menus. We find that, among other things, plan menu generosity is higher for certain union affiliations, lower for substitute teachers and part-time employees, decreasing in district average house price index, and decreasing in the percentage of registered Republicans in a school district. None of these relationships are inconsistent with our understanding of the process by which district benefits decisions are made.

We exploit this identifying variation directly within our structural model, but we can also use it in a more isolated way to produce reduced-form estimates of moral hazard.

Appendix A.4 presents an instrumental variables analysis of moral hazard using two-stage least squares. The estimates yield a moral hazard "elasticity" that can be directly compared with others in the literature. Our overall estimate of the elasticity of demand for healthcare spending with respect to the end-of-year average price of healthcare is -0.27, broadly similar to the benchmark estimate of -0.2 from the RAND experiment (Manning et al., 1987; Newhouse, 1993). We also find suggestive evidence of heterogeneity in moral hazard effects, which is an important aspect of our research question and our structural model.

IV Empirical Model

IV.A Parameterization

We parameterize household utility and the distribution of health states, allowing us to represent our theoretical model fully in terms of data and parameters to estimate. We extend the theoretical model to account for the fact that in our empirical setting, there are multiple insurers, consumers are households consisting of individuals, consumers may value a dollar of premiums and a dollar of out-of-pocket cost differently, and consumers make repeated plan choices over time.

Household Utility. Following Cardon and Hendel (2001) and Einav et al. (2013), we parameterize valuation of healthcare spending to be quadratic in its difference from the health state. Household k's valuation of spending level m given health state realization l is given by

$$b(m, l, \omega_k) = (m - l) - \frac{1}{2\omega_k}(m - l)^2, \tag{4}$$

where ω_k governs the curvature of the benefit of additional spending and, ultimately, the degree to which optimal utilization will vary across coverage levels. Given an out-of-pocket cost function $c_{jt}(m)$ for plan j in year t, privately optimal total healthcare spending is given by $m^*_{jt}(l, \omega_k) = \text{argmax}_m \left(b(m, l, \omega_k) - c_{jt}(m)\right)$.[32] Solving yields $m^*_{jt}(l, \omega_k) = \omega_k(1 - c'_{jt}(m^*_{jt})) + l$.

This parameterization is attractive because it produces reasonable predicted behavior under nonlinear insurance contracts, and it is tractable enough to be used inside an opti-

[32]The out-of-pocket cost function $c_{jt}(m)$ is indexed by t because cost-sharing parameters vary within a plan across years. Note that $c_{jt}(m)$ also varies by household type (individual versus family), but we omit an additional index to save on notation.

mization routine.[33] Additionally, ω_k can be usefully interpreted as the incremental spending induced when moving a household from no insurance (when marginal out-of-pocket cost is one and $m^* = l$) to full insurance (when marginal out-of-pocket cost is zero and $m^* = \omega + l$). Substituting for m^*, we denote the benefit of optimal utilization as $b^*_{jt}(l, \omega_k)$ and the associated out-of-pocket cost as $c^*_{jt}(l, \omega_k)$. Households face uncertainty about payoffs only through uncertainty in $b^*_{jt}(l, \omega_k) - c^*_{jt}(l, \omega_k)$.[34]

As in our theoretical model, we assume that households have constant absolute risk aversion preferences. Facing uncertainty about their healthcare needs, household k in year t derives the following expected utility from plan choice j:

$$U_{kjt} = \int - \exp\big(- \psi_k x_{kjt}(l)\big) dF_{kft}(l), \tag{5}$$

where ψ_k is the coefficient of risk aversion, x_{kjt} is the payoff associated with realization of health state l, and F_{kft} is the distribution of health states. Health state distributions can vary by insurer $f(j)$ in order to capture differences in provider prices across insurers.

The payoff of health state realization l when enrolled in plan j is given by

$$x_{kjt}(l) = -p_{kjt} + \alpha^{OOP} \left(b^*_{jt}(l, \omega_k) - c^*_{jt}(l, \omega_k)\right) + \delta^{f(j)}_{kj} + \gamma^{inertia}_{kjt} + \beta \mathbf{X}_{kjt} + \sigma_\epsilon \epsilon_{kjt}, \tag{6}$$

where p_{kjt} is the household's plan premium (net of the employer contribution); $b^*_{jt}(l, \omega_k) - c^*_{jt}(l, \omega_k)$ is the payoff from optimal utilization measured in units of out-of-pocket dollars; $\delta^{f(j)}_{kj}$ are insurer fixed effects that control for brand and other insurer characteristics, $\gamma^{inertia}_{kjt}$ are a set of fixed effects for both the plan and the insurer a household was enrolled in the previous year; and $\mathbf{X}_{kjt}$ is a set of additional covariates that can affect household utility.[35] The payoff x_{kjt} is measured in units of premium dollars. Out-of-pocket costs can be valued differently from premiums through parameter α^{OOP}.[36] Finally, ϵ_{kjt} represents a

[33] The model predicts that if a consumer realizes a health state just under the plan deductible, she will take advantage of the proximity to cheaper healthcare and consume a bit more (putting her into the coinsurance region). Figure A.3 provides a depiction of optimal spending behavior predicted by this model.

[34] Under our parameterization, $b^*_{jt}(l, \omega_k) = \frac{\omega_k}{2}(1 - c'_{jt}(m^*_{jt})^2)$. Because both b^*_{jt} and c^*_{jt} are increasing in ω, a larger ω will contribute to a less risky distribution of payoffs. All else equal, this would work to align willingness to pay and efficient coverage level. An important motivation for the inclusion of unobservable heterogeneity in risk aversion is to allow it to vary flexibly with respect to the amount of moral hazard.

[35] $\mathbf{X}_{kjt}$ includes HRA or HSA contributions HA_{kjt}; vision and dental plan contributions VD_{kjt}; and a fixed effect $\nu^{NarrowNet}_{jt}$ for one plan (Moda - 2) that had a narrow provider network (in 2011 and 2012 only). The associated parameters for health account and vision/dental contributions are α^{HA} and α^{VD}, respectively.

[36] We cannot distinguish between potential reasons why premiums may be valued differently from out-

household-plan-year idiosyncratic shock, with magnitude σ_ϵ to be estimated. We assume that the shocks are independently and identically distributed Type 1 Extreme Value. In each year, households choose the plan that maximizes expected utility from the set of plans $\mathcal{J}_{kt}$ available to them: $j^*_{kt} = \text{argmax}_{j \in \mathcal{J}_{kt}} U_{kjt}$.

Distribution of Health States. We assume that individuals face a lognormal distribution of health states and households face the sum of draws from each individual in the household. Because there is no closed-form expression for the distribution of the sum of draws from lognormal distributions, we represent a household's distribution of health states using a lognormal that approximates. We derive the parameters of the approximating distribution using the Fenton-Wilkinson method.[37] This novel means of modeling the household distribution of health states allows us to fully exploit the large amount of heterogeneity in household composition that exists in our data. Importantly given the size of our data, it also allows us to closely fit observed spending distributions using a smaller number of parameters than would be required if demographic covariates were aggregated to the household level. We estimate the parameters of individuals' health state distributions, allowing parameters to vary with individual demographics.

An individual i faces uncertain health state $\tilde{l}^i$, which has a shifted lognormal distribution with support $(-\kappa_{it}, \infty)$:

$$\log(\tilde{l}^i + \kappa_{it}) \sim N(\mu_{it}, \sigma^2_{it}).$$

The shift is included to capture the mass of individuals with zero spending observed in the data. If κ_{it} is positive, then negative health states are permitted, which may imply zero spending.[38] Parameters μ_{it}, σ_{it}, and κ_{it} are parameterized to vary with individual demographics, including health risk score, which can vary over time.

A household k faces uncertain health state $\tilde{l}$, which has a shifted lognormal distribution with support $(-\kappa_{kt}, \infty)$: $\log(\tilde{l} + \kappa_{kt}) \sim N(\mu_{kt}, \sigma^2_{kt})$. Under the approximation, household-level parameters μ_{kt}, σ_{kt}, and κ_{kt} can be calculated as functions of the individual-level parameters μ_{it}, σ_{it}, and κ_{it}. Variation in μ_{kt}, σ_{kt}, and κ_{kt} across households, as well as within households over time, arises from variation in household composition: the number

of-pocket costs. We expect the tax deductibility of premiums would push α^{OOP} up, while systematic underestimation of out-of-pocket spending would push α^{OOP} down.

[37] Additional details can be found in Appendix B.1

[38] A household that realizes a negative health state will have zero spending, as long as ω_k is not so large that optimal spending becomes positive. Operationally, this entails amending the optimal spending policy to be $m^*_{jt}(l, \omega_k) = \max(0, \omega_k(1 - c'_{jt}(m^*_{jt})) + l)$.

of individuals and each individual's demographics. In addition to this observable heterogeneity, we incorporate unobserved heterogeneity in household health though parameter μ_{kt}. In this way, adverse selection (on unobservables) is permitted, since households can hold private information about their health that can drive both plan choices and spending outcomes.

Finally, to account for the fact that there are multiple insurers in our empirical setting, we introduce an additional set of parameters ϕ_f to serve as "exchange rates" for monetary health states across insurers. These parameters are intended to capture differences in total healthcare spending that are driven by differences in provider prices across insurers.[39] For example, the same doctor's office visit might lead to different amounts of total spending across insurers simply because each insurer paid the doctor a different price. We do not want such variation to be attributed to differences in underlying health. We therefore capture it in a structured way by estimating insurer-level parameters that multiply realized health states, transforming them from underlying "quantities" into the monetary spending amounts we observe in the claims data. A household's money-metric health state l is the product of an insurer-level multiplier ϕ_f and the underlying "quantity" health state $\tilde{l}$, where $\tilde{l}$ is lognormally distributed depending only on household characteristics. Taken together, the distribution F_{kft} is defined by

$$
\begin{gathered}
l = \phi_f \tilde{l}, \\
\log(\tilde{l} + \kappa_{kt}) \sim N(\mu_{kt}, \sigma_{kt}^2).
\end{gathered}
$$

IV.B Identification

We aim to recover the joint distribution across households of willingness to pay, risk protection, and the social cost of moral hazard associated with different levels of insurance. Variation in these objects arises from variation in either household preferences (the risk-aversion and moral-hazard parameters) or in households' distribution of health states. Our primary identification concerns are (i) distinguishing preferences from private information about health, (ii) distinguishing taste for out-of-pocket spending (α^{OOP}) from risk aversion, and (iii) identifying heterogeneity in the risk-aversion and moral-hazard parameters.

We first explain how ω, which captures moral hazard, is distinguished from unobserved

[39] In reality, ϕ_f will also capture other multiplicative differences across insurers, such as care management protocols or provider practice patterns. Even so, we think it likely that most of the variation comes from differences in average provider prices across insurers. Our estimates of ϕ_f conform to our priors on provider price variation across insurers (most notably, that Kaiser pays the lowest prices).

variation in μ_{kt}, which captures adverse selection. In the data, there is a strong positive correlation between plan generosity and total healthcare spending (see Figure 2a). A large part of this relationship can be explained by observable household characteristics. But even conditional on observables, there is still residual positive correlation. This residual correlation could be attributable to either the effect of lower out-of-pocket prices driving utilization (moral hazard) or private information about health affecting both utilization and coverage choice (adverse selection). The key to distinguishing between these explanations is the variation in plan menus.

Both within and across school districts, we observe similar households facing different menus of plans.[40] As a result, some households are more likely to choose higher coverage only because of their plan menu. The amount of moral hazard is identified by the extent to which households facing more generous plan menus also have higher healthcare spending. On the other hand, we also observe similar households facing similar menus of plans, but still making different plan choices. This variation identifies the degree of private information about health, as well as the magnitude of the idiosyncratic shock σ_ϵ. Conditional on observables and the predicted effects of moral hazard, if households that inexplicably choose more generous coverage also inexplicably realize higher healthcare spending, this variation in plan choice will be attributed to private information about health. Otherwise, any residual unexplained variation in plan choice will be attributed to the idiosyncratic shock.

Both risk aversion (ψ) and the relative valuation of premiums and out-of-pocket spending (α^{OOP}) affect households' preference for more- or less-generous insurance, but do not affect healthcare spending. To distinguish between them, we use cases in which observably different households face similar plan menus. Risk aversion is identified by the degree to which households' taste for higher coverage is positively related to uncertainty in out-of-pocket spending, holding expected out-of-pocket cost fixed. The relative valuation of premiums and out-of-pocket spending is identified by the rate at which households trade off premiums with expected out-of-pocket cost, holding uncertainty in out-of-pocket cost fixed.

Unlike the preceding arguments, identification of unobserved heterogeneity in risk aversion and the moral hazard parameter relies on the panel nature of our data. Plan menus, household characteristics, and plan characteristics change over time. We therefore observe the same households making choices under different circumstances. If we had a large number of observations for each household and sufficient variation in circumstances, the preceding

[40] Our identification argument for moral hazard is similar to that of Cardon and Hendel (2001).

arguments could be applied household by household, and we could nonparametrically identify the distribution of ψ and ω by recovering household-specific estimates. In reality, we have at most five observations for each household. We ask less of this data by assuming that the distribution of unobserved heterogeneity is multivariate normal. The variance and covariance of the unobserved components of household types are identified by the extent to which different households consistently act in different ways. For example, if some households consistently make choices that reflect high risk aversion and other (observationally equivalent) households consistently make choices that reflect low risk aversion, this will show up as variance in the unobserved component of the risk-aversion parameter.

IV.C Estimation

We project the parameters of the individual health state distributions μ_{it}, σ_{it}, and κ_{it} on time-varying individual demographics:

$$\begin{aligned}\mu_{it} &= \boldsymbol{\beta}^{\mu}\mathbf{X}_{it}^{\mu},\\ \sigma_{it} &= \boldsymbol{\beta}^{\sigma}\mathbf{X}_{it}^{\sigma},\\ \kappa_{it} &= \boldsymbol{\beta}^{\kappa}\mathbf{X}_{it}^{\kappa}.\end{aligned} \tag{7}$$

$\mathbf{X}_{it}^{\mu}$, $\mathbf{X}_{it}^{\sigma}$, and $\mathbf{X}_{it}^{\kappa}$ contain indicators for the 0–30th, 30–60th, 60–90th, and 90–100th percentiles of individual health risk scores.[41] $\mathbf{X}_{it}^{\mu}$ and $\mathbf{X}_{it}^{\kappa}$ also contain a linear term in risk score, which is estimated separately for the 90–100th percentile group. $\mathbf{X}_{it}^{\mu}$ also contains an indicator for whether the individual is under 18 years old and an indicator for whether the individual is a female between the ages of 18 and 30.

Using the derivations shown in Appendix B.1, the parameters of households' health state distributions are a function of individual parameters:

$$\begin{aligned}\sigma_{kt}^2 &= \log[1+\left[\sum_{i\in\mathcal{I}_k}\exp(\mu_{it}+\frac{\sigma_{it}^2}{2})\right]^{-2}\sum_{i\in\mathcal{I}_k}(\exp(\sigma_{it}^2)-1)\exp(2\mu_{it}+\sigma_{it}^2)],\\ \bar{\mu}_{kt} &= -\frac{\sigma_{kt}^2}{2}+\log[\sum_{i\in\mathcal{I}_k}\exp(\mu_{it}+\frac{\sigma_{it}^2}{2})],\\ \kappa_{kt} &= \sum_{i\in\mathcal{I}_k}\kappa_{it},\end{aligned} \tag{8}$$

[41]The distributions of risk scores are highly right-skewed, so these groupings fit the data better than true quartiles.

where $\mathcal{I}_k$ represents the set of individuals in household k. Private information about health is incorporated with normally distributed unobservable heterogeneity in μ_{kt}. The household-specific mean of μ_{kt} is given by $\bar{\mu}_{kt}$, and the variance is given by σ^2_μ. A large σ^2_μ means that households have substantially more information about their health than can be explained by observables.

We assume that μ_{kt}, ψ_k, and ω_k are jointly normally distributed:

$$\begin{bmatrix} \mu_{kt} \\ \omega_k \\ \log(\psi_k) \end{bmatrix} \sim N\left(\begin{bmatrix} \bar{\mu}_{kt} \\ \boldsymbol{\beta}^\omega \mathbf{X}^\omega_k \\ \boldsymbol{\beta}^\psi \mathbf{X}^\psi_k \end{bmatrix}, \begin{bmatrix} \sigma^2_\mu & & \\ \sigma^2_{\omega,\mu} & \sigma^2_\omega & \\ \sigma^2_{\psi,\mu} & \sigma^2_{\omega,\psi} & \sigma^2_\psi \end{bmatrix} \right). \tag{9}$$

There is both observed (through the mean vector) and unobserved (through the covariance matrix) heterogeneity in each parameter. Covariates $\mathbf{X}^\omega_k$ and $\mathbf{X}^\psi_k$ include an indicator for whether the household has children and a constant.[42]

We model inertia at both the plan and insurer level: $\gamma^{inertia}_{kjt} = \gamma^{plan}_k \mathbf{1}_{k,j=j(t-1)} + \gamma^{ins}_k \mathbf{1}_{k,f=f(t-1)}$. We allow γ^{plan}_k to vary linearly with household age and allow the intercept to vary by whether the household has children.[43] To capture whether sicker households face higher barriers to switching insurers (and therefore provider networks), we allow γ^{ins}_k to vary linearly with household risk score.[44] Insurer fixed effects $\delta^{f(j)}_k$ vary by household age and whether a household has children, and we allow the intercepts to vary by geographic region in order to capture the relative attractiveness of insurer provider networks across different parts of the state (as well as other sources of geographical heterogeneity in insurer preferences). We normalize the fixed effect for Moda to be zero. Since the parameters of the individual health state distributions can vary freely, the "provider price" parameters require normalization: ϕ_{Moda} is normalized to one.

We estimate the model via maximum likelihood. Our estimation approach follows Revelt and Train (1998) and Train (2009), with the important distinction that we model a discrete/continuous choice. Our construction of the discrete/continuous likelihood function follows Dubin and McFadden (1984). The likelihood function for a given household is the conditional density of its observed sequence of total healthcare spending, given its observed sequence of plan choices. We use Gaussian quadrature to integrate numerically over the

[42] If a household has children in some years but not others, we assign it to its modal status.

[43] Household age is calculated as the mean age of all adults in a household across all years.

[44] Additionally, in 2013, Moda rebranded and changed the names of all of its plans and added a plan, in a way that did not result in a direct mapping between all 2012 and 2013 plans. To capture this flexibly, we estimate a separate insurer-level inertia parameter for Moda plans in 2013.

distribution of unobserved heterogeneity, as well as the distributions of household health states. Additional details on the estimation procedure are provided in Appendix B.2.

V Results

V.A Model Estimates

Table 3 presents parameter estimates. Column 3 presents our primary specification, as described in Section IV. Columns 1 and 2 present simpler specifications that are useful in understanding and validating the model. The table excludes insurer fixed effects and health state distribution parameters; these can be found in Table A.11.

Table 3. Parameter Estimates

	(1)		(2)		(3)	
Variable	Parameter	Std. Err.	Parameter	Std. Err.	Parameter	Std. Err.
Employee Premium ($000s)	$-1.000^\dagger$		$-1.000^\dagger$		$-1.000^\dagger$	
OOP spending, $-\alpha^{OOP}$	−1.504	0.024	−1.519	0.024	−1.348	0.028
HRA/HSA contributions, α^{HA}	0.292	0.023	0.293	0.023	0.250	0.023
Vision/dental contributions, α^{VD}	1.346	0.025	1.340	0.025	1.143	0.037
Plan inertia intercept, γ^{plan}	4.272	0.095	5.009	0.059	4.265	0.098
Plan inertia * (Age−40), γ^{plan}	0.019	0.002	0.073	0.006	0.018	0.002
Plan inertia * **1**[Children], γ^{plan}	0.189	0.040	1.208	0.119	0.188	0.041
Insurer inertia intercept,γ^{ins}	6.097	0.116	4.605	0.231	6.030	0.120
Insurer inertia * Risk score, γ^{ins}	0.182	0.026	0.501	0.074	0.117	0.026
Moda-specific inertia, 2013	1.824	0.196	1.924	0.199	1.555	0.198
Narrow net. plan, $\nu^{NarrowNet}$	−2.662	0.165	−2.665	0.165	−2.459	0.169
Kaiser provider prices, ϕ_K	0.669	0.007	0.831	0.006	0.766	0.000
Providence provider prices, ϕ_P	1.038	0.017	1.096	0.017	1.061	0.006
Risk aversion intercept, $\boldsymbol{\beta}^{\psi}$	−0.495	0.059	−0.597	0.065	0.313	0.049
Risk aversion * **1**[Children], $\boldsymbol{\beta}^{\psi}$	−0.344	0.070	−0.221	0.062	−1.103	0.096
Std. dev. of risk aversion, σ_ψ	0.921	0.037	0.997	0.102	0.603	0.131
Std. dev. of private health info., σ_μ	0.853	0.003	0.314	0.049	0.271	0.005
Moral hazard intercept, $\boldsymbol{\beta}^{\omega}$					1.133	0.000
Moral hazard * **1**[Children], $\boldsymbol{\beta}^{\omega}$					0.615	0.000
Std. dev. of moral hazard, σ_ω					0.145	0.073
Corr(μ, ψ), $\rho_{\mu,\psi}$	0.354	0.000	0.168	0.088	0.710	0.102
Corr(ψ, ω), $\rho_{\psi,\omega}$					−0.168	0.045
Corr(μ, ω), $\rho_{\mu,\omega}$					0.027	0.013
Scale of idiosyncratic shock, σ_ϵ	2.516	0.027	2.519	0.027	2.406	0.028
Insurer * {Region, Age, **1**[Child.]}	Yes		Yes		Yes	
Heterogeneity in spending dists.			Yes		Yes	
Number of observations	679,773		679,773		679,773	

Notes: The table presents estimates for selected parameters; Table A.11 presents estimates for the remaining parameters. Standard errors are derived from the analytical Hessian of the likelihood function. Column 3 presents our primary estimates, while columns 1 and 2 present alternative specifications. All models are estimated on an unbalanced panel of 44,562 households over five years. Coefficients of absolute risk aversion are relative to thousands of dollars. Estimates from column 3 are the inputs into the calculation in Section V.B. To make non-interacted coefficients more readily interpretable, we use (Age−40). †By normalization.

Column 1 presents a version of the model in which there is no moral hazard and no heterogeneity in health across individuals. That is, ω is fixed at zero, and we do not allow μ_{it}, σ_{it}, or κ_{it} to vary with observable individual demographics. However, unobservable heterogeneity in household health (through σ_μ) is still permitted. In column 2, we introduce observable heterogeneity in health. A key difference across columns 1 and 2 is the magnitude of the adverse selection parameter σ_μ, which falls by more than half. When rich observable heterogeneity in health is introduced to the model, the estimated amount of unobservable heterogeneity in health falls substantially. Moral hazard is introduced in column 3. Here, an important difference is the increase in the estimated amount of risk aversion. With moral hazard, the model can explain a larger part of the dispersion in spending for observably similar households. This implies that households are facing less uncertainty than previously thought, and that more risk aversion is necessary to explain the same plan choices. Because estimated risk aversion increases, the relative valuation of premiums and out-of-pocket costs (α^{OOP}), which had been compensating for low risk aversion, falls.

In column 3, we estimate an average moral hazard parameter ω of $1,115 among individuals and $1,542 among families.[45] Recall that ω represents the additional total spending that would be induced by moving a household from no insurance to full insurance. For scale, average total spending is $4,702 for individual households and $11,044 for families. Our estimates imply that moving from a plan with a 50 percent coinsurance rate to full insurance would result in an increase in total healthcare spending equal to 11 percent of mean spending for individuals and 7 percent for families.

We estimate a large degree of risk aversion. Our estimates imply a mean (median) coefficient of absolute risk aversion of 1.12 (0.84) across households.[46] Put differently, to make households indifferent between (i) a payoff of zero and (ii) an equal-odds gamble between gaining $100 and losing $X, the mean (median) value of $X in our population is $90.17 ($92.94).[47] We note, however, that our estimates of risk aversion are with respect to both financial risk *and* riskiness in the value from healthcare utilization (through b^*_{jt}), so they are not directly comparable to estimates that consider only financial risk. The standard deviation of the uncertain portion of payoffs ($b^*_{jt} - c^*_{jt}$) with respect to the distribution of health states is $853 on average across household-plan-years. The standard deviation of

[45] For comparison, the average ω estimated by Einav et al. (2013) is $1,330.

[46] We measure monetary variables in thousands of dollars; dividing our estimated coefficients of absolute risk aversion by 1,000 makes them comparable to estimates that use risk measured in dollars.

[47] A risk-neutral household would have $X equal to $100, and an infinitely risk-averse household would have $X equal to $0. Using the same example, Handel (2013) reports a mean $X of $91.0; Einav et al. (2013) report a mean $X of $84.0; and Cohen and Einav (2007) report a mean $X of $76.5.

out-of-pocket costs alone (c^*_{jt}) is \$1,358. To avoid a normally distributed lottery (in units of $b^*_{jt} - c^*_{jt}$) with mean zero and standard deviation \$853, the median household would be willing to pay \$305.

The importance of unobserved heterogeneity varies for health, risk aversion, and moral hazard.[48] The estimated amount of private information about health is fairly small once we account for the full set of household observables and moral hazard: Unobserved heterogeneity in μ_{kt} accounts for 8 percent of the total variation in μ_{kt} across household-years.[49] Unobserved heterogeneity in the moral hazard parameter accounts for 9 percent of its total variation across households. On the other hand, unobserved heterogeneity in risk aversion accounts for 54 percent of its total variation.

Conditional on observables, we find that households that are idiosyncratically risk averse are also idiosyncratically less prone to moral hazard ($\rho_{\psi,\omega} < 0$) and also have private information that they are unhealthy ($\rho_{\mu,\psi} > 0$). We find that households with private information that they are unhealthy are also idiosyncratically more prone to moral hazard ($\rho_{\mu,\omega} > 0$). Accounting for both unobservable *and* observable variation, we find that risk aversion and moral hazard have a strong negative correlation of -0.90. Among households with (without) children, expected health state $\mathbb{E}[\tilde{l}]$ has a correlation of 0.15 (0.13) with risk aversion, and a correlation of 0.05 (0.08) with the moral hazard parameter. Figure A.4 plots the unconditional joint distribution of these three key dimensions of household type.

Our estimates imply substantial disutility from switching insurers and plans. Average disutility across households from switching insurers is \$6,372, with a standard deviation of \$91. Average disutility from switching plans (but not insurers) is \$4,466, with a standard deviation of \$1,739. We estimate that insurer inertia is increasing in household risk score, and that plan inertia is increasing in household age and is on average \$188 higher for households with children.[50] The exceptionally large magnitudes of our inertia coefficients reflect, in large part, the infrequency with which households switch plans and insurers, as shown in Table 1. Only 3.3 percent of household-years ever voluntarily switch insurers, and only 13.6 percent of household-years ever voluntarily switch plans.

---

[48]Following Revelt and Train (2001), we derive each household's posterior type distribution using Bayes' rule, conditioning on their observed choices and the population distribution. For the purposes of examining total variation in types across households (accounting for both observed and unobserved heterogeneity), we assign each household the expectation of their type with respect to their posterior distribution. This procedure is described in detail in Appendix B.3.

[49]Limited selection on unobservables is consistent with the findings of Cardon and Hendel (2001).

[50]We do not investigate the micro-foundations of our estimates of household disutility from switching; see Handel (2013) for a full treatment of inertia in health insurance.

Finally, the estimates in column 3 indicate that households weight out-of-pocket expenditures 34.8 percent more than plan premiums. We believe this could be driven by a variety of factors, including (i) household premiums are tax deductible, while out-of-pocket expenditures are not, and (ii) employee premiums are very low (at the median, zero), perhaps rendering potential out-of-pocket costs in the thousands of dollars relatively more salient. A single household in Oregon with an income of $80,000 paid an effective state plus federal income tax rate of 28.9 percent in 2013. Using this tax rate, a dollar of out-of-pocket spending (after tax) would be equivalent to 1.41 dollars of premiums (pre-tax). We also find that households value a dollar in HSA/HRA contributions on average 75 percent less than a dollar of premiums. This is consistent with substantial hassle costs associated with these types of accounts, as documented by Reed et al. (2009) and McManus et al. (2006).

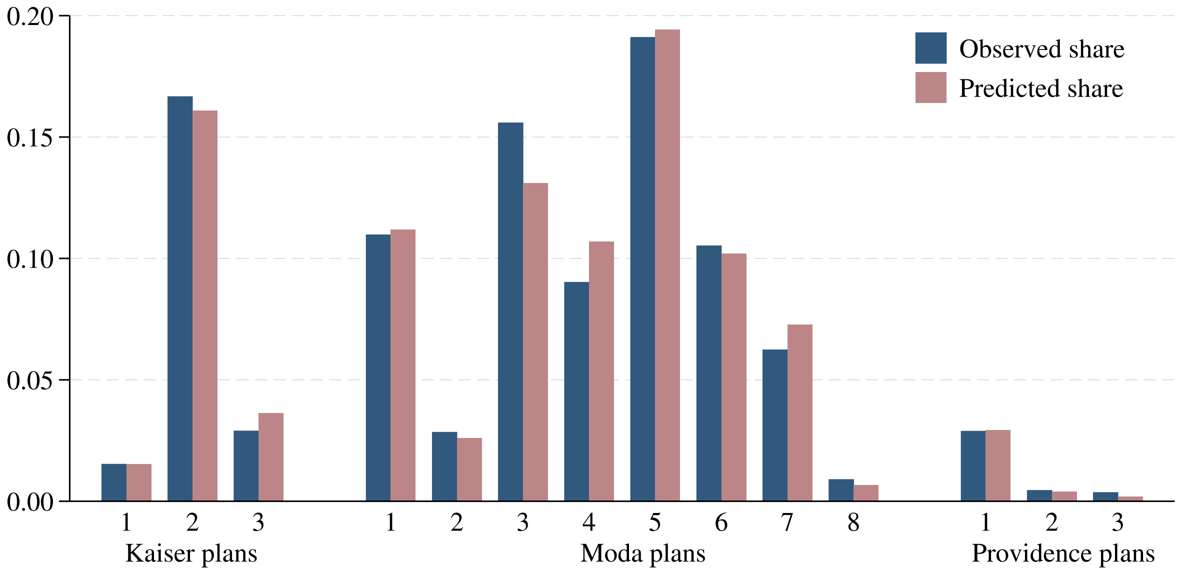
Model Fit. We conduct two procedures to evaluate model fit, corresponding to the two stages of the model. First, we compare households' predicted plan choices with those observed in the data. Figure 3 displays the predicted and observed market shares for each plan, pooled across all years in our sample.[51] Shares are matched exactly at the insurer level due to the presence of insurer fixed effects, but are not matched exactly plan by plan. Predicted choice probabilities over plans within an insurer are driven by plan prices, inertia, and households' valuation of different levels of coverage through their expectation of out-of-pocket spending, their value of risk protection, and their expectation of utility from the consumption of healthcare services. Given the relative inflexibility of the model with respect to household choice of coverage level within an insurer, the fit is quite good.

In our second exercise, we compare the predicted distributions of households' total healthcare spending to the distributions of total healthcare spending we observe in the data. In a given year, each household faces a predicted distribution of health states and a corresponding plan-specific distribution of total healthcare spending, as defined by our model and estimated parameters. To construct the predicted distribution of total spending in a population of households, we take a random draw from the predicted distribution of each household corresponding to the household's chosen plan. Figure 4 presents kernel density plots of the predicted and observed distributions of total spending among household-years enrolled by each insurer.[52] The vertical lines in each plot represent the mean of the respec-

[51] Figure A.5 provides corresponding comparisons separately for each year. As another metric, the model predicts 72 percent of household plan choices correctly (i.e., assigns the highest predicted probability to the correct plan). If households were modeled as choosing randomly from their plan choice set, 23 percent of plan choices would be predicted correctly (i.e., the average choice set size is approximately four plans).

[52] The distributions shown are conditional on predicted/observed spending greater than zero. The observed

Figure 3. Model Fit: Plan Choices



Notes: The figure shows predicted and observed market shares at the plan level. All years are pooled, so an observation is a household-year. Predicted shares are calculated using the estimates in column 3 of Table 3.

tive distribution. Overall, across all household-year observations, average total healthcare spending is observed to be \$10,754 and predicted to be \$10,738. Figure A.6 presents similar comparisons by family size and quartile of household risk score. The spending distribution fit is good both overall, as well as in subsamples of households, reflecting our flexible modeling approach for individual and household spending distributions.

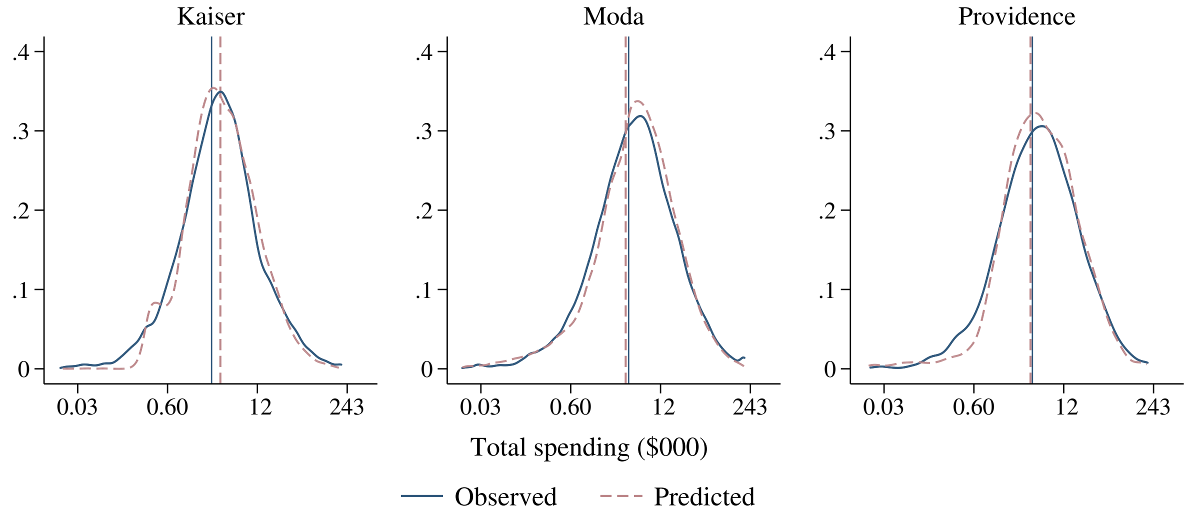
V.B Willingness to Pay and Social Surplus

We can now construct each household's willingness to pay for different levels of coverage, as well as the social surplus generated by each household's allocation to different levels of coverage. Our focus is on whether, in this population, it is efficient to offer vertical choice. We consider a set of vertically differentiated contracts that have a deductible, coinsurance rate, out-of-pocket maximum design. The lowest level of coverage we consider is a Catastrophic contract with a deductible and out-of-pocket maximum of \$10,000.[53]

probability of zero household spending is 3.7 percent, while the predicted probability is 4.7 percent.

[53] We place a lower bound on coverage level because, in reality, three assumptions implicit in our theoretical model would be violated for very low coverage (or no coverage). Namely, we have assumed (i) the full cost of healthcare m does not vary across contracts, (ii) consumers will never utilize "too little" healthcare, and (iii) consumer wealth exceeds all potential out-of-pocket cost realizations. There are clear and important violations of each of these assumptions at very low coverage in the real world. While we argue that ours is the right model for considering choice over coverage levels in a relatively high range, we readily admit that it is not well equipped to make normative comparisons of good insurance to little or no insurance. In

Figure 4. Model Fit: Healthcare Spending



Notes: The figure shows kernel density plots of the predicted and observed distribution of total healthcare spending on a log scale, separately among households enrolled with each of the three insurers, conditional on predicted/observed spending greater than zero. All years are pooled, so an observation is a household-year. Vertical lines represent the mean of the respective distribution. Predicted distributions are estimated using parameter estimates from column 3 in Table 3.

The highest level of coverage is full insurance. We consider five "evenly spaced" contracts spanning this range: Since they roughly correspond to the levels of coverage offered on Affordable Care Act (ACA) exchanges, we refer to them as full insurance, Gold, Silver, Bronze, and Catastrophic.[54] Their actuarial values are 1.00, 0.84, 0.72, 0.61, and 0.53; their out-of-pocket cost functions are depicted in Figure A.7. We focus our graphical analysis and discussion of results on only these five contracts because they can tractably convey our main findings. In Section C.1, we revisit the specification of the set of potential contracts by adding more contracts and by considering alternative contract designs.[55]

Willingness to Pay. We make several simplifications to our empirical model in order to map it from our setting in Oregon back to our theoretical model, while maintaining parameterizations and estimated distributions of consumer types. To start, we put aside intertemporal variation in households' distributions of health states and focus on the first year each household appears in the data. We also use the provider price parameter $\phi = 1$ (corresponding to that of Moda). This leaves each household with a single type: $\{F_k, \psi_k, \omega_k\}$, where F_k

truth, we believe such comparisons involve considerations that would stretch the limits of any economic model.

[54]The deductibles, coinsurance rates, and out-of-pocket maximums for the intermediate contracts are $1,169, 21%, $2,564 for Gold; $3,060, 34%, $4,872 for Silver; and $5,771, 48%, $7,436 for Bronze.

[55]We do not consider the set of Moda plans because they are not truly vertically differentiated (see Figure A.2). While this poses no particular problem for estimation, it means they are not the relevant subject for our research question in counterfactuals.

is a shifted lognormal distribution described by parameters $\{\mu_k, \sigma_k, \kappa_k\}$.[56] With respect to payoffs (equation (6)), we (i) hold all nonfinancial features fixed, so any insurer fixed effects cancel; (ii) suppose households choose from the new menu of contracts for the first time, making inertia irrelevant; (iii) assume the idiosyncratic shock is not utility-relevant;[57] and (iv) set α^{OOP} to one so that premiums and out-of-pocket costs are valued one-for-one.[58]

Willingness to pay for marginally more generous insurance is equal to the difference in certainty equivalent between a (higher coverage) focal contract j and a (lower coverage) reference contract j_0, when both have price zero. Certainty equivalents are given by

$$\begin{aligned} CE_{kj} &= -\psi_k^{-1} \log(-U_{kj}) \\ &= \bar{x}_{kj} - \psi_k^{-1} \log \left(\int \exp\left(-\psi_k(x_{kj}(l) - \bar{x}_{kj}\right) dF_k(l) \right), \end{aligned}$$

where $x_{kj}(l)$ is the payoff associated with health state l in contract j, and $\bar{x}_{kj}$ is the expectation of $x_{kj}(l)$ with respect to the distribution of l. With attention restricted to the dimension of coverage level, willingness to pay depends only on the benefit of healthcare spending b, out-of-pocket costs c, and riskiness in both:

$$\begin{aligned} WTP_{kj} &= CE_{kj} - CE_{k,j_0} \\ &= \bar{c}_{k,j_0} - \bar{c}_{kj} + \bar{b}_{kj} - \bar{b}_{k,j_0} + \Psi_{kj}, \end{aligned}$$

where $\bar{c}_{kj}$ is the expectation of out-of-pocket costs $c_j(m_j^*(l, \omega_k))$ with respect to the distribution of l, and $\bar{b}_{kj}$ is similarly defined. As in our theoretical expression for WTP, we pull out the mean and lump deviations into Ψ_{kj}, the value of risk protection. Whereas our theoretical reference contract x_0 was the null contract, our empirical reference contract j_0 is the Catastrophic contract. We hereinafter refer to "willingness to pay" for a given contract,

---

[56]We assign household types by integrating over each household's posterior distribution of types. We likewise calculate household-specific willingness to pay and social surplus using this procedure. We omit these steps in this section because the notation is cumbersome, but it is provided in Appendix B.3.

[57]Our model allows for rich heterogeneity in preferences over financially differentiated contracts, so we are comfortable with the interpretation that any remaining choice determinants contained in ϵ can be considered "mistake-making" (Ketcham et al., 2012; Handel and Kolstad, 2015; Bhargava, Loewenstein and Sydnor, 2017) or "monkey-on-the-shoulder tastes" (Akerlof and Shiller, 2015), and so can be omitted from the social welfare calculation. We take this approach to get back to vertical differentiation. The counterfactual thought experiment becomes one of supposing consumers had access to a tool that would perfectly aid them in avoiding mistakes and expressing their true preferences. The question is whether we need such a tool.

[58]Otherwise, welfare could be created by moving a dollar of spending between premiums and out-of-pocket cost, which we find undesirable. If we leave α^{OOP} as estimated, optimal levels of coverage increase, since out-of-pocket costs are so disliked.

but bear in mind that this is *marginal* willingness to pay with respect to this particular reference point.

Figure 5 presents the distribution of willingness to pay among family households.[59] Households are ordered on the horizontal axis according to their willingness to pay. As in a demand curve, the highest willingness-to-pay households are on the left. Figure 5, as well as the figures that follow, is composed of connected binned scatter plots. Households at each percentile of willingness to pay are binned together, and the average value of the vertical axis variable is plotted. These 100 points are then connected with a line.[60] The left panel shows the willingness to pay curves for our candidate contracts. As the plans are vertically differentiated, all households are willing to pay more for higher coverage. The highest willingness-to-pay households are willing to pay \$10,000 more for full insurance than for Catastrophic.

As in equation (2), willingness to pay can be decomposed into three parts: expected reduced out-of-pocket cost holding behavior fixed (the "transfer"), expected payoff from moral hazard spending, and the value of risk protection. Recall that only the latter two components are relevant to social welfare. The right panel of Figure 5 presents this decomposition for the Gold plan. We find that the transfer represents the majority of willingness to pay for most households.[61] However, this varies across the distribution of willingness to pay: For households with the lowest willingness to pay, only one third is made up by the transfer, while for households with the highest willingness to pay, nearly all is made up by the transfer. These highest willingness-to-pay households are willing to pay a \$7,500 premium just to avoid paying \$7,500 in expected out-of-pocket costs. Importantly, this means that allocating them to higher coverage does not generate any social surplus.

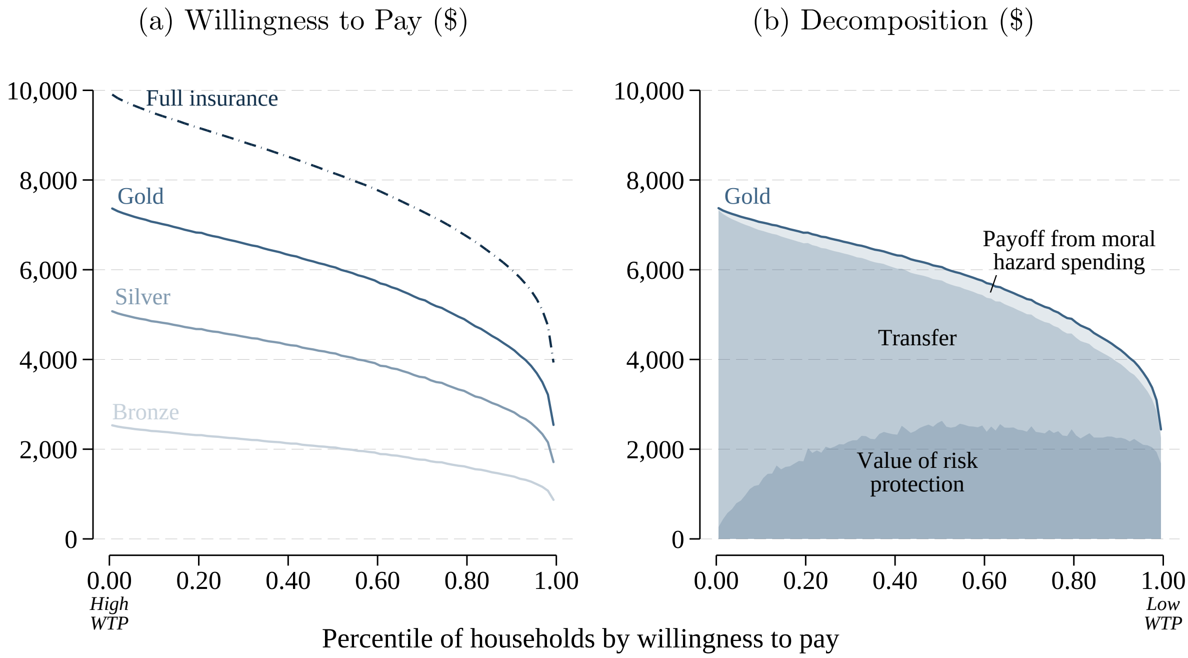
Social Surplus. With willingness to pay, we can determine households' privately optimal choices given any premiums. We next specify socially optimal choices. As in Section II, the social surplus generated by allocating a household to a given contract is the difference

[59] We focus on family households because families make up 75 percent of the sample and because our set of potential contracts is chosen to mimic the coverage levels offered to families.

[60] Households are in fact ordered by their willingness to pay for full insurance. Because their ordering is nearly identical across contracts, the lines in the left panel are monotonically decreasing and appear smooth (otherwise, they would be jagged). The consistent ordering of households across contracts is what permits a graphical analysis of multiple contracts analogous to the two-contract example in Figure 1. To illustrate the consistency of the ordering, Figure A.8 shows a household-level plot of willingness to pay.

[61] In an interesting parallel, this result corresponds to findings by Einav et al. (2020) that the transfer makes up the majority of hospitals' private incentives to participate in a bundled payment program.

Figure 5. Willingness to Pay



Notes: The figure shows the distribution across households of (a) willingness to pay and (b) the decomposition of willingness to pay. The left panel consists of four connected binned scatter plots, with respect to 100 bins of households ordered by willingness to pay. The right panel consists of three connected binned scatter plots, with the area between each line shaded to indicate the component represented. Both willingness to pay and its components are measured in dollars relative to the Catastrophic plan.

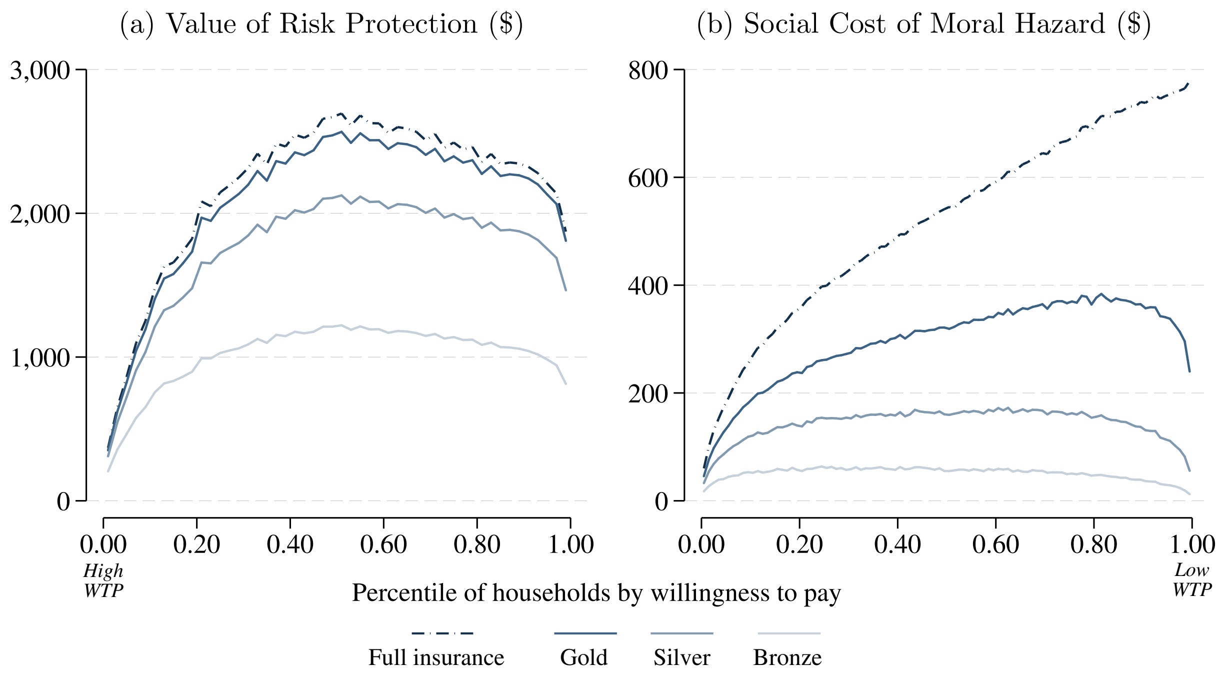
between their willingness to pay and expected insurer cost:

$$SS_{kj} = \underbrace{\Psi_{kj}}_{\substack{\textit{Value of risk}\\ \textit{protection}}} - \underbrace{\left((\bar{k}_{kj} - \bar{k}_{k,j_0}) - (\bar{c}_{k,j_0} - \bar{c}_{kj} + \bar{b}_{kj} - \bar{b}_{k,j_0})\right)}_{\substack{\textit{Social cost}\\ \textit{of moral hazard}}},$$

where $\bar{k}_{kj}$ is the expectation of insured spending $k_j(m_j^*(l, \omega_k))$ with respect to the distribution of l. The value of risk protection varies in the population to the extent there is variation in risk aversion and in the probability that households realize health states that would result in different levels of out-of-pocket cost across contracts. The social cost of moral hazard varies in the population to the extent that there is variation in the moral hazard parameter and in the probability that households realize health states that would result in different marginal out-of-pocket cost across contracts.

To understand the contribution of each of these components to the overall relationship between willingness to pay and social surplus, we first plot them separately. Figure 6a shows the distribution across households of the value of risk protection. We find that the majority of the social welfare gains from more generous insurance are driven by households with intermediate levels of willingness to pay. This "shape" of risk protection could be driven by either the distribution across households of risk aversion or of risk. We investigate by examining the joint distribution of risk aversion and willingness to pay (see Figure A.9a). While there is substantial variation in risk aversion, average risk aversion is monotonically increasing in willingness to pay. The inverted U-shape in Figure 6a must therefore be driven by the shape of risk.[62]

Figure 6. Components of Social Surplus



Notes: The figure shows the distribution across households of (a) the value of risk protection and (b) the marginal social cost of moral hazard; both are relative to the Catastrophic contract. Each panel is composed of four connected binned scatter plots, with respect to 50 bins of households ordered by willingness to pay.

The inverted U-shape of risk is driven by the concavity of the contracts we consider. Very sick households (or households with many children) are very likely to realize health states above the out-of-pocket maximum of every possible contract, leaving them little uncertainty about out-of-pocket costs. Very healthy households are very likely to realize health states below all deductibles, rendering the contracts roughly identical in both uncertainty and expectation. The households that do face variation across contracts in uncertainty about out-of-pocket costs are those for which much of the density of their health state distribution lies in the range in which out-of-pocket costs vary across both contracts and health states.[63]

[62] In Appendix C.2, we further investigate the importance of preference variation to our main results.

[63] Figure A.10 confirms this pattern by showing the distributions of health states faced by households across the distribution of willingness to pay.

Figure 6b shows the distribution of the social cost of moral hazard. It provides two important insights. First, high willingness-to-pay households on average barely change their behavior across this range of coverage levels.[64] While they may have already been consuming more healthcare in the Catastrophic contract than they would have absent insurance, the marginal effect of higher coverage is minimal. On the other hand, households with low willingness to pay do, on average, change their behavior over this range of coverage levels. Overall, this pattern is driven by the interaction of the health state distributions and concave contracts (treatment intensity), as well as by the fact that moral hazard parameters are decreasing in willingness to pay (treatment effect).[65] The second insight is that the Gold contract can recover about half of the social cost of moral hazard induced by full insurance. The $1,000 deductible is enough to undo much of the full insurance social cost of moral hazard, while, as seen in Figure 6a, giving up only a small amount of risk protection.

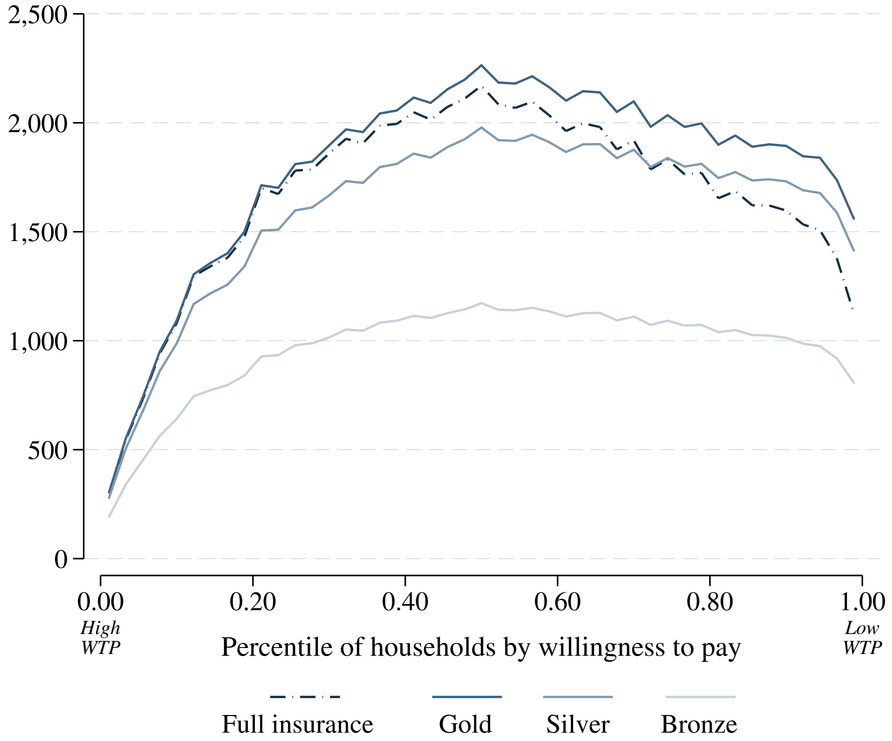
Finally, we construct the social surplus curves by subtracting Figure 6b from Figure 6a. Figure 7 presents the marginal social surplus generated by allocating households to each contract relative to the Catastrophic contract. The plot consists of four connected binned scatter plots, with respect to 50 (to reduce noise) quantiles of willingness to pay. At each quantile, the curves measure the average social surplus generated if all households at that quantile were allocated to a given contract. Since households can be screened only by their willingness to pay, this is exactly what is relevant for determining optimal prices.

Social surplus curves for all contracts lay everywhere above zero, meaning that Catastrophic is the worst contract from a social welfare perspective at any level of willingness to pay. The Bronze plan is strictly second worst. Of the others, the Gold contract generates weakly greater surplus than any other contract at every level of willingness to pay. This figure is the empirical analog of the theoretical examples given in Figure 1. The Catastrophic plan is the "low" contract, and the four others are "high" contracts. Vertical choice should only be offered if consumers with higher willingness to pay should have higher coverage. As in the theoretical example, this statement corresponds to a crossing of upper-envelope social

[64] This finding is driven by the embedded assumption that moral hazard will not be expressed as long as end-of-year marginal out-of-pocket cost does not vary across contracts. While there is substantial empirical evidence that consumers do respond to spot prices (e.g. Aron-Dine et al., 2015; Dalton, Gowrisankaran and Town, 2015), here we do not find evidence of moral hazard among high-risk households, which were very likely to hit even the highest out-of-pocket maximums (see Table A.8). Even so, if the data did support a moral hazard response among these households, the model would load the effect onto the moral hazard parameter ω, compensating a weak treatment with a stronger treatment effect.

[65] Variation in treatment intensity can be inferred from the health state distributions at different levels of willingness to pay, shown in Figure A.10. Variation in treatment effect can be seen in the relationship between the moral hazard parameter and willingness to pay, shown in Figure A.9b.

Figure 7. Social Surplus ($)



Notes: The figure shows the distribution across households of social surplus relative to the Catastrophic contract. The figure is composed of four connected binned scatter plots, with respect to 50 bins of households ordered by willingness to pay.

surplus curves, with the higher-coverage contract to the left. Here, the upper envelope of social surplus curves is composed of a single contract. An efficiency-maximizing regulator would find it optimal to forgo choice and offer only the Gold contract.[66]

While Gold is the efficient contract *on average* at every level of willingness to pay, it is not the efficient contract for every household. Figure A.11 shows the heterogeneity in households' efficient contracts. Full insurance is the efficient contract for 20 percent of households, Gold is efficient for 70 percent, Silver is efficient for 10 percent, and Bronze is efficient for less than 1 percent of households. While efficient coverage level does vary, it is not predicted well by willingness to pay. The optimal feasible allocation under community-rated prices would therefore achieve social surplus equal to the integral of the Gold social surplus curve.

[66]Figure 7 is nearly all the information one would need to make this determination, but it is not all because households are not perfectly consistently ordered in willingness to pay across contracts. We therefore also formally confirm by numerically solving for optimal prices.

VI Counterfactual Pricing Policies

We compare outcomes under five pricing policies: (i) regulated pricing with community rating, (ii) regulated pricing with type-specific prices, (iii) competitive pricing with community rating, (iv) competitive pricing with type-specific prices, and (v) subsidies to support full vertical choice. Regulated pricing is the baseline policy considered in this paper, in which the regulator can observe the distribution of consumer types and can set premiums. Competitive pricing is the case in which competition among private firms drives premiums to equal average costs on a plan-by-plan basis, rendering the market susceptible to unraveling due to adverse selection. Subsidies to support full vertical choice is a policy whereby prices are set with the intention of supporting the availability of (read: enrollment in) every contract.

We consider two scenarios, (ii) and (iv), in which premiums can vary by consumer attributes. If observable dimensions of household type are predictive of their efficient coverage level, allowing plan menus to be tailored to specific types may improve allocations. We divide households into four groups: childless households under age 45, childless households over age 45, households under age 45 with children, and households over age 45 with children.[67] We use age and whether the household has children because these are used in practice on ACA exchanges and are also important observables with which the parameters of our model may vary.

Welfare Outcomes. Table 4 summarizes outcomes under each of our five pricing policies. It shows the percent of households Q enrolled in each contract at the optimal allocation feasible under the policy, the percent of first-best social surplus achieved, and the expected insurer cost per household AC among households in each contract (measured in thousands of dollars). Social surplus is normalized to zero for the Catastrophic contract. We benchmark outcomes against the first-best allocation of households to contracts (as depicted in Figure A.11).[68] The first-best allocation generates \$1,802 in social surplus per household relative to the counterfactual of allocating all households to Catastrophic. Expected total healthcare spending per household at the first-best allocation is \$12,090, and expected insurer cost per household is \$10,387.

[67]Among family households (anything except the employee alone), 6 percent are childless and under age 45, 26 percent are childless and over age 45, 53 percent have children and are under age 45, and 15 percent have children and are over age 45.

[68]This allocation cannot be supported by prices unless prices can vary by all aspects of consumer type, including risk aversion and the moral hazard parameter.

Table 4. Outcomes of Alternative Pricing Policies

	Policy	% of First Best Surplus		Potential Contracts Full	Gold	Silver	Bronze	Ctstr.
*	First best	1.000	*Q:*	0.20	0.70	0.10	<0.01	–
			AC:	13.63	8.66	15.76	31.90	–
(i)	Regulated pricing with community rating	0.965	*Q:*	–	1.00	–	–	–
			AC:	–	10.18	–	–	–
(ii)	Regulated pricing with type-specific prices	0.977	*Q:*	0.27	0.73	–	–	–
			AC:	14.25	9.50	–	–	–
(iii)	Competitive pricing with community rating	–	*Q:*	–	–	–	–	1.00
			AC:	–	–	–	–	6.15
(iv)	Competitive pricing with type-specific prices	0.285	*Q:*	–	–	0.06	0.26	0.68
			AC:	–	–	4.00	8.62	5.95
(v)	Subsidies to support vertical choice	0.782	*Q:*	0.01	0.07	0.63	0.28	0.01
			AC:	60.01	31.26	8.17	1.89	0.24

Notes: The table summarizes outcomes under the five pricing policies we consider as well as the first-best outcome, among the 32,382 family households. Q represents the percent of households enrolled in each plan; AC represents average expected insurer cost (in thousands of dollars) among households enrolled in a given plan. At the first-best allocation, per-household social surplus is $1,802, and average expected insurer cost is $10,387. Social surplus is normalized to zero at the Catastrophic contract.

Alternative (i) is our baseline policy, in which the regulator can set prices but is restricted to community rating. As indicated by Figure 7, under this scenario it is welfare maximizing to offer only Gold. The average expected insurer cost of all households in the Gold contract is $10,182. Interestingly, although 30 percent of households are misallocated, this policy generates 96.5 percent of the welfare generated by the first-best allocation. Among the households for whom the Gold plan was not optimal, there is little variation in social surplus between the three most generous contracts.[69] Among all households, the welfare gains from more generous insurance are similarly flat among the top contracts: If the regulator were to put everyone in a single contract, the percent of first-best surplus achieved by Bronze is 53 percent, by Silver 89 percent, by Gold 96 percent, and by full insurance 93 percent. In dollars, the per-household welfare gain from moving all households from Bronze to Silver is $649, while the gain from Silver to Gold is only $126.

Because pricing policy (i) is almost as efficient as the first-best outcome, there is little scope for improvement by varying prices by consumer types in alternative (ii). Even so,

[69]This is in particular true among the 20 percent of households for whom the optimal contract is full insurance. Allocating these households instead to Gold generates 99 percent of the social surplus achieved by full insurance. These households have almost none of their potential spending in the range over which Gold and full insurance differ in marginal out-of-pocket cost.

we do find that allowing the regulator to discriminate can improve allocational efficiency. Among households without children, it is efficient to offer choice between full insurance and Gold. Among households with children, it is again efficient to offer only Gold. It becomes possible to productively offer full insurance because the high willingness-to-pay households with children, to whom it is not efficient to provide such high coverage, can be excluded.

Alternative (iii) considers competitive pricing with community rating. We calculate the competitive equilibrium proposed by Azevedo and Gottlieb (2017).[70] We find that in this population, a separating equilibrium cannot be supported, and the market fully unravels to the Catastrophic contract. The associated premium and expected insurer cost per household is \$6,151. While choice is permitted under this policy, the market cannot deliver it. Alternative (iv) considers the allocations that could be supported under competitive pricing if the market could be segmented. We find that childless households below age 45 can support a pooling equilibrium at the Silver contract, and childless households above age 45 can support a pooling equilibrium at Bronze. Both markets for households with children still unravel to Catastrophic coverage.

The first four policies are natural benchmarks, but none turn out to feature vertical choice. Choice is banned under regulated pricing, and it is prevented by adverse selection under competitive pricing. But in reality, vertical choice does exist. It is sustained in U.S. health insurance markets, including in the market we study in Oregon, in large part by a variety of subsidy and tax policies. To mimic this status quo outcome, policy (v) implements premiums that can support enrollment in every contract. We target enrollment shares that match the true metal-tier shares observed on ACA exchanges in 2018.[71] The targeted shares are those shown in Table 4. The premiums that can support these shares and break even in aggregate are \$13,571 for full insurance, \$11,034 for Gold, \$8,805 for Silver, \$6,991 for Bronze, and \$6,035 for Catastrophic.[72] Because households with mid-range willingness to pay (for whom social surplus increases steeply at low coverage levels; see Figure 7) now choose Silver instead of Bronze or Catastrophic, this allocation substantially increases welfare relative to the competitive outcome.

[70] Like the authors, we use a mass of behavioral consumers equal to 1 percent of the population of households. See Azevedo and Gottlieb (2017) for additional details.

[71] Shares are pulled from Kaiser Family Foundation "Marketplace Plan Selections by Metal Level," available at https://www.kff.org/health-reform/state-indicator/marketplace-plan-selections-by-metal-level. We map Platinum coverage to full insurance.

[72] We refer to this policy as "subsidies" to support vertical choice because it can be thought of as mimicking a reality in which the regulator announces consumer subsidies it will provide for each contract, and then private firms compete. The subsidy (or tax) that would need to be announced for each contract is equal to the difference between the desired premium and the resulting average cost (see Table 4).

Distributional Outcomes. The population faces an unavoidable healthcare spending bill of \$11,359 per household. It is unavoidable because it arises even if all households have the least generous insurance (Catastrophic). While full insurance provides the benefit of additional risk protection, it also raises the population's healthcare spending bill due to moral hazard, to \$12,410 per household.

The spending bill is funded by a combination of out-of-pocket costs and insured costs. Insured costs are in turn funded by premiums or taxes. We do not distinguish between the two: An increase in premiums on all contracts by \$5 is equivalent to a tax of \$5.[73] If all households had Catastrophic coverage, in expectation 47 percent of the spending bill would be paid out-of-pocket, and 53 percent would be insured. If all households had full insurance, 100 percent of spending would be insured. There are therefore large differences between policies in the source of funding for the population's healthcare spending bill, and in turn, how evenly the spending bill is shared across households. If all households had full insurance, the spending bill would be split perfectly evenly in the population.[74] If all households had no insurance, each household would pay their own expected cost.

Figure 8 shows distributional outcomes under three of our candidate policies: (i) regulated pricing ("All Gold"), (iii) competitive pricing ("All Catastrophic"), and (v) subsidies to support vertical choice ("Vertical Choice"). Panel (a) shows the distribution of households' (expected) healthcare spending bill, premium plus expected out-of-pocket cost, given households' chosen contracts. Households are again ordered on the horizontal axis according to their willingness to pay.[75] For example, under "All Catastrophic," the premium is \$6,151, and the highest willingness-to-pay households have expected out-of-pocket costs of \$9,578, implying a healthcare spending bill of \$15,729. The lowest willingness-to-pay households have expected out-of-pocket costs of \$1,238, implying a healthcare spending bill of \$7,389. The population's healthcare spending bill is split more evenly when households have higher coverage.

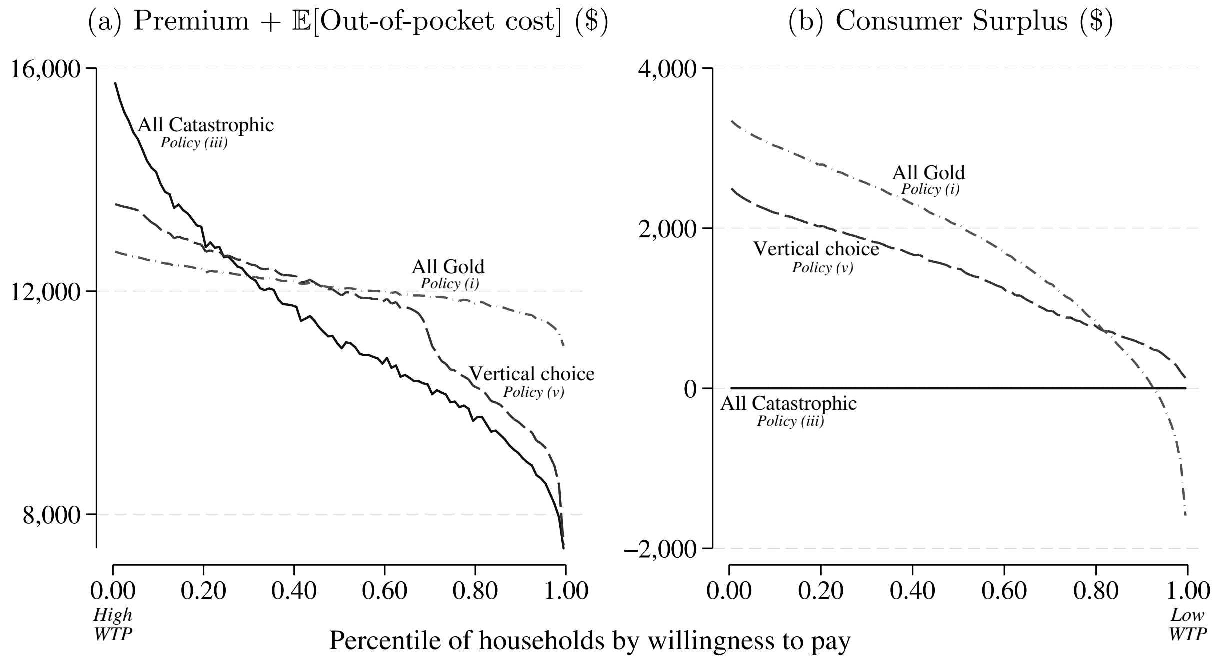
Panel (b) shows the distribution of consumer surplus. For each household, consumer surplus is the difference between their marginal willingness to pay and their marginal premium. The marginal premium is the difference between the premium of the chosen contract under

[73] The equivalence is due to our assertion that consumers choose exactly one contract. This could result from a sufficiently compelling mandate, or a direct tax on incomes that precludes the ability to opt out.

[74] The premium for a single contract would in reality be assessed as a tax. In that case, premiums would not be split evenly, but according to the prevailing income tax system.

[75] All households at a particular level of willingness to pay choose the same plan and thus pay the same premium, but there is still variation in expected out-of-pocket cost among households at a given level of willingness to pay. The plot is therefore a connected binned scatter plot, similar to the previous figures.

Figure 8. Distributional Outcomes



Notes: The figure shows the distribution across households of (a) premiums plus expected out-of-pocket costs and (b) consumer surplus, under three policies considered in Table 4. Since willingness to pay is relative to the Catastrophic contract, consumer surplus is similarly marginal. Consumer surplus equals marginal willingness to pay less marginal premium. The premium for the single plan is $6,151 under "All Catastrophic" and $10,182 under "All Gold." Premiums under "Vertical choice" are $13,571 for full insurance, $11,034 for Gold, $8,805 for Silver, $6,991 for Bronze, and $6,035 for Catastrophic. Households are arranged on the horizontal axis according to their willingness to pay. Panel (a) is composed of three connected binned scatter plots, while panel (b) is composed of true line plots.

the focal policy, and the premium of the Catastrophic contract when all households are allocated to it ($6,151). Since willingness to pay is measured relative to the Catastrophic contract, consumer surplus is similarly marginal. While expected healthcare spending may vary across households at a given level of willingness to pay, consumer surplus does not: Willingness to pay uniquely determines both components of consumer surplus.[76] The sum of consumer surplus across households is the total social surplus generated by a given policy. The difference between the "All Gold" consumer surplus curve in Figure 8b and the Gold contract's social surplus curve in Figure 7 is that the former shows who receives the surplus, while the latter shows who generates it; the integrals of the two curves are the same.

We find that 91 percent of households prefer optimal regulation under policy (i) to the alternative of an unregulated (and unraveled) market. We find that all households prefer vertical choice under policy (v) to the alternative of an unregulated market. Strikingly, we

[76] A consumer's type θ in turn uniquely determines willingness to pay. Figure 8b shows the relative value of being assigned to different levels of willingness to pay, perhaps in some prior lottery. Figures A.9 and A.10, which describe the types of households at different levels of willingness to pay, give a sense of where different draws from θ would place a household in willingness to pay.

also find that 82 percent of households prefer optimal regulation to vertical choice. While a shift to optimal regulation from vertical choice would make 18 percent of households worse off, only 7 percent of households would be at least $500 worse off. The shift would raise welfare by $330 per household per year.

VII Conclusion

This paper presents a framework for evaluating the efficiency of choice over coverage levels in health insurance markets. Our framework incorporates consumer heterogeneity along multiple dimensions, endogenous healthcare utilization through moral hazard, and menus of nonlinear insurance contracts among which traded contracts are endogenous. We emphasize the importance of distinguishing between the components of willingness to pay that are only redistributive and the components that generate social surplus from insurance. We also emphasize that the redistributive component plays a large role in determining feasible allocations. Health is persistent, but contracts (at least in the U.S.) often span only a short time.[77] The implication is that a large part of insurable spending can be foreseen, and it may not be possible to align the private incentive to maximize one's own transfer and the social incentive to mitigate residual uncertainty. The presence of moral hazard means that the problem is more complicated than simply mandating full insurance for all.

We show that the key condition for vertical choice to be efficient is whether consumers with higher willingness to pay have higher efficient levels of coverage. In reverse, this implies that a lowest-coverage plan should only be offered if the lowest willingness-to-pay consumers should have it. The lowest coverage we consider is a "Catastrophic" high-deductible health plan. We find that low willingness-to-pay consumers are sufficiently risk averse and facing sufficient risk to warrant higher coverage, so we conclude that such low coverage should not be offered in this market. On the other hand, a highest-coverage plan should only be offered if the highest willingness-to-pay consumers should have it. The highest coverage we consider is full insurance, and we find that it would be more efficient for the high willingness-to-pay consumers to have lower coverage. Between these extremes, we find that private values for coverage are not positively correlated with social values, and thus that choice over coverage level should not be offered. The optimal plan menu is a single plan with an actuarial value of 84 percent. Reassuringly from a policy perspective, we also find that the welfare stakes

[77] Ghili et al. (2019) consider long-term contracts in health insurance markets. It would be interesting to consider the welfare effects of vertical choice in a dynamic setting.

of misallocation are low in the neighborhood of the optimal contract.

We focus our attention on a range of coverage levels over which uncertainty about healthcare utilization represents a purely financial gamble. Important considerations our model does not address arise when consumers face liquidity constraints (Ericson and Sydnor, 2018) or when consumers are protected from large losses by limited liability in addition to by insurance (Gross and Notowidigdo, 2011). These distortions would become more pronounced outside the range of coverage levels we consider, and it would be interesting to explore their effects on our conclusions. In addition, the socially optimal level of healthcare utilization in our model is the level a consumer would choose absent insurance. If healthcare providers charge supracompetitive prices, or if there are externalities with respect to healthcare utilization, it may be the case that using insurance to induce additional utilization is desirable. Finally, a central simplification in our model is that healthcare is a homogenous good over which consumers choose only the quantity to consume. In reality, healthcare is multidimensional, and the time and space over which utilization decisions are made is complex. We see the extension of our model to capture other dimensions of healthcare utilization as an important direction for future research.

References

Abaluck, Jason, and Jonathan Gruber. 2011. "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program." *American Economic Review*, 101(4): 1180–1210.

Abaluck, Jason, and Jonathan Gruber. 2016. "Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance." *American Economic Review*, 106(8): 2145–84.

Abaluck, Jason, and Jonathan Gruber. 2017. "Improving the Quality of Choices in Health Insurance Markets." *NBER Working Paper 22917.*

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

Akerlof, George A., and Robert J. Shiller. 2015. *Phishing for Phools: The Economics of Manipulation and Deception.* Princeton University Press.

Apesteguia, Jose, and Miguel A. Ballester. 2018. "Monotone Stochastic Choice Models: The Case of Risk and Time Preferences." *Journal of Political Economy*, 126(1): 74–106.

Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen. 2015. "Moral Hazard in Health Insurance: Do Dynamic Incentives Matter?" *The Review of Economics and Statistics*, 97(4): 725–741.

Arrow, Kenneth J. 1965. "Uncertainty and the Welfare Economics of Medical Care: Reply (The Implications of Transaction Costs and Adjustment Lags)." *The American Economic Review*, 55(1/2): 154–158.

Arrow, Kenneth J. 1968. "The Economics of Moral Hazard: Further Comment." *The American Economic Review*, 58(3): 537–539.

Azevedo, Eduardo M., and Daniel Gottlieb. 2017. "Perfect Competition in Markets With Adverse Selection." *Econometrica*, 85(1): 67–105.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. 2017. "Choose to Lose: Health Plan Choices from a Menu with Dominated Option*." *The Quarterly Journal of Economics*, 132(3): 1319–1372.

Bogin, Alexander, William Doerner, and William Larson. 2019. "Local House Price Dynamics: New Indices and Stylized Facts." *Real Estate Economics*, 47(2): 365–398.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. 2017. "What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics*." *The Quarterly Journal of Economics*, 132(3): 1261–1318.

Bundorf, M. Kate, Jonathan Levin, and Neale Mahoney. 2012. "Pricing and welfare in health plan choice." *American Economic Review*, 102(7): 3214–3248.

Bundorf, M. Kate, Maria Polyakova, Cheryl Stults, Amy Meehan, Roman Klimke, Ting Pun, Albert Solomon Chan, and Ming Tai-Seale. 2019. "Machine-Based Expert Recommendations And Insurance Choices Among Medicare Part D Enrollees." *Health Affairs*, 38(3): 482–490. PMID: 30830808.

Cardon, James H., and Igal Hendel. 2001. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *The RAND Journal of Economics*, 32(3): 408–427.

Carlin, Caroline, and Robert Town. 2008. "Sponsored Health Plans." *ReVision*, 0–67.

Cobb, Barry, Rafael Rumí, and Antonio Salmerón. 2012. "Approximating the Distribution of a Sum of Log-normal Random Variables." *Proceedings of the 6th European Workshop on Probabilistic Graphical Models, PGM 2012*.

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

Cutler, David M, and Sarah J Reber. 1998. "Paying For Health Insurance: The Trade-off Between Competition and Adverse Selection." *The Quarterly Journal of Economics*, 113(2): 433–466.

Dafny, Leemore, Kate Ho, and Mauricio Varela. 2013. "Let Them Have Choice: Gains from Shifting Away from Employer-Sponsored Health Insurance and toward an Individual Exchange." *American Economic Journal: Economic Policy*, 5(1): 32–58.

Dalton, Christina M, Gautam Gowrisankaran, and Robert Town. 2015. "Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D." National Bureau of Economic Research Working Paper 21104.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *The American Economic Review*, 67(3): 297–308.

Dubin, Jeffrey A., and Daniel L. McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica*, 52(2): 345–362.

Einav, Liran, Amy Finkelstein, and Jonathan Levin. 2010. "Beyond Testing: Empirical Models of Insurance Markets." *Annual Review of Economics*, 2(1): 311–336.

Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen. 2013. "Selection on moral hazard in health insurance." *American Economic Review*, 103(1): 178–219.

Einav, Liran, Amy Finkelstein, Yunan Ji, and Neale Mahoney. 2020. "Voluntary Regulation: Evidence from Medicare Payment Reform." National Bureau of Economic Research Working Paper 27223.

Einav, Liran, Amy N. Finkelstein, and Mark Cullen. 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *The Quarterly Journal of Economics*, CXV(3).

Ericson, Keith Marzilli, and Justin R Sydnor. 2018. "Liquidity Constraints and the Value of Insurance." National Bureau of Economic Research Working Paper 24993.

Ericson, Keith Marzilli, and Justin Sydnor. 2017. "The Questionable Value of Having a Choice of Levels of Health Insurance Coverage." *Journal of Economic Perspectives*, 31(4): 51–72.

Fenton, L. F. 1960. "The sum of log-normal probability distributions in scatter transmission systems." *IRE Transactions on Communication Systems*, 8: 57–67.

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

Geruso, Michael. 2017. "Demand heterogeneity in insurance markets: Implications for equity and efficiency." *Quantitative Economics*, 8(3): 929–975.

Ghili, Soheil, Benjamin R Handel, Igal Hendel, and Michael D Whinston. 2019. "The Welfare Effects of Long-Term Health Insurance Contracts." National Bureau of Economic Research Working Paper 23624.

Gross, Tal, and Matthew J. Notowidigdo. 2011. "Health insurance and the consumer bankruptcy decision: Evidence from expansions of Medicaid." *Journal of Public Economics*, 95(7): 767 – 778.

Gruber, Jonathan, Ben Handel Jonathan Kolstad Sam Kina. 2019. "Managing Intelligence: Skilled Experts and AI in Markets for Complex Products."

Handel, Benjamin, Igal Hendel, and Michael Whinston. 2015. "Equilibria in Health Exchanges : Adverse Selection vs . Reclassification Risk." *Econometrica*, 83(4): 1261–1313.

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

Handel, Benjamin R., and Jonathan T. Kolstad. 2015. “Health Insurance for ”Humans”: Information Frictions, Plan Choice, and Consumer Welfare.” *American Economic Review*, 105(8): 2449–2500.

Ho, Kate, and Robin Lee. 2019. “Health Insurance Menu Design: Managing the Spending Coverage Tradeoff.” *Presentation. Conference Celebrating the Scholarly Career of Mark Satterthwaite*, Kellogg School of Management.

Ketcham, Jonathan D., Claudio Lucarelli, Eugenio J. Miravete, and M. Christopher Roebuck. 2012. “Sinking, Swimming, or Learning to Swim in Medicare Part D.” *American Economic Review*, 102(6): 2639–73.

Kowalski, Amanda E. 2015. “Estimating the tradeoff between risk protection and moral hazard with a nonlinear budget set model of health insurance.” *International Journal of Industrial Organization*, 43: 122 – 135.

Lustig, Joshua. 2008. “The Welfare Effects of Adverse Selection in Privatized Medicare.” *Department of Economics, Institute for Business and Economic Research, UC Berkeley, Department of Economics, Working Paper Series.*

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz. 1987. “Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment.” *The American Economic Review*, 77(3): 251–277.

McManus, Margaret A., Stephen Berman, Thomas McInerny, and Suk-fong Tang. 2006. “Weighing the Risks of Consumer-Driven Health Plans for Families.” *Pediatrics*, 117(4): 1420–1424.

Miranda, Mario J., and Paul L. Fackler. 2002. *Applied Computational Economics and Finance.* MIT Press.

Newhouse, Joseph. 1993. “Free for All? Lessons from the RAND Health Insurance Experiment.” *Cambridge, MA. Harvard University Press.*

Pauly, Mark V. 1968. “The Economics of Moral Hazard: Comment.” *American Economic Review*, 58(3): 531–537.

Pauly, Mark V. 1974. “Overinsurance and Public Provision of Insurance : The Roles of Moral Hazard and Adverse Selection.” *The Quaterly Journal of Economics*, 88(1): 44–62.

Reed, Mary, Vicki Fung, Mary Price, Richard Brand, Nancy Benedetti, Stephen F. Derose, Joseph P. Newhouse, and John Hsu. 2009. “High-Deductible Health Insurance Plans: Efforts To Sharpen A Blunt Instrument.” *Health Affairs*, 28(4): 1145–1154. PMID: 19597214.

Revelt, David, and Kenneth Train. 1998. “Mixed Logit with Repeated Choices: Households’ Choices of Appliance Efficiency Level.” *The Review of Economics and Statistics*, 80(4): 647–657.

Revelt, David, and Kenneth Train. 2001. “Customer-Specific Taste Parameters and Mixed Logit: Households’ Choice of Electricity Supplier.” , (0012001).

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics*, 90(4): 629–649.

Tilipman, Nicholas. 2018. "Narrow Physician Networks, Switching Costs, and Product Variety in Employer Markets."

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation: Second Edition.* Cambridge University Press.

Zeckhauser, Richard. 1970. "Medical Insurance: A Case Study of the Tradeoff Between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory*, 2(1): 10–26.

For Online Publication

Appendix A

A.1 Calculation of Willingness to Pay

The expected utility of a type-θ consumer with initial income $\hat{y}$ for contract x at premium p is given by $U(x,p,\theta)$, as defined in equation (1) and repeated here:

$$U(x,p,\theta) = \mathbb{E}_l[\ u_\psi(\hat{y} - p - c^*(l,\omega,x) + b^*(l,\omega,x))].$$

The corresponding certainty equivalent $CE(x,p,\theta)$ solves $u(CE(x,p,\theta)) = U(x,p,\theta)$. It can be expressed as:

$$\begin{aligned}
CE(x,p,\theta) &= u_\psi^{-1}(U(x,p,\theta)) \\
&= EV(x,\theta) + \hat{y} - p + u_\psi^{-1}(U(x,p,\theta)) - EV(x,\theta) + p - \hat{y} \\
&= EV(x,\theta) + \hat{y} - p - RP(x,p,\theta),
\end{aligned}$$

where $EV(x,\theta) + \hat{y} - p$ is the expected payoff and $RP(x,p,\theta)$ is the risk premium associated with the lottery. In particular,

$$\begin{aligned}
EV(x,\theta) &= \mathbb{E}_l[\ b^*(l,\omega,x) - c^*(l,\omega,x)\ ] \\
&= \mathbb{E}_l[\ b_0(l,\omega) - c_0(l,\omega,x) + v(l,\omega,x)\ ],\ \text{and} \\
RP(x,p,\theta) &= EV(x,\theta) + \hat{y} - p - u_\psi^{-1}(U(x,p,\theta)).
\end{aligned} \tag{10}$$

A consumer's willingness to pay for contract x relative to the null contract x_0 is equal to $\tilde{p}$ that solves:

$$
\begin{aligned}
CE(x, \tilde{p}, \theta) &= CE(x_o, p_o, \theta) \\
EV(x, \theta) + \hat{y} - \tilde{p} - RP(x, \tilde{p}, \theta) &= EV(x_0, \theta) + \hat{y} - p_0 - RP(x_0, p_0, \theta) \\
\tilde{p} - p_0 &= EV(x, \theta) - EV(x_0, \theta) + RP(x_0, p_0, \theta) - RP(x, \tilde{p}, \theta),
\end{aligned}
$$

where p_0 is the price of the null contract. To obtain a closed-form expression for willingness to pay, we assume constant absolute risk aversion, and thus that the risk premium RP does not depend on residual income.[78] In this case, marginal willingness to pay for contract x relative to the null contract is given by:

$$
\begin{aligned}
WTP(x, \theta) &= EV(x, \theta) - EV(x_0, \theta) + RP(x_0, \theta) - RP(x, \theta) \\
&= \mathbb{E}_l[\ c_0(l, \omega, x_0) - c_0(l, \omega, x) + v(l, \omega, x)\ ] + \Psi(x, \theta),
\end{aligned}
$$

where $\Psi(x, \theta) = RP(x_0, \theta) - RP(x, \theta)$. If the null contract provides a riskier distribution of payoffs than contract x, $\Psi(x, \theta)$ will be positive. The last step uses the facts that (i) $c^*(l, \omega, x_0) = c_0(l, \omega, x_0)$, and (ii) $\mathbb{E}_l[v(l, \omega, x_0)] = 0$ because there is no moral hazard spending in the null contract.

A.2 Estimation of Plan Cost-sharing Features

A crucial input to our empirical model is the cost-sharing function of each plan. While Table 2 describes plans using the deductible and in-network out-of-pocket maximum, plans are in reality characterized by a much more complex set of payment rules, including copayments, specialist visit coinsurance, out-of-network fees, and fixed charges for emergency room visits. To structurally model moral hazard, we make the giant simplification that healthcare is a homogenous good over which the consumer chooses only the quantity to consume, and we model this decision as being based in part on out-of-pocket cost. To that end, our empirical model requires as an input a univariate function that maps total healthcare spending into out-of-pocket cost.

A natural choice might be to use the deductible, nonspecialist coinsurance rate, and in-network out-of-pocket maximum. However, in our setting, the out-of-pocket cost function

[78] In equation (10), $\hat{y} - p$ cancels out completely. This assumption is most reasonable when marginal premiums between relevant plans are small relative to income.

described by these features does not correspond well to what we observe in the claims data. In particular, we often observe out-of-pocket spending amounts that exceed plans' in-network out-of-pocket maximum. Because of this, we take a different approach.

We define plan cost-sharing functions by three parameters: a deductible, a coinsurance rate, and an out-of-pocket maximum. Taking the true deductibles as given (since these correspond well to the data), we estimate a coinsurance rate and an out-of-pocket maximum that minimizes the sum of squared residuals between predicted and observed out-of-pocket cost. We observe realized total healthcare spending for each household in the claims data. Predicted out-of-pocket cost is calculated by applying the deductible and supposed coinsurance rate and out-of-pocket maximum. "Observed" out-of-pocket cost is either observed directly in the claims data (if a household chose that plan) or else calculated counterfactually.[79] We carry out this procedure separately for each plan, year, and family status (individual or family).[80]

Figure A.1 shows the data used to estimate the cost-sharing features of a particular plan (Moda - 3 for individual households in 2012). Total healthcare spending is on the horizontal axis and out-of-pocket cost is on the vertical axis. Each gray open circle indicates a household.[81] The dark-colored dots are a binscatter plot of the gray open circles data, using 100 data points. The observed, basic cost-sharing features of the plan are a deductible of $300, nonspecialist coinsurance rate of 20 percent, and in-network out-of-pocket maximum of $2,000. It is clear that the data do not correspond well to a $2,000 out-of-pocket maximum. The red line shows the "estimated" cost-sharing function: The estimated coinsurance rate is 20.5 percent and the estimated out-of-pocket maximum is $3,218. Table A.3 presents the estimated cost-sharing features for all plans in all years. Figure A.2 shows graphically the estimated out-of-pocket cost functions for Moda plans in 2009.

A.3 Variation in Plan Menu Generosity

Measuring Plan Menu Generosity. With "plan menu generosity," we want to capture the likelihood that a household would choose generous health insurance coverage when presented with that menu. At a simple level, if plan menus consisted of only a single plan,

[79] We calculate counterfactual out-of-pocket spending using the "claims calculator" developed for this setting by Abaluck and Gruber (2016).

[80] So that the cost-sharing estimates are not affected by large outliers, we drop observations where out-of-pocket spending was above $20,000 or total healthcare spending was above $100,000.

[81] Because there are thousands of households, the plot only shows the dots for a 20 percent random sample.

the assignment to higher coverage would obviously constitute a "more generous menu" than the assignment to lower coverage. Similarly, if plan choice sets were all the same and only employee premiums varied, lower premiums would clearly correspond to a more generous menu. However, in our setting, plan menus are more complex. They contain multiple plans and many possible permutations of plan choice sets, and plans vary by their actuarial value, the identity of their insurer, their associated employee premium, and their potential HSA/HRA and vision/dental contribution. All of these factors likely influence households' plan choices.

In order to construct usable measures of plan menu generosity, we transform these multi-dimensional objects using a conditional logit model that excludes all household observables. This specification allows us to predict the probability that a given household would choose a given plan when presented with a given plan menu *as if* the household had been acting like the average household in the data. Variation in the resulting predicted choice probabilities is driven only by variation in plan menus, and not by variation in (observed or unobserved) household characteristics.

Abstracting from the dimension of time for now, we define $plan_{jk}$ as an indicator for the plan j chosen by household k. We estimate the following conditional logit model:

$$plan_{jk} = \underset{j \in \mathcal{J}_d}{\operatorname{argmax}}(\alpha p_{jd} + \alpha^{VD} p_{jd}^{VD} + \alpha^{HA} p_{jd}^{HA} + \nu_j + \epsilon_{jk}), \tag{11}$$

where $\mathcal{J}_d$ is the set of plans available in the school district-family type-occupation type combination d (to which household k belongs), p_{jd} is the employee premium, p_{jd}^{VD} is the vision/dental subsidy, and p_{jd}^{HA} is the HSA/HRA contribution. Plan characteristics are captured nonparametrically by plan fixed effects ν_j. All household-specific determinants of plan choice are contained in the error term ϵ_{jk}. Estimated parameters are presented in Table A.7, separately for each year of our data. As expected, households dislike premiums, prefer higher HSA/HRA and vision/dental subsidies, and prefer higher-coverage plans to lower-coverage plans.

We use the choice probabilities implied by equation (11) to construct our measures of plan menu generosity. Given plan menu $\mathbf{menu}_d \equiv \{p_{jd}, p_{jd}^{VD}, p_{jd}^{HA}, \nu_j\}_{j \in \mathcal{J}_d}$, we denote the predicted probability that a household would choose plan j as ρ_{jd}.[82] Our measures of plan menu generosity are the probability a household would choose a given insurer and the expected actuarial value of a household's plan choice conditional on insurer, respectively

[82]Formally: $\rho_{jd} = \frac{\exp(U_{jd})}{\sum_{g \in \mathcal{J}_d} \exp(U_{gd})}$, where $U_{jd} = \alpha p_{jd} + \alpha^{VD} p_{jd}^{VD} + \alpha^{HA} p_{jd}^{HA} + \nu^j$.

given by:

$$\rho_{fd} = \sum_{j \in \mathcal{J}_d^f} \rho_{jd},$$
$$\widehat{AV}_{fd} = \sum_{j \in \mathcal{J}_d^f} (\frac{\rho_{jd}}{\rho_{fd}}) AV_j, \tag{12}$$

where $\mathcal{J}_d^f$ is the set of plans in $\mathbf{menu}_d$ offered by insurer f. Since this is where the majority of the variation in coverage level lies, we focus on explaining plan menu generosity using the predicted actuarial value among Moda plans. In our reduced-form analysis of moral hazard in Appendix A.4, we use the measure for all insurers.

Explaining Plan Menu Generosity. We first compare plan menu generosity to observed household health (Table A.4). We can in all years reject the hypothesis that household risk scores are correlated with plan menu generosity, conditional on family structure. We also consistently find that plan menus are most generous for single employee coverage and least generous for employee plus family coverage. This is consistent with our understanding of OEBB's benefit structure and is common in employer-sponsored health insurance.

We further explore what covariates, in addition to family structure, *do* explain variation in plan menu generosity. Table A.5 presents three additional regressions of predicted actuarial value on employee-level covariates (part-time versus full-time status, occupation type, and union affiliation), as well as on school district-level covariates (home price index and percent of Republicans). Employees are either part-time or full-time. There are eight mutually exclusive employee occupation types; the regressions omit the type "Licensed Administrator."[83] There are five mutually exclusive union affiliations, and employees may not be affiliated with a union; the regressions omit the non-union category. We calculate the average home price index (HPI) in a school district by taking the average zip-code level home price index across employees' zip-code of residence.[84] *Pct. Republican* measures the percent of households in a school district that are registered as Republicans as of 2016.[85]

We find that plan menus are less generous for part-time employees, are substantially less generous for substitute teachers, and are more generous for employees at community

[83] "Licensed" refers to the possession of a teaching license.

[84] We use 5-digit zip-code-level home price indices from Bogin, Doerner and Larson (2019). The data and paper are accessible at http://www.fhfa.gov/papers/wp1601.aspx.

[85] Data on percent of registered voters by party is available at the county level; we construct school-district-level measures by taking the average over employees' county of residence. Voter registration data in Oregon can be downloaded at https://data.oregon.gov/api/views/6a4f-ecbi.

colleges. Certain union affiliations are also predictive of more or less generous plan menus. Across school districts, plan menu generosity is decreasing in both the logged home price index and the percent of registered Republicans.

A.4 Reduced-form Estimates of Moral Hazard

While our primary sample consists of data from 2009–2013, we conduct our reduced-form analysis of moral hazard using only data from 2008.[86] The OEBB marketplace began operating in 2008, so that year all employees chose from this set of plans for the first time. This "active choice" year permits us to look cleanly at how plan choices and healthcare spending depended on plan menus without also having to account for how prior-year plan menus affected current-year plan choices. While our structural model will capture these dynamics, we feel they are better avoided at this stage.

We estimate how plan menus—choice sets and prices—affect plan choices, and in turn how plan choices affect total healthcare spending, as described by equations (13) and (14):

$$plan_k = f(\mathbf{menu}_d, \mathbf{X}_k, \xi_k), \tag{13}$$

$$y_k = g(plan_k, \mathbf{X}_k, \xi_k). \tag{14}$$

Here, $plan_k$ represents the plan chosen by household k, $\mathbf{menu}_d$ represents the plan menu available to the school district-family type-occupation type combination d (to which household k belongs), $\mathbf{X}_k$ are observable household characteristics, ξ_k are unobservable household characteristics, and y_k is total healthcare spending. Because household characteristics appear in both equations, the standard challenge in estimating the effect of $plan_k$ on y_k is that a household's chosen plan is correlated with its unobservable characteristics ξ_k. Our identifying assumption is that plan menus are independent of household unobservables ξ_k conditional on household observables $\mathbf{X}_k$.

We parameterize $plan_k$ to be an indicator variable for the identity of the insurer and a continuous variable for the plan actuarial value. We then parameterize equation (14) according to

$$\log(y_k) = \delta_f \mathbf{1}_{f(k)=f} + \gamma \log(1 - AV_{j(k)})\mathbf{1}_{f(k)=Moda} + \beta \mathbf{X}_k + \xi_k, \tag{15}$$

[86]The cost-sharing features of 2008 plans are presented in Table A.2; they are very similar to the plans offered in 2009. We apply the same sample construction criteria to our 2008 sample, except that households must be present for one prior year. Summary statistics on the 2008 sample can be found in Table A.6.

where $\mathbf{1}_{f(k)=f}$ is an indicator for the insurer chosen by household k and $AV_{j(k)}$ is the actuarial value of the plan chosen by household k. The parameter δ_f represents insurer-specific treatment effects on total spending.[87] Our parameter of interest is γ, which represents the responsiveness of total spending to plan generosity, holding the insurer fixed (at Moda).[88] We follow the literature in formulating the model so that γ represents the elasticity of total healthcare spending with respect to the average out-of-pocket price per dollar of total spending.[89]

We estimate equation (15) using two-stage least squares, instrumenting for the chosen insurer ($\mathbf{1}_{f(k)=f}$) and actuarial value ($AV_{j(k)}$) using $\mathbf{menu}_d$. As instruments, we use the measures of plan menu generosity constructed in Appendix A.3. Namely, we instrument for $\mathbf{1}_{f(k)=f}$ using using ρ_{fd} and for $\log(1-AV_{j(k)})\mathbf{1}_{f(k)=Moda}$ using $\log(1-\widehat{AV}_{d,Moda})\rho_{d,Moda}$. Table A.8 reports the estimates. We report only the coefficient of interest (γ), but all specifications also contain insurer fixed effects, as well as controls for household risk score and family structure. The first column presents the parameters estimated without instruments, and the second column presents the instrumental variables estimates. Comparing the coefficients in columns 1 and 2, we find that moral hazard explains 46 percent of the observed relationship between plan generosity and total healthcare spending. Our overall estimate of the elasticity of demand for healthcare spending in the population is -0.27. The standard benchmark estimate from the RAND health insurance experiment is -0.2 (Manning et al., 1987; Newhouse, 1993).

Heterogeneity. Columns 3 and 4 of Table A.8 introduce heterogeneity in γ by household health. For each household type (individual or family), we classify households into quartiles based on household risk score, where Q_n denotes the quartile of risk (Q_4 is highest risk). We construct separate instruments for each of the eight household types by estimating the logit model in equation (11) for only that subsample of households.[90] We find noisy but large differences in γ across household risk quartiles and between individual and family households.

Variation in γ could reflect either heterogeneity in the intensity of treatment (extent of exposure to varying marginal prices of healthcare across plans), or heterogeneity in

[87] These may arise due to "supply side" effects arising from differences in provider prices, provider networks, or care management practices, or due to "demand side" effects from differences in average plan generosity.

[88] We do not try to estimate a moral hazard elasticity among the plans offered by Kaiser and Providence because there is so little variation in coverage level.

[89] To accommodate the fact that 2 percent of households have zero spending, we add 1 to total spending.

[90] Estimates for each subsample are presented in Table A.9.

treatment effect (different responsiveness to varying marginal prices of healthcare across plans), or both. While this analysis cannot distinguish between these two effects, we find suggestive evidence that the heterogeneity at least in part reflects differential treatment intensity. The remainder of this section presents an analysis that compares the realized spending outcomes of households in different risk quartiles with the variation in plan cost-sharing features that gives rise to different end-of-year marginal out-of-pocket prices. We find that the household types for which we estimate higher γ are also more likely to be exposed to varying marginal out-of-pocket costs. Distinguishing variation in treatment intensity from variation in treatment effect is an important advantage of our structural model.

Variation in Treatment Intensity. We explore the extent to which heterogeneity in moral hazard can be explained by variation in the intensity of treatment. Assignment to a lower or higher coverage plan could affect total spending by exposing consumers to lower or higher out-of-pocket costs. However, if a consumer is so healthy that they would almost always be consuming healthcare at levels below the deductible of both plans, there is in fact no variation in coverage level for that consumer. The same could be true of very sick households that, knowing they will always spend the out-of-pocket maximum, face the same marginal out-of-pocket cost in both plans.

Table A.10 compares the realized spending outcomes of households in different risk quartiles with the variation in plan cost-sharing features that gives rise to different marginal out-of-pocket prices. The top panel of Table A.10 shows the observed distributions of total spending for the four quartiles of risk for individual and family households. The bottom panel shows the (in-network) deductible and out-of-pocket maximum for each of the Moda plans in 2008. It shows, for example, that individual households in the first health quartile have the majority of the density of their spending distribution around or below the deductibles, while individual households in the third and fourth quartiles have the majority of their spending around or above the out-of-pocket maximums.

The patterns of heterogeneity in our estimates of moral hazard in Table A.8 correspond well to the likely variation in marginal out-of-pocket prices facing each type of household. For example, we estimate the largest amount of moral hazard for the second quartile of individual households, whose spending distribution more closely spans the range over which there would in fact be marginal out-of-pocket price variation across plans. Likewise for family households, those in the fourth quartile nearly all have spending above the highest out-of-pocket maximum, and we do not estimate any moral hazard within this group. While this exercise is merely suggestive, it points to the fact that an important dimension of heterogeneity is the extent to which households are exposed to differential out-of-pocket spending across nonlinear insurance contracts.

Appendix B Estimation Details

B.1 Fenton-Wilkinson Approximation

Because there is no closed-form solution for the distribution of the sum of lognormal random variables, the Fenton-Wilkinson approximation is widely used in practice.[91] Under this approximation, the distribution of the sum of draws from independent lognormal distributions can be represented by a lognormal distribution. The parameters of the approximating distribution are chosen such that its first and second moments match the corresponding moments of the true distribution of the sum of lognormals. In our application, the sum of lognormals is the household's health state distribution, and the lognormals being summed are the individuals' health state distributions. An individual's health state $\tilde{l}^i$ is assumed have a shifted lognormal distribution:

$$\log(\tilde{l}^i + \kappa_i) \sim N(\mu_i, \sigma_i^2).$$

All parameters may vary over time (since individual demographics vary over time), but t subscripts are omitted here for simplicity. The moment-matching conditions for the distribution of the household-level health state $\tilde{l}$ are:

$$E(\tilde{l} + \kappa_k) = \sum_{i \in \mathcal{I}_k} E(\tilde{l}^i + \kappa_i), \tag{16}$$

$$Var(\tilde{l} + \kappa_k) = \sum_{i \in \mathcal{I}_k} Var(\tilde{l}^i + \kappa_i), \tag{17}$$

$$E(\tilde{l}) = \sum_{i \in \mathcal{I}_k} E(\tilde{l}^i), \tag{18}$$

where $\mathcal{I}_k$ is the set of individuals in household k. Equation (16) sets the mean of the household's distribution equal to the sum of the means of each individual's distribution. Equation (17) matches the variance. Because we have a third parameter to estimate (the

[91]See Fenton (1960), and for a summary, Cobb, Rumí and Salmerón (2012).

shift, κ_k), we use a third moment-matching condition to match the first moment of the unshifted distribution, shown in equation (18).

Under the approximating assumption that $\tilde{l} + \kappa_k$ is distributed lognormally, and substituting the analytical expressions for the mean and variable of a lognormal distribution, these equations become:

$$\begin{aligned}
\exp(\mu_k + \frac{\sigma_k^2}{2}) &= \sum_{i \in \mathcal{I}_k} \exp(\mu_i + \frac{\sigma_i^2}{2}) \\
\big(\exp(\sigma_k^2) - 1\big) \exp(2\mu_k + \sigma_k^2) &= \sum_{i \in \mathcal{I}_k} \big(\exp(\sigma_i^2) - 1\big) \exp(2\mu_i + \sigma^2) \\
\exp(\mu_k + \frac{\sigma_k^2}{2}) - \kappa_k &= \sum_{i \in \mathcal{I}_k} \exp(\mu_i + \frac{\sigma_i^2}{2}) - \kappa_i
\end{aligned}$$

Given a guess of the parameters to be estimated (the individual-level parameters), this leaves three equations in three unknowns, and we can solve for the household-level parameters. The solutions for μ_k, σ_k^2, and κ_k are:

$$\begin{aligned}
\sigma_k^2 &= \log[1 + \left[\sum_{i \in \mathcal{I}_k} \exp(\mu_i + \frac{\sigma_i^2}{2})\right]^{-2} \sum_{i \in \mathcal{I}_k} (\exp(\sigma_i^2) - 1) \exp(2\mu_i + \sigma_i^2)] \\
\mu_k &= -\frac{\sigma_k^2}{2} + \log[\sum_{i \in \mathcal{I}_k} \exp(\mu_i + \frac{\sigma_i^2}{2})] \\
\kappa_k &= \sum_{i \in \mathcal{I}_k} \kappa_i
\end{aligned}$$

Given these algebraic solutions for the parameters of a household's health state distribution, we can work backward to estimate which individual-level parameters best explain the observed data on individual-level demographics and household-level healthcare spending. A key advantage of using this approximation instead of simply simulating the true distribution of the sum of lognormals is that we can use quadrature to integrate the distributions of health states, thereby limiting the number of support points needed for numerical integration.

B.2 Estimation Algorithm

We estimate the model using a maximum likelihood approach similar to that described by Revelt and Train (1998) and Train (2009), with the appropriate extension to a discrete/continuous choice model in the style of Dubin and McFadden (1984). The maximum likelihood estimator selects the parameter values that maximize the conditional probability density of households' observed total healthcare spending, given their plan choices.

The model contains four dimensions of unobservable heterogeneity: risk aversion, household health, the moral hazard parameter, and the T1-EV idiosyncratic shock. The last we can integrate analytically, but the first three we must integrate numerically; we denote these as $\beta_{kt} = \{\psi_k, \mu_{kt}, \omega_k\}$. We denote the full set of parameters to be estimated as θ, which, among other things, contains the parameters of the distribution of β_{kt}. Given a guess of θ, we simulate the distribution of β_{kt} using Gaussian quadrature with 27 support points, yielding simulated points $\beta_{kts}(\theta) = \{\psi_{ks}, \mu_{kts}, \omega_{ks}\}$, as well as weights W_s.[92,93] For each simulation draw s, we then calculate the conditional density at households' observed total healthcare spending and the probability of households' observed plan choices.

Conditional Probability Density of Healthcare Spending. We have data on realized healthcare spending m_{kt} for each household and year. We aim to construct the distribution of healthcare spending for each household-year implied by the model and guess of parameters. We start by constructing individual-level health state distribution parameters μ_{it}, σ_{it}, and κ_{it} from θ and individual demographics, as described in equation (7). We then construct household-level health state distribution parameters μ_{kts}, σ_{kt}, and κ_{kt} using the formulas in equation (8) and the draws of $\beta_{kts}(\theta)$. The model predicts that upon realizing their health state l, households choose total healthcare spending m by trading off the benefit of healthcare utilization with its out-of-pocket cost. Specifically, accounting for the fact that negative health states may imply zero spending, the model predicts optimal healthcare spending $m^*_{jt}(l, \omega_{ks}) = \max(0\ ,\ \omega_{ks}(1 - c'_{jt}(m^*)) + l)$ if household k were enrolled in plan j in year t. Inverting the expression, the health state realization l_{kjts} that would have given rise to observed spending m_{kt} under moral hazard parameter ω_{ks} is given by

$$l_{kjts} : \begin{cases} l_{kjts} < 0 & m_{kt} = 0 \\ l_{kjts} = m_{kt} - \omega_{ks}(1 - c'_{jt}(m_{kt})) & m_{kt} > 0. \end{cases}$$

[92] Note that the mean vector of β_{kts} is a fixed function of θ and household demographics.

[93] We use the Matlab program *qnwnorm* to implement this method, with three points in each dimension of unobserved heterogeneity. The program can be obtained as part of Mario Miranda and Paul Fackler's CompEcon Toolbox; for more information, see Miranda and Fackler (2002).

Note that $c'_{jt}(m^*) = 1$ when $m_{kt} = 0$. Household health state is distributed according to

$$l = \phi_f \tilde{l}$$
$$\log(\tilde{l} + \kappa_{kt}) \sim N(\mu_{kts}, \sigma^2_{kt}).$$

There are two possibilities to consider. First, if m_{kt} is equal to zero, the implied health state realization l_{kjts} is negative. Given monetary health state realization l_{kjts}, the implied "quantity" health state realization is equal to $\tilde{l}_{kjts} = \phi_f^{-1} l_{kjts}$, where f is the insurer offering plan j. Since $\phi_f > 0$, the probability of observing $l_{kjts} < 0$ is the probability of observing $\tilde{l}_{kjts} \leq \kappa_{kt}$. Second, if m_{kt} is greater than zero, it is useful to define $\lambda_{kjts} = \phi_f^{-1} l_{kjts} + \kappa_{kt}$, which itself is distributed lognormally (no shift). The density of m_{kt} in this case is given by the density of λ_{kjts}. Taken together, the probability density of total healthcare spending m conditional on plan, parameters, and household observables $\mathbf{X}_{kt}$ is given by $f_m(m_{kt}|c_{jt}, \beta_{kts}, \theta, \mathbf{X}_{kt}) = P(m = m_{kt}|c_{jt}, \beta_{kts}, \theta, \mathbf{X}_{kt})$, where

$$f_m(m_{kt}|c_{jt}, \beta_{ks}, \theta, \mathbf{X}_{kt}) = \begin{cases} \Phi\left(\frac{\log(\kappa_{kt}) - \mu_{kt}}{\sigma_{kt}}\right) & m_{kt} = 0, \\ \phi_f^{-1} \Phi'\left(\frac{\log(\lambda_{kjts}) - \mu_{kt}}{\sigma_{kt}}\right) & m_{kt} > 0, \end{cases}$$

and $\Phi(\cdot)$ is the standard normal cumulative distribution function. For a given guess of parameters, there are certain values of m_{kt} for which the probability density is zero. In order to rationalize the data at all possible parameter guesses, in practice we use a convolution of $f_m(m_{kt}|c_{jt}, \beta_{ks}, \theta, \mathbf{X}_{kt})$ and a uniform distribution over the range [-1e-75, 1e75].[94]

Probability of Plan Choices. We next calculate the probability of a household's observed plan choice. Given θ and β_{kts}, we simulate the distribution of health states l_{kjtsd} using $D = 30$ support points:

$$l_{kjtsd} = \phi_f \big(\exp(\mu_{kts} + \sigma_{kt} Z_d) - \kappa_{kt} \big),$$

where Z_d is a vector of points that approximates a standard normal distribution using Gaussian quadrature, and W_d (to be used soon) are the associated weights. We then calculate the optimal healthcare spending choice m_{kjtsd} associated with each potential health state realization, according to $m^*_{kjtsd} = \max(0\ ,\ \omega_{ks}(1 - c'_{jt}(m^*_{kjtsd})) + l_{kjtsd})$. Because marginal out-of-pocket cost depends on where the out-of-pocket cost function is evaluated,

[94] We have experimented with varying these bounds and found that this does not affect parameter estimates as long as the uniform density is sufficiently small.

there is not a closed-form solution for m^*_{kjtsd}. Instead, we derive cutoff values on the health state that determine which out-of-pocket cost "region" a household will find optimal.

Plans in our empirical setting are characterized by a deductible, a coinsurance rate, and an out-of-pocket maximum. Because the plans are piecewise linear (in three pieces), one must only test three candidate values of $c'(m)$, and then compare optimized utility in each case. Specifically, $c'(m) = 1$ if spending m is in the deductible region, $c'(m) = c$ in the coinsurance region, and $c'(m) = 0$ in the out-of-pocket maximum region. By performing a generic version of this calculation, we can construct the relevant cutoff values for the health state. Define a plan to consist of a deductible D, a coinsurance rate C, and an out-of-pocket maximum O. Define $A = C^{-1}(O - D(1 - C))$ to be the level of total spending above which the consumer would reach their out-of-pocket maximum. Under moral hazard parameter ω, the relevant cutoff values are

$$
\begin{aligned}
Z_1 &= D - \omega(1 - C)/2, \\
Z_2 &= O - \omega/2, \\
Z_3 &= A - \omega(1 - C/2),
\end{aligned}
$$

where $Z_1 \leq Z_2 \leq Z_3$ so long as $O \geq D$ and $C \in [0, 1]$. There are two types of plans to consider. If D and A are sufficiently far apart (there is a sufficiently large coinsurance region), then only the cutoffs Z_1 and Z_3 matter, and it may be optimal to be in any of the three regions, depending on where the health state is relative to those two cutoff values. If D and A are close together, it will never be optimal to be in the coinsurance region (better to burn right though it and into the free healthcare of the out-of-pocket maximum region), and the cutoff Z_2 will determine whether the deductible or out-of-pocket maximum region is optimal. If the realized health state is negative, optimal spending will equal zero. In sum, optimal spending m^* conditional on health state realization l, moral hazard parameter ω, and plan characteristics $\{D, C, O\}$ is given by

$$
\text{If} \quad A - D > \omega/2: \qquad m^* = \begin{cases} \max(0, l) & l \leq Z_1, \\ l + \omega(1 - C) & Z_1 < l \leq Z_3, \\ l + \omega & Z_3 < l; \end{cases}
$$

$$
\text{If} \quad A - D \leq \omega/2: \qquad m^* = \begin{cases} \max(0, l) & l \leq Z_2, \\ l + \omega & Z_2 < l. \end{cases}
$$

Derivations are available upon request. A graphical example (of the case in which the coinsurance region is sufficiently large) is shown in Figure A.3b. All plans in our empirical setting have $A - D > \omega/2$ at reasonable values of ω.

With distributions of privately optimal total healthcare spending m^*_{kjtsd} in hand for each household, plan, year, and draw of β_{ks}, we can calculate households' expected utility from enrolling in each potential plan. We construct the numerical approximation to equation (5) using the quadrature weights W_d:

$$U_{kjts} = -\sum_{d=1}^{D} W_d \cdot \exp\left(-\psi_k x_{kjts}(l_{kjtsd})\right),$$

where the monetary payoff x is calculated as in equation (6). To avoid numerical issues arising from double-exponentiation, we estimate the model in certainty-equivalent units of U_{kjts}:

$$U^{CE}_{kjts} = \bar{x}_{kjts} - \frac{1}{\psi_k} \log\left(\sum_{d=1}^{D} W_d \cdot \exp\Big(-\psi_k\big(x_{kjts}(l_{kjtsd}) - \bar{x}_{kjts}\big)\Big)\right),$$

where $\bar{x}_{kjts} = \mathbb{E}_d[x_{kjts}(l_{kjtsd})]$. Another reason for estimating the model in certainty equivalents is that it becomes simple to denominate the logit error term in dollars rather than in utils. This ensures that our choice model is "monotone," meaning that the probability of preferring a less-risky plan is everywhere increasing in risk aversion; see Apesteguia and Ballester (2018) for a full treatment of this issue.

Choice probabilities, conditional on β_{kts}, are given by the standard logit formula:

$$L_{kjts} = \frac{\exp(U^{CE}_{kjts}/\sigma_\epsilon)}{\sum_{i\in\mathcal{J}_{kt}} \exp(U^{CE}_{kits}/\sigma_\epsilon)}.$$

Likelihood Function. The numerical approximation to the likelihood of the sequence of choices and healthcare spending amounts for a given household is given by

$$LL_k = \sum_{j=1}^{J} d_{kjt} \sum_{s=1}^{S} W_s \prod_{t=1}^{T} f_m(m_{kt}|\theta, \beta_{kts}, c_{jt}, \mathbf{X}_{kt}) L_{kjts},$$

where $d_{kjt} = 1$ if household k chose plan j in year t and zero otherwise. The log-likelihood function for parameters θ is

$$LL(\theta) = \sum_{k=1}^{K} \log\left(LL_k\right).$$

B.3 Recovering Household-specific Types

We assume that household types $\beta_{kt}(\theta) = \{\psi_k, \mu_{kt}, \omega_k\}$ are distributed multivariate normal with observable heterogeneity in the mean vector, according to equation (9). After estimating the model and obtaining $\hat{\theta}$, we want to use each household's observed outcomes (plan choices and healthcare spending amounts) to back out which type they are likely to be. Let $g(\beta|\hat{\theta})$ denote the population distribution of types. Let $h(\beta|\hat{\theta}, y)$ denote the density of β conditional on parameters $\hat{\theta}$ and a sequence of observed plan choices and healthcare spending amounts y. Using what Revelt and Train (2001) term the "conditioning of individual tastes" method, we recover households' posterior distribution of β using Bayes' rule:

$$h(\beta|\hat{\theta}, y) = \frac{p(y|\beta)g(\beta|\hat{\theta})}{p(y|\hat{\theta})}.$$

Taking the numerical approximations, $p(y|\hat{\theta})$ is simply the household-specific likelihood function LL_k for an observed sequence of plan choices and spending amounts; $g(\beta|\hat{\theta})$ is the quadrature weights W_s on each simulated point; and $p(y|\beta)$ is the *conditional* household likelihood function LL_{ks}:

$$LL_{ks} = \sum_{j=1}^{J} d_{kjt} \prod_{t=1}^{T} f_m(m_{kt}|\theta, \beta_{ks}, c_{jt}, \mathbf{X}_{kt}) L_{kjts}.$$

Taken together, the numerical approximation to each household's posterior distribution of unobserved heterogeneity is given by

$$h_{ks}(\beta|\hat{\theta}, y_k) = \frac{LL_{ks} \cdot W_s}{LL_k},$$

where $\sum_s h_{ks}(\beta|\hat{\theta}, y_k) = 1$.

For the purposes of examining total variation in types across households (accounting for both observed and unobserved heterogeneity), we assign each household the expectation of their type with respect to their posterior distribution. We also use the household-specific distributions over types to calculated expected quantities of interest for each household. In particular, we calculate WTP_{kjt} and SS_{kjt} as

$$WTP_{kjt} = \sum_s h_{ks}(\beta|\hat{\theta}, y_k) WTP_{kjts},$$
$$SS_{kjt} = \sum_s h_{ks}(\beta|\hat{\theta}, y_k) SS_{kjts}.$$

B.4 Joint Distribution of Household Types

The joint distribution of household types is of central importance to this paper. Here, we investigate the distribution implied by our primary estimates in column 3 of Table 3. For each household, we first calculate the expectation of their type with respect to their posterior distribution of unobservable heterogeneity:

$$\psi_k = \sum_s h_{ks}(\beta|\hat{\theta}, y_k)\psi_{ks},$$
$$\omega_k = \sum_s h_{ks}(\beta|\hat{\theta}, y_k)\omega_{ks}.$$

In place of μ_{kt}, a more relevant measure of household health is the expected health state, i.e., expected total spending absent moral hazard. Using the expectation of a shifted lognormal variable and price parameter $\phi = 1$, the expected health state $\bar{l}_{kt}$ is given by

$$\bar{l}_{kt} = \sum_s h_{ks}(\beta|\hat{\theta}, y_k)(\exp(\mu_{kts} + \frac{\sigma_{kt}^2}{2}) - \kappa_{kt}).$$

To limit our focus to one type for each household, we look at $\bar{l}_{kt}$ for the first year each household appears in the data. Figure A.4 presents the joint distribution of household types along the dimensions of risk aversion (ψ), moral hazard (ω), and expected health state (log($\mathbb{E}$[Health state])). We measure the health state on a log scale for readability.

Appendix C Additional Material

C.1 Alternative Plan Designs

In the main text, we consider a potential set of five contracts that vary in deductible, coinsurance rate, and out-of-pocket maximum. Here, we consider alternative sets of potential contracts. We first expand the set of five contracts to a denser set of 40 contracts, still ranging from full insurance to Catastrophic. We then consider "linear contracts" that have only coinsurance, and "stop-loss contracts" that have only out-of-pocket maximums.

Adding additional contracts. Though choice is not efficient among our primary set of five contracts, it could be efficient if more contracts were available. We consider a denser set of 40 contracts that span the range between Catastrophic coverage and full insurance. The out-of-pocket cost functions for this full set of contracts is shown in Figure A.7b. The expanded set includes our five original contracts.

From Figure 7, we know that the Gold contract's social surplus curve lies everywhere above those of Silver and full insurance. This means that even in our expanded set of contracts, the regulator will never want to offer coverage as low as Silver or as high as full insurance. It says nothing, however, about the efficiency of choice over contracts in a closer neighborhood to Gold.

We solve numerically for optimal premiums in the set of 40 contracts. We find that in a small neighborhood of the Gold contract, it is efficient to offer choice. Figure A.12 shows willingness to pay and social surplus curves for the four contracts the regulator would offer. The four contracts are the Gold contract, the next-most-generous contract ("Gold+"), and the two next-less-generous contracts ("Gold−" and "Gold−−").[95] The social surplus curves in panel A.12b reveal the reason: The higher willingness-to-pay households should have higher coverage. Optimal premiums result in 22 percent of households choosing Gold+, 11 percent choosing Gold, 43 percent choosing Gold−, and 24 percent choosing Gold−−. This allocation achieves social surplus equal to $1,829 per household, a $27 (1.5 percent) increase over what is achieved by the Gold contract alone. Given the particularly small amounts of social surplus at stake across these contracts, it likely becomes first-order to consider any fixed costs associated with offering each contract or with offering choice at all.

Linear contracts. A fundamental driver of our findings is that, due to out-of-pocket maximums, consumers with high expected healthcare spending face little uncertainty. Absent these caps on consumer liability, we indeed find that expectation and variance are positively correlated (this can be seen indirectly in Figure A.10). As a result, with the disclaimer that our model is likely not well suited to such comparisons (c.f. footnote 53), we find that vertical choice is efficient among linear contracts. However, we also find that a

[95]The deductibles, coinsurance rates, and out-of-pocket maximums for these contracts are $998, 20%, $2,308 for Gold+; $1,169, 21%, $2,564 for Gold; $1,348, 23%, $2,821 for Gold−; and $1,563, 24%, $3,077 for Gold−−.

linear contracts with an out-of-pocket maximum always delivers higher social surplus than its uncapped counterpart, across the distribution of willingness to pay. A regulator would therefore never want to bind itself to offering only uncapped contracts.

We therefore consider a set of capped linear contracts. Consumers pay a fixed coinsurance rate while their total healthcare spending is lower than $20,000; after that point, they are fully insured. We consider coinsurance rates of 10%, 20%, 30%, and 50%, implying out-of-pocket maximums of $2,000, $4,000, $6,000, and $10,000. These contracts are depicted in Figure A.13b. Whereas only health state draws under about $10,000 would result in out-of-pocket uncertainty in our five original contracts, now the range of draws up to about $20,000 yield uncertainty.

Figure A.14 shows the distribution across household of willingness to pay and social surplus for this set of contracts. The larger coinsurance range results in a larger value of risk protection among high willingness-to-pay households, driving them to have full insurance as their efficient contract. Solving numerically for optimal premiums, the optimal allocation has 66 percent of households in full insurance, 32 percent of households in the 10% contract, and 2 percent of households in the 20% contract.[96] This allocation yields social surplus per household of $2,154, a $352 (19.5 percent) increase over what is achieved by our headline menu of contracts. While optimal contract design is only adjacent to the focus of this paper, it is interesting that removing the deductible has such a large effect on welfare.

Stop-loss contracts. For comparison, we also consider pure stop-loss contracts, with no insurance up to a point, and full insurance thereafter. We consider a $2,000, $4,500, $7,500, and $10,000 stop-loss amount (out-of-pocket maximum). The $10,000 contract is equivalent to the Catastrophic contract in the main text. These contracts are depicted in Figure A.13b.

The associated willingness to pay and social surplus curves are depicted in Figure A.15. Their shapes are familiar. As indicated by the social surplus curves in panel A.15b, optimal premiums lead all households to choose the $2,000 out-of-pocket maximum contract. This allocation delivers social surplus equal to $1,783 per household.

[96] These shares do not correspond perfectly to the visual implications of Figure A.14b because there is more disorder in willingness to pay across contracts, as indicated by the jaggedness in Figure A.14a.

C.2 Alternative Distribution of Consumer Types

We investigate the importance of preference heterogeneity to our results. Figure A.4 shows the estimated distribution of consumers types $\theta = \{\psi, \omega, F\}$ on which our main-text analysis is based. Here, we fix risk aversion ψ and the moral hazard parameter ω, leaving only variation across households in F. We fix ψ at 0.599 and ω at 1.751, their median values among family households.

Figure A.16 shows the distribution across households of willingness to pay and social surplus that arises from this new distribution of households types, using our original menu of contracts. The new distribution of types changes households' willingness to pay and social surplus, as well as the order of households on the horizontal axis. Variation in willingness to pay is now even more heavily determined by variation in mean reduced out-of-pocket costs. The pattern of low willingness to pay and high value of risk protection becomes even more pronounced. There is still little difference in social surplus across contracts for households with high willingness to pay. Overall, we find that it remains optimal to offer only the Gold contract. Since the original argument for offering choice centers on the ability to cater to heterogeneous preferences, this is not particularly surprising.

Table A.1. Sample Construction

Criteria	2009	2010	2011	2012	2013
Individuals in membership file	161,502	162,363	156,113	156,042	157,799
Not eligible for coverage	7,370	8,265	8,422	8,719	8,388
Retiree, COBRA, or oldest member over 65	13,180	12,567	12,057	11,603	11,840
Partial year coverage	17,115	18,649	19,283	21,281	23,074
Covered by multiple plans	1,447	1,947	2,038	2,239	2,336
Opted out	3,241	4,205	4,321	4,576	4,529
Not in intact family	8,389	9,188	9,181	8,925	10,265
No prior year of data	6,175	3,947	2,455	3,104	3,702
Missing premium or contribution data	25,653	28,466	22,755	23,284	30,401
Final total	78,932	75,129	75,601	72,311	63,264

Notes: This table shows the counts of individuals dropped due to each sample selection criterion. Drops are made in the order in which criteria appear. All observations in 2008 are dropped because there is no year of prior data.

Table A.2. Plan Characteristics

2008

Plan	Actuarial Value	Avg. Employee Premium ($)	Full Premium ($)	Deductible ($)	OOP Max. ($)	Market Share
Kaiser - 1	0.97	682	9,768	0	1,200	0.07
Kaiser - 2	0.96	313	9,334	0	2,000	0.10
Moda - 1	0.92	1,086	11,051	300	500	0.28
Moda - 2	0.89	648	10,613	300	1,000	0.06
Moda - 3	0.88	363	10,097	600	1,000	0.11
Moda - 4	0.86	461	9,674	900	1,500	0.07
Moda - 5	0.82	273	8,888	1,500	2,000	0.12
Moda - 6	0.78	320	8,032	3,000	3,000	0.03
Moda - 7	0.68	37	6,141	3,000	10,000	0.00
Providence - 1	0.96	1,005	10,645	900	1,200	0.14
Providence - 2	0.95	933	10,563	900	2,000	0.02

2010

Plan	Actuarial Value	Avg. Employee Premium ($)	Full Premium ($)	Deductible ($)	OOP Max. ($)	Market Share
Kaiser - 1	0.96	701	11,586	0	2,400	0.17
Kaiser - 2	0.95	582	11,231	0	3,000	0.03
Moda - 1	0.89	3,876	15,794	600	1,200	0.10
Moda - 2	0.86	2,867	14,579	600	1,500	0.01
Moda - 3	0.85	1,833	13,300	600	1,800	0.17
Moda - 4	0.84	897	11,904	900	2,000	0.12
Moda - 5	0.82	528	10,890	1,500	2,000	0.21
Moda - 6	0.78	311	9,795	3,000	3,000	0.09
Moda - 7	0.75	106	7,472	3,000	10,000	0.02
Providence - 1	0.91	4,702	16,680	1,200	1,200	0.04
Providence - 2	0.89	4,314	16,245	1,800	1,800	0.01

Table A.2. Plan Characteristics, cont.

2011

Plan	Actuarial Value	Avg. Employee Premium ($)	Full Premium ($)	Deductible ($)	OOP Max. ($)	Market Share
Kaiser - 1	0.95	520	11,051	0	2,400	0.16
Kaiser - 2	0.92	348	10,126	300	4,000	0.04
Moda - 1	0.86	3,414	15,622	600	4,500	0.06
Moda - 2	0.84	1,009	12,391	900	6,000	0.00
Moda - 3	0.84	1,208	12,688	900	6,000	0.15
Moda - 4	0.83	603	11,334	1,200	6,300	0.09
Moda - 5	0.82	367	10,188	1,500	6,600	0.24
Moda - 6	0.78	190	8,764	3,000	6,600	0.15
Moda - 7	0.75	130	7,806	3,000	10,000	0.05
Providence - 1	0.87	2,835	14,882	300	3,600	0.02
Providence - 2	0.84	2,066	13,891	900	6,000	0.00

2012

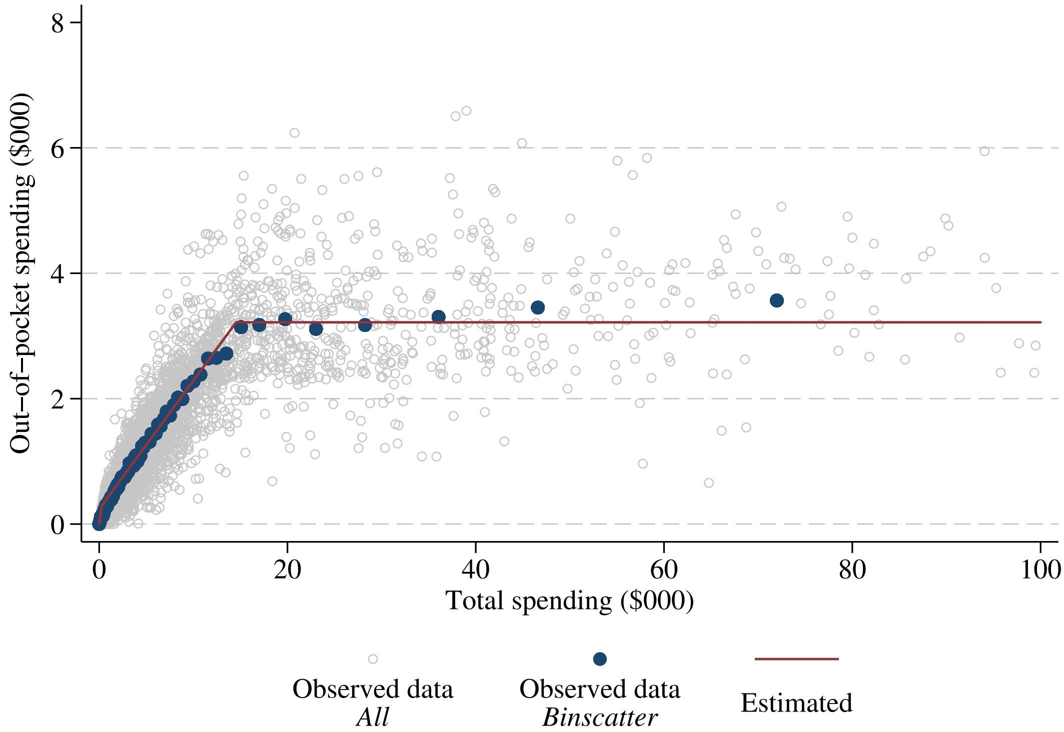
Plan	Actuarial Value	Avg. Employee Premium ($)	Full Premium ($)	Deductible ($)	OOP Max. ($)	Market Share
Kaiser - 1	0.95	1,478	13,408	0	2,400	0.18
Kaiser - 2	0.93	843	12,278	450	4,000	0.04
Moda - 1	0.87	5,677	18,514	600	4,500	0.06
Moda - 2	0.85	2,164	14,299	900	6,000	0.01
Moda - 3	0.85	2,995	15,359	900	6,000	0.12
Moda - 4	0.84	1,899	13,902	1,200	6,300	0.06
Moda - 5	0.83	1,082	12,670	1,500	6,600	0.22
Moda - 6	0.79	501	11,139	3,000	6,600	0.17
Moda - 7	0.76	148	8,395	3,000	10,000	0.11

2013

Plan	Actuarial Value	Avg. Employee Premium ($)	Full Premium ($)	Deductible ($)	OOP Max. ($)	Market Share
Kaiser - 1	0.95	1,815	14,203	0	3,000	0.20
Kaiser - 2	0.94	998	12,895	600	4,400	0.03
Moda - 1	0.87	6,537	19,675	600	6,000	0.03
Moda - 2	0.85	3,069	15,765	1,050	7,200	0.08
Moda - 3	0.84	1,152	13,157	1,500	7,800	0.22
Moda - 4	0.82	692	12,212	2,250	8,400	0.06
Moda - 5	0.80	493	11,427	3,000	9,000	0.11
Moda - 6	0.78	344	10,480	3,750	12,000	0.05
Moda - 7	0.77	151	8,574	3,000	10,000	0.13
Moda - 8	0.76	224	9,474	4,500	15,000	0.05

Notes: Average employee premium is taken across all employees, even those who did not choose a particular plan. Full premium reflects the premium negotiated between OEBB and the insurer; the full premium shown is for the employee plus spouse family type. The deductible and out-of-pocket maximum shown are for in-network services for a family household.

Figure A.1. Example of Plan Cost-sharing Features Estimation



Notes: The figure shows the data used to estimate the cost-sharing features for plan Moda - 3 for individual households in 2012. Each gray dot represents a household observation. The blue dots are a binned scatter plot of the gray data, using 100 points. The basic cost-sharing features of the plan (as observed in plan documents) are a deductible of $300, nonspecialist coinsurance rate of 20 percent, and in-network out-of-pocket maximum of $2,000. The red line shows the "estimated" cost-sharing function of the plan, which minimizes the sum of squared errors between predicted and observed out-of-pocket spending. The estimated coinsurance rate is 20.5 percent and the estimated out-of-pocket maximum is $3,218.

Table A.3. Estimated Plan Characteristics

2009	Individuals			Families		
Plan	Ded.	Coins.	OOP Max.	Ded.	Coins.	OOP Max.
Kaiser - 1	0	0.03	564	0	0.03	645
Kaiser - 2	0	0.03	684	0	0.04	760
Kaiser - 3	0	0.03	734	0	0.04	791
Moda - 1	100	0.10	1,613	300	0.10	2,009
Moda - 2	100	0.18	1,922	300	0.15	2,662
Moda - 3	200	0.20	2,081	600	0.15	3,062
Moda - 4	300	0.19	2,796	900	0.15	3,835
Moda - 5	500	0.22	3,164	1,500	0.16	4,296
Moda - 6	1,000	0.22	3,713	3,000	0.12	5,422
Moda - 7	1,500	0.42	4,693	3,000	0.30	8,086
Providence - 1	300	0.02	790	900	0.00	900
Providence - 2	300	0.03	867	900	0.00	986
Providence - 3	300	0.04	1,116	900	0.01	1,296

2010	Individuals			Families		
Plan	Ded.	Coins.	OOP Max.	Ded.	Coins.	OOP Max.
Kaiser - 1	0	0.03	697	0	0.04	805
Kaiser - 2	0	0.04	820	0	0.05	885
Moda - 1	200	0.14	2,526	600	0.12	3,430
Moda - 2	200	0.21	2,846	600	0.18	3,967
Moda - 3	200	0.21	3,189	600	0.18	4,299
Moda - 4	300	0.22	3,109	900	0.18	4,079
Moda - 5	500	0.22	3,321	1,500	0.16	4,572
Moda - 6	1,000	0.22	3,844	3,000	0.12	5,684
Moda - 7	1,500	0.19	4,913	3,000	0.15	7,579
Providence - 1	400	0.05	1,523	1,200	0.02	1,851
Providence - 2	600	0.06	1,998	1,800	0.02	2,473

2011	Individuals			Families		
Plan	Ded.	Coins.	OOP Max.	Ded.	Coins.	OOP Max.
Kaiser - 1	0	0.04	883	0	0.06	974
Kaiser - 2	100	0.06	1,340	300	0.06	1,831
Moda - 1	200	0.22	2,608	600	0.18	4,316
Moda - 2	300	0.22	3,201	900	0.17	5,094
Moda - 3	300	0.22	3,246	900	0.17	5,202
Moda - 4	400	0.22	3,324	1,200	0.17	5,367
Moda - 5	500	0.22	3,529	1,500	0.16	5,727
Moda - 6	1,000	0.22	4,061	3,000	0.13	6,728
Moda - 7	1,500	0.21	4,914	3,000	0.15	7,663
Providence - 1	100	0.18	2,164	300	0.16	3,496
Providence - 2	300	0.15	2,911	900	0.13	4,378

Notes: The table shows plan deductibles, estimated coinsurance rates, and estimated out-of-pocket maximums. The estimation procedure is described in Section A.2.

Table A.3. Estimated Plan Characteristics, cont.

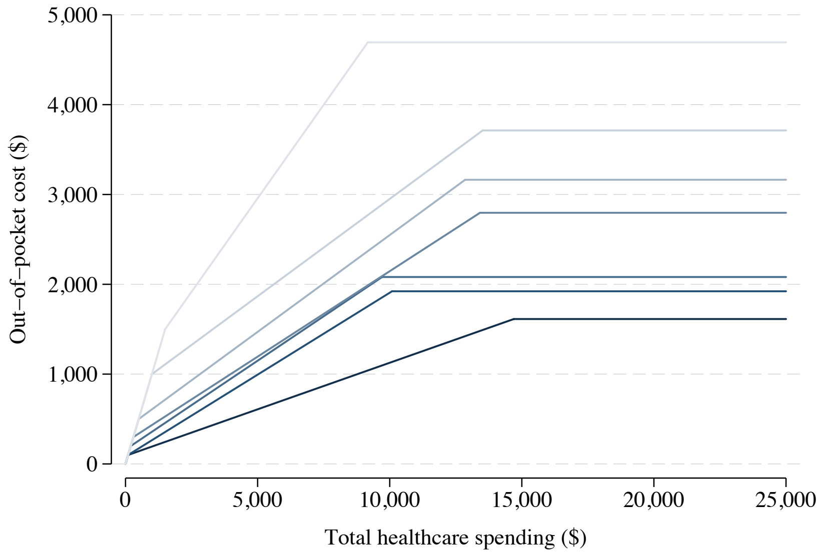
2012	Individuals			Families		
Plan	Ded.	Coins.	OOP Max.	Ded.	Coins.	OOP Max.
Kaiser - 1	0	0.04	911	0	0.06	995
Kaiser - 2	150	0.07	1,709	450	0.05	2,160
Moda - 1	200	0.21	2,571	600	0.17	4,154
Moda - 2	300	0.21	3,187	900	0.17	4,981
Moda - 3	300	0.20	3,218	900	0.17	5,025
Moda - 4	400	0.21	3,291	1,200	0.16	5,104
Moda - 5	500	0.21	3,493	1,500	0.16	5,498
Moda - 6	1,000	0.21	4,000	3,000	0.12	6,608
Moda - 7	1,500	0.21	4,927	3,000	0.15	7,662

2013	Individuals			Families		
Plan	Ded.	Coins.	OOP Max.	Ded.	Coins.	OOP Max.
Kaiser - 1	0	0.04	911	0	0.06	1,040
Kaiser - 2	200	0.03	867	600	0.01	951
Moda - 1	200	0.20	3,237	600	0.17	4,893
Moda - 2	350	0.20	3,842	1,050	0.16	5,647
Moda - 3	500	0.20	4,175	1,500	0.15	6,160
Moda - 4	750	0.20	4,704	2,250	0.14	6,989
Moda - 5	1,000	0.19	5,186	3,000	0.12	7,714
Moda - 6	1,250	0.19	6,414	3,750	0.12	9,187
Moda - 7	1,500	0.21	4,865	3,000	0.15	7,650
Moda - 8	1,500	0.19	7,620	4,500	0.11	10,614

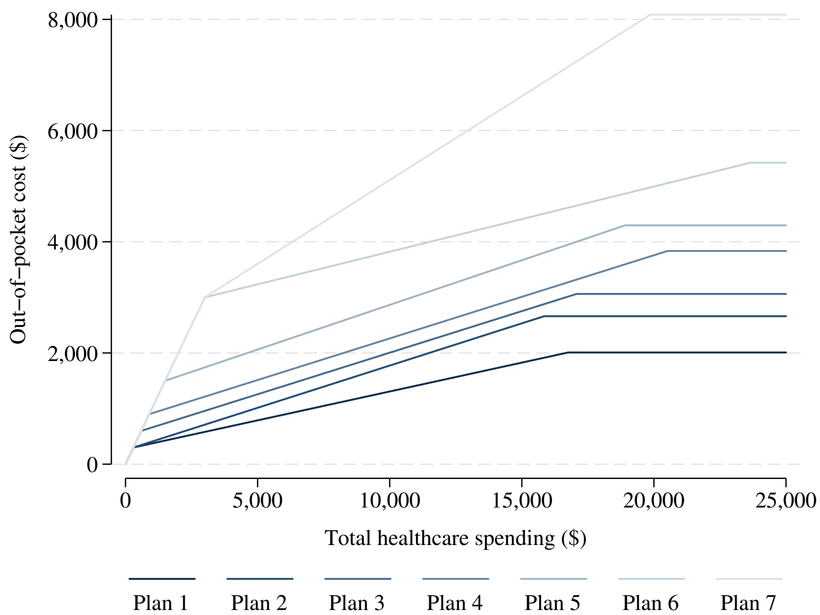
Notes: The table shows plan deductibles, estimated coinsurance rates, and estimated out-of-pocket maximums. The estimation procedure is described in Section A.2.

Figure A.2. Out-of-pocket Cost Functions for Moda Plans, 2009

(a) Individuals



(b) Families



Notes: The figure shows the estimated out-of-pocket cost functions for Moda plans in 2009. Cost-sharing features are estimated separately for individuals and families; the procedure is described in Section A.2. Note that the plans are not truly vertically differentiated. Graphically, this would mean that a weakly better plan is everywhere weakly below and weakly flatter than a worse plan, which is not the case here.

Table A.4. Plan Menu Generosity and Household Health

	2008	2009	2010	2011	2012	2013
Household Risk Score	-0.006	0.017	0.020	0.002	0.006	0.000
	(0.039)	(0.016)	(0.011)*	(0.009)	(0.010)	(0.012)
Family Type						
Employee Alone	0.000†	0.000†	0.000†	0.000†	0.000†	0.000†
Employee + Spouse	-1.389	-1.369	-1.498	-1.040	-1.626	-1.612
	(0.077)***	(0.040)***	(0.029)***	(0.025)***	(0.026)***	(0.031)***
Employee + Child	-0.542	-0.634	-0.907	-0.616	-1.092	-0.937
	(0.084)***	(0.053)***	(0.039)***	(0.031)***	(0.031)***	(0.037)***
Employee + Family	-1.792	-1.882	-1.804	-1.306	-2.147	-2.102
	(0.064)***	(0.037)***	(0.028)***	(0.023)***	(0.025)***	(0.029)***
Dependent variable mean	88.7	88.5	84.6	82.7	83.3	82.6
R^2	0.020	0.084	0.154	0.115	0.242	0.220
Number of observations	37,666	31,074	29,538	29,279	27,897	24,283

Notes: The level of observation is the household. The dependent variable is plan menu generosity, as measured by predicted actuarial value conditional on choosing Moda, $\widehat{AV}_{d,Moda}$. $\widehat{AV}_{d,Moda}$ is calculated according to equation (12), and it is multiplied by 100 to increase parameter magnitudes. Household risk score is the mean risk score among all individuals in a household, and has been z-scored such that the variable has a mean of zero and a standard deviation of one within each year. As we do not have data before 2008, the 2008 regression uses risk scores calculated using 2008 claims data. *p <0.10, **p <0.05, ***p <.01. †By normalization.

Table A.5. Explaining Plan Menu Generosity: 2008

	(1)	(2)	(3)	(4)
Household Risk Score	-0.006	0.016	0.011	0.025
	(0.039)	(0.039)	(0.038)	(0.040)
Family Type				
Employee Alone	0.000†	0.000†	0.000†	0.000†
Employee + Spouse	-1.389	-1.374	-1.251	-1.085
	(0.077)***	(0.083)***	(0.083)***	(0.085)***
Employee + Child	-0.542	-0.535	-0.478	-0.462
	(0.084)***	(0.085)***	(0.084)***	(0.082)***
Employee + Family	-1.792	-1.819	-1.688	-1.437
	(0.064)***	(0.071)***	(0.071)***	(0.074)***
Part-time		-0.428	-0.448	-0.867
		(0.133)***	(0.133)***	(0.139)***
Occupation Type				
Admin.		-1.745	-1.883	-2.685
		(0.455)***	(0.459)***	(0.501)***
Classified		-0.598	-0.469	-0.155
		(0.283)**	(0.414)	(0.457)
Comm. Coll. Fac.		0.553	1.138	1.044
		(0.287)*	(0.430)***	(0.470)**
Comm. Coll. Non-Fac.		0.671	0.457	0.077
		(0.288)**	(0.288)	(0.302)
Confidential		-2.759	-2.883	-3.133
		(0.855)***	(0.856)***	(0.915)***
Licensed		0.001	1.645	1.628
		(0.278)	(0.459)***	(0.505)***
Substitute		-11.051	-9.312	-9.354
		(0.283)***	(0.457)***	(0.496)***
Union Affiliation				
AFT			0.251	-0.398
			(0.374)	(0.432)
IAFE			0.758	1.222
			(0.404)*	(0.458)***
OACE			2.671	1.617
			(0.389)***	(0.449)***
OEA			-1.799	-1.765
			(0.434)***	(0.491)***
OSEA			-0.086	-0.426
			(0.395)	(0.449)
District characteristics				
ln(HPI)				-0.876
				(0.085)***
Pct. Republican				-14.077
				(0.467)***
Dependent variable mean	88.7	89.0	89.1	98.3
R^2	0.020	0.031	0.046	0.073
Number of observations	37,666	37,666	37,666	35,698

Notes: The level of observation is the household. The dependent variable is plan menu generosity, as measured by predicted actuarial value conditional on choosing Moda, $\widehat{AV}_{d,Moda}$. $\widehat{AV}_{d,Moda}$ is calculated according to equation (12), and it is multiplied by 100 to increase parameter magnitudes. Household risk score is the mean risk score among all individuals in a household, and has been z-scored such that the variable has a mean of zero and a standard deviation of one within each year. As we do not have data before 2008, the 2008 regression uses risk scores calculated using 2008 claims data. *p <0.10, **p <0.05, ***p <.01. †By normalization.

Table A.6. Household Summary Statistics (2008)

Sample demographics	2008
Number of households	45,012
Number of enrollees	116,267
Enrollee age, mean (med.)	38.2
	(35.8)
Premiums	
Employee premium ($), mean (med.)	596
	(0)
Full premium ($), mean (med.)	10,107
	(10,605)
Household health spending	
Total spending ($), mean (med.)	9,956
	(4,485)
Out-of-pocket ($), mean (med.)	957
	(620)

Notes: Summary statistics are shown for households in the 2008 analysis sample used in our reduced-form analysis. Enrollees are employees plus their covered dependents. Statistics for premiums are for households' chosen plans, rather than for all possible plans. Sample medians are shown in parentheses.

Table A.7. Plan Choice Logit Model

	2008	2009	2010	2011	2012	2013
Employee premium ($000)	-0.789	-0.674	-0.505	-0.372	-0.515	-0.490
	(0.017)	(0.014)	(0.008)	(0.010)	(0.008)	(0.008)
HRA/HSA contrib. ($000)	0.112		0.358	0.134	0.269	0.534
	(0.759)		(0.044)	(0.024)	(0.019)	(0.015)
Vision/dental contrib. ($000)	0.654	0.408	0.480	0.794	0.553	0.710
	(0.021)	(0.022)	(0.019)	(0.017)	(0.017)	(0.017)
Kaiser - 1	-0.771	-0.728				
	(0.026)	(0.030)				
Kaiser - 2	-1.287	-1.112	-0.846	-0.469	-0.375	-0.074
	(0.031)	(0.032)	(0.034)	(0.035)	(0.034)	(0.044)
Kaiser - 3		-1.563	-1.042	-0.985	-1.629	-1.820
		(0.384)	(0.056)	(0.051)	(0.048)	(0.058)
Moda - 1	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$
Moda - 2	-1.113	-1.184	-0.911	-2.088	-2.578	-0.593
	(0.026)	(0.032)	(0.058)	(0.163)	(0.072)	(0.045)
Moda - 3	-1.226	-1.110	-0.518	-0.373	-0.389	-0.957
	(0.022)	(0.025)	(0.029)	(0.034)	(0.033)	(0.046)
Moda - 4	-1.751	-1.540	-1.356	-1.192	-1.554	-2.261
	(0.028)	(0.030)	(0.034)	(0.037)	(0.039)	(0.055)
Moda - 5	-1.951	-1.881	-1.341	-0.878	-0.999	-2.391
	(0.034)	(0.037)	(0.040)	(0.039)	(0.037)	(0.055)
Moda - 6	-2.785	-2.871	-2.205	-1.406	-1.917	-3.182
	(0.048)	(0.051)	(0.050)	(0.043)	(0.046)	(0.065)
Moda - 7	-4.391	-4.260	-3.388	-1.959	-3.007	-3.492
	(0.098)	(0.098)	(0.074)	(0.050)	(0.060)	(0.073)
Moda - 8						-3.679
						(0.068)
Providence - 1	0.001	0.048	0.135	-0.778		
	(0.019)	(0.028)	(0.038)	(0.053)		
Providence - 2	-0.600	-0.314				
	(0.043)	(0.049)				
Providence - 3		-0.048	-0.159	-0.939		
		(0.078)	(0.083)	(0.436)		
Number of observations	163,431	121,744	116,541	114,527	163,278	163,683

Notes: This table presents parameter estimates from the conditional logit model described by equation (11), presented separately for each year. An observation is a household-plan. Moda - 1 (the highest coverage Moda plan) is the omitted plan. †By normalization.

Table A.8. Estimates of Moral Hazard

	OLS *All*	IV *All*	IV *Individuals*	IV *Families*
	(1)	(2)	(3)	(4)
$\log(1-AV_{j(k)})\mathbf{1}_{f(k)=Moda}$	-0.580	-0.269		
	(0.053)***	(0.084)***		
$\log(1-AV_{j(k)})\mathbf{1}_{f(k)=Moda} \times Q_1$			-0.220	-0.415
			(0.290)	(0.131)***
$\log(1-AV_{j(k)})\mathbf{1}_{f(k)=Moda} \times Q_2$			-0.410	-0.235
			(0.189)**	(0.088)***
$\log(1-AV_{j(k)})\mathbf{1}_{f(k)=Moda} \times Q_3$			-0.253	-0.218
			(0.136)*	(0.090)**
$\log(1-AV_{j(k)})\mathbf{1}_{f(k)=Moda} \times Q_4$			-0.017	0.074
			(0.346)	(0.145)
R^2	0.19	0.19	0.44	0.37
Observations	35,146	35,146	8,962	26,184

Notes: This table shows the OLS and IV estimates of equation (15), describing the relationship between household total spending and plan generosity. The unit of observation is a household, and the dependent variable is log of 1 + total spending. In columns 3 and 4, coefficients can vary by household risk quartile Q_n; Q_4 is the sickest households. Columns 1 and 2 are estimated on all households, while columns 3 and 4 are estimated only on individual or family households, respectively. All specifications also include insurer fixed effects and controls for household risk score and family structure. Standard errors (in parentheses) are clustered by household plan menu, of which there are 533 among individual households and 1,750 among family households. We can reject the hypothesis that the four coefficients are equal at the 10 percent level for families, but not for individuals. *p <0.10, **p <0.05, ***p <.01.

Table A.9. Plan Choice Logit Model by Family Status and Risk Quartile, 2008

	Ind. Q_1	Fam. Q_1	Ind. Q_2	Fam. Q_2	Ind. Q_3	Fam. Q_3	Ind. Q_4	Fam. Q_4
Employee premium ($000)	-1.602***	-1.014***	-1.345***	-1.019***	-1.401***	-0.949***	-1.302***	-0.870***
	(0.128)	(0.047)	(0.114)	(0.049)	(0.113)	(0.053)	(0.108)	(0.056)
Vision/dental contrib. ($000)	1.301***	0.943***	1.254***	0.884***	1.089***	0.621***	1.042***	0.495***
	(0.092)	(0.061)	(0.094)	(0.065)	(0.094)	(0.071)	(0.099)	(0.076)
HSA/HRA contrib. ($000)				-6.871		2.774***		-6.703
				(318.561)		(1.068)		(526.706)
Kaiser - 1	-0.074	1.351**	-1.452**	-0.856	1.069	0.863	2.149***	0.525
	(0.420)	(0.531)	(0.671)	(0.747)	(0.799)	(0.918)	(0.782)	(0.801)
Kaiser - 2	0.575	1.765***	-0.960	-0.278	1.483*	1.376	2.468***	1.135
	(0.410)	(0.517)	(0.657)	(0.731)	(0.791)	(0.899)	(0.774)	(0.789)
Moda - 1	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$	$0.000^{\dagger}$
Moda - 2	-1.175***	-0.425***	-1.077***	-1.011***	-0.498*	-0.571**	-0.644**	-0.930***
	(0.185)	(0.161)	(0.242)	(0.215)	(0.260)	(0.254)	(0.270)	(0.214)
Moda - 3	-0.865***	-0.298	-0.880***	-1.162***	-0.290	-0.395	-0.108	-0.810**
	(0.202)	(0.240)	(0.332)	(0.334)	(0.372)	(0.399)	(0.383)	(0.333)
Moda - 4	-1.265***	-0.331	-1.535***	-1.719***	-0.370	-0.535	-0.100	-1.194**
	(0.280)	(0.349)	(0.477)	(0.488)	(0.534)	(0.584)	(0.553)	(0.486)
Moda - 5	-1.083***	-0.065	-1.419**	-1.896**	0.386	-0.119	0.623	-1.029
	(0.407)	(0.527)	(0.713)	(0.740)	(0.805)	(0.885)	(0.832)	(0.737)
Moda - 6	-1.053*	-0.086	-1.903*	-2.678**	0.515	-0.517	1.390	-1.634
	(0.592)	(0.770)	(1.048)	(1.084)	(1.171)	(1.295)	(1.210)	(1.082)
Moda - 7	-2.060**	0.093	-3.330*	-5.027***	0.880	-0.940	1.879	-1.986
	(0.997)	(1.304)	(1.757)	(1.854)	(1.968)	(2.225)	(2.058)	(1.842)
Providence - 1	-0.251	1.141*	-1.448*	-0.696	0.474	2.210**	0.840	-0.613
	(0.566)	(0.659)	(0.863)	(0.850)	(0.920)	(0.938)	(0.922)	(0.747)
Providence - 2	0.300	1.533**	-1.022	-0.194	1.017	2.809***	1.215	-0.121
	(0.536)	(0.639)	(0.836)	(0.830)	(0.894)	(0.915)	(0.915)	(0.728)
Number of osbervations	8,487	25,054	8,367	25,416	8,285	25,393	8,077	25,326

Notes: The table presents the results of estimating equation (11) separately by quartile of household risk score within individual and family households in 2008. The columns indicate which sample is being used: Individuals (Ind.) versus families (Fam.) and the household risk quartile Q_n, where Q_4 is the sickest households. The omitted plan fixed effect is for Moda plan 1 (the most generous Moda plan). The coefficient on employee premium (measured in thousands of dollars) is normalized to -1. *p <0.10, **p <0.05, ***p <.01. †By normalization.

Table A.10. Spending Distributions and Moda Plan Characteristics, 2008

Panel A: Total Spending Distributions by Risk Quartile

	Percentile of total spending				
Risk quartile	10th	25th	50th	75th	90th
Individuals					
$Q1$	0	30	381	851	1,454
$Q2$	293	721	1,286	1,984	3,025
$Q3$	782	1,688	2,861	4,266	5,987
$Q4$	1,869	4,134	7,155	12,765	21,240
Families					
$Q1$	418	985	1,959	3,508	6,718
$Q2$	1,489	2,567	4,212	6,584	10,984
$Q3$	3,373	5,261	7,811	11,745	17,301
$Q4$	5,096	9,820	15,401	22,637	29,615

Panel B: Plan Characteristics

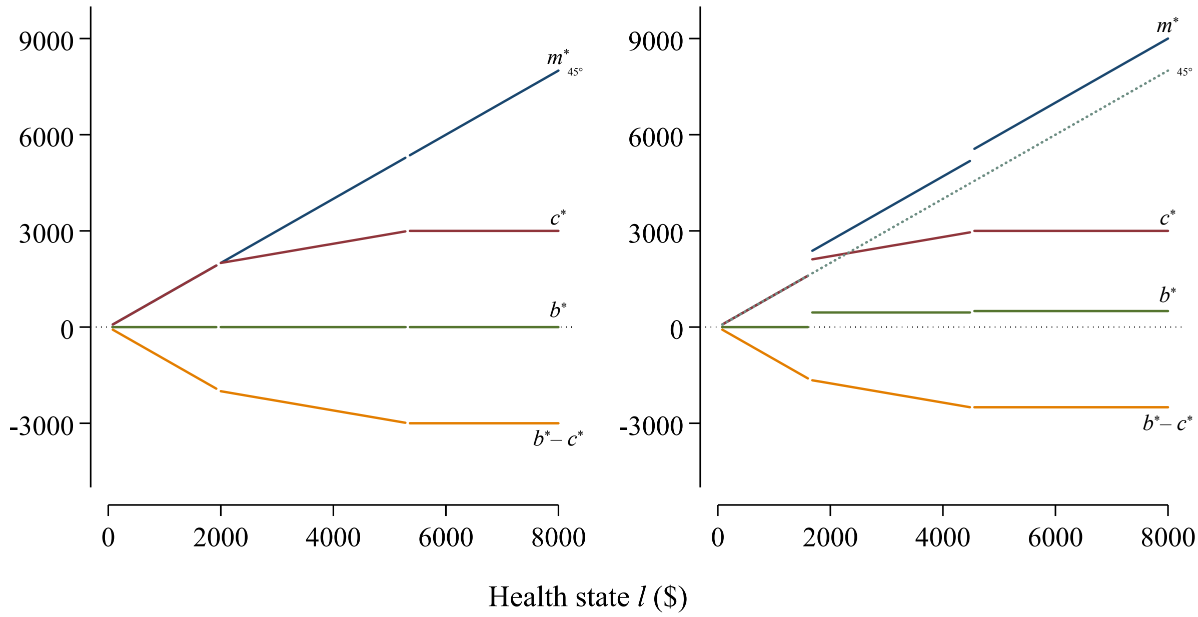
	Moda plan						
	Plan 1	Plan 2	Plan 3	Plan 4	Plan 5	Plan 6	Plan 7
Individuals							
Deductible	100	100	200	300	500	1,000	1,500
OOP Max.	500	1,000	1,000	1,500	2,000	2,000	5,000
Families							
Deductible	300	300	600	900	1,500	3,000	3,000
OOP Max.	500	1,000	1,000	1,500	2,000	3,000	10,000

Notes: The table shows the distributions of household realized total healthcare spending and the plan characteristics of Moda plans in 2008. Panel A shows the spending distributions, by quartile of household risk score within Individual and Family households. Panel B shows the in-network deductible and out-of-pocket maximum (OOP Max.) for each of the Moda plans.

Figure A.3. Healthcare Spending Choice Example

(a) No Moral Hazard ($\omega \approx \$0$)

(b) Moral Hazard ($\omega = \$1{,}000$)



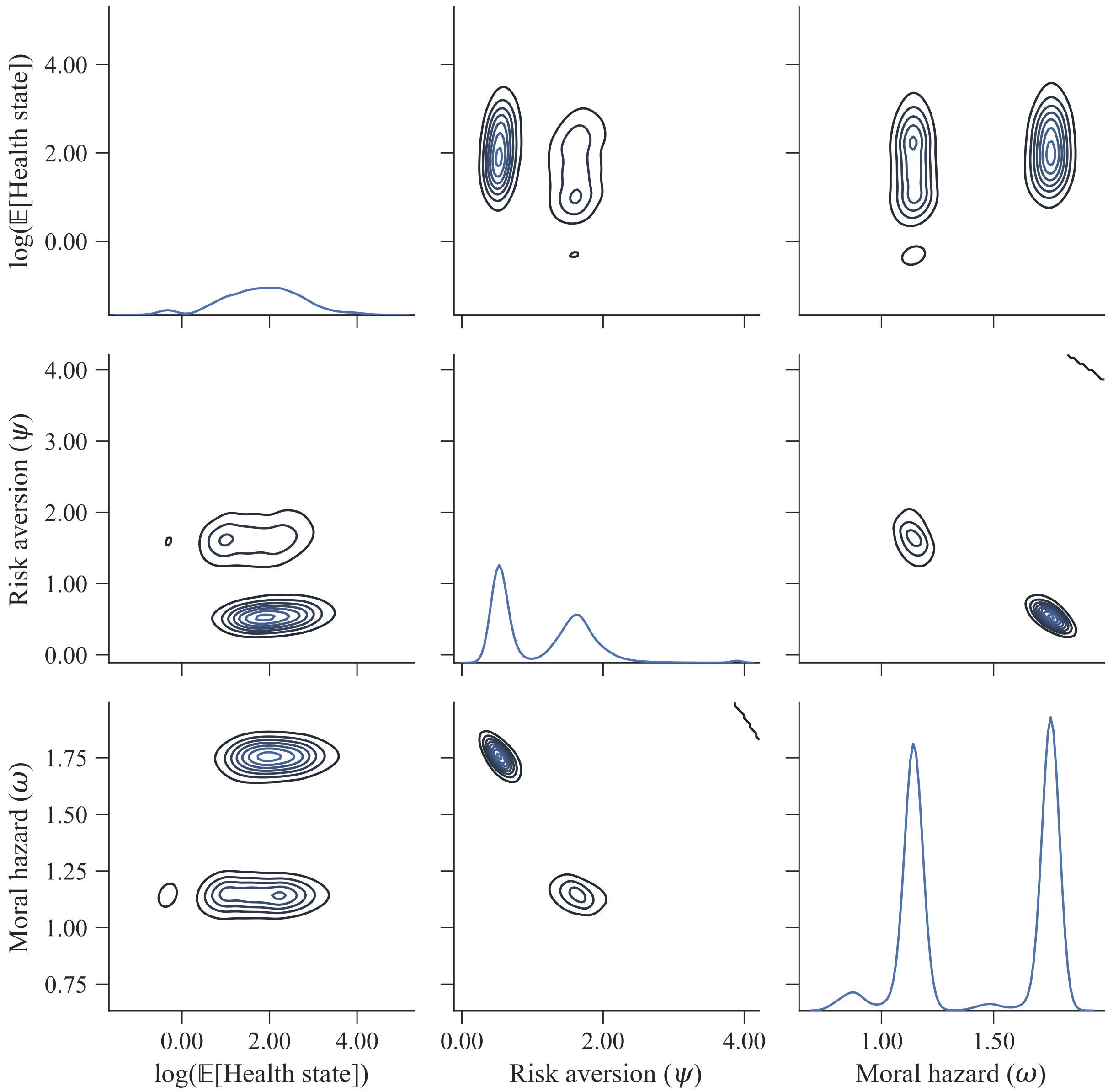
Notes: The figure shows optimal healthcare spending predicted by our specification of household preferences over healthcare utilization (equation (4)), given an insurance contract with a deductible of \$2,000, a coinsurance rate of 30%, and an out-of-pocket maximum of \$3,000. Predicted behavior is shown under (a) no moral hazard and (b) under some moral hazard ($\omega = \$1{,}000$). The horizontal axis shows possible health state realizations. Optimal total healthcare spending is m^*: When there is no moral hazard, it is optimal to have total spending equal to the health state. The vertical axis also shows the corresponding out-of-pocket costs c^*, utility from healthcare utilization b^*, and net payoff from healthcare utilization $b^* - c^*$. Households face a lottery over payoff $b^* - c^*$.

Table A.11. Additional Parameter Estimates

	(1)		(2)		(3)	
Variable	Parameter	Std. Err.	Parameter	Std. Err.	Parameter	Std. Err.
Insurer fixed effects						
Kaiser * (Age−40) ($000s)	−0.073	0.005	−0.078	0.005	−0.071	0.005
Providence * (Age−40) ($000s)	−0.073	0.008	−0.122	0.009	−0.074	0.008
Kaiser * **1**[Children] ($000s)	−1.608	0.119	−1.509	0.120	−0.546	0.124
Providence * **1**[Children] ($000s)	−1.373	0.174	−2.116	0.199	−0.480	0.177
Kaiser * Region 1 ($000s)	−1.692	0.093	−1.477	0.091	−1.976	0.095
Kaiser * Region 2 ($000s)	−5.112	0.254	−4.949	0.254	−5.343	0.252
Providence * Region 1 ($000s)	−4.420	0.156	−3.899	0.158	−4.530	0.159
Providence * Region 2 ($000s)	−5.727	0.211	−5.301	0.213	−5.701	0.213
Providence * Region 3 ($000s)	−5.153	0.233	−4.716	0.235	−5.633	0.234
Health state distributions						
κ	0.167	0.002				
κ * Risk QT 1			0.123	0.004	0.184	0.000
κ * Risk QT 2			0.174	0.004	0.201	0.000
κ * Risk QT 3			0.162	0.004	0.302	0.000
κ * Risk QT 4			0.095	0.037	0.182	0.022
κ * Risk QT <4 * Risk score			0.156	0.023	0.270	0.017
μ	0.618	0.006				
μ * Female 18–30			0.142	0.014	0.059	0.016
μ * Age $<$ 18			0.020	0.014	−0.015	0.016
μ * Risk QT 1			−0.267	0.025	−0.421	0.021
μ * Risk QT 2			0.555	0.012	0.212	0.010
μ * Risk QT 3			0.709	0.008	0.420	0.007
μ * Risk QT 4			1.355	0.015	1.279	0.013
μ * Risk QT <4 * Risk score			1.025	0.016	1.184	0.018
μ * Risk QT 4 * Risk score			0.311	0.005	0.326	0.004
σ	1.117	0.002				
σ * Risk QT 1			1.408	0.010	1.450	0.008
σ * Risk QT 2			1.129	0.005	1.392	0.004
σ * Risk QT 3			1.067	0.003	1.244	0.003
σ * Risk QT 4			0.992	0.005	1.047	0.005

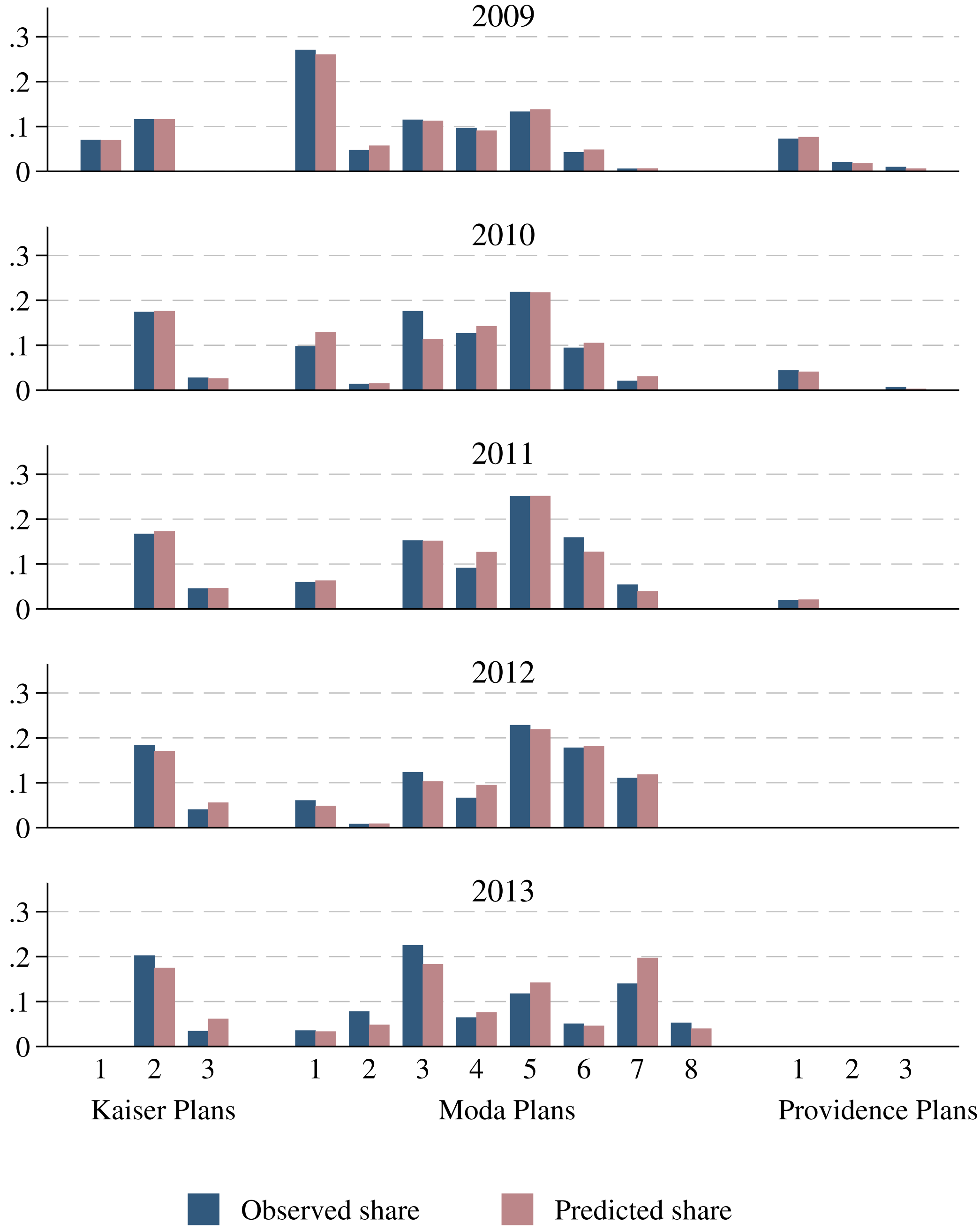
Notes: This table presents the parameter estimates that were not presented in Table 3. Column 1 estimates a model without observable heterogeneity in health, while columns 2 and 3 include this. "Risk QT #" is an indicator for an individual's risk quartile, where "Risk QT 4" is the sickest individuals. To make non-interacted coefficients more readily interpretable, we use (Age−40). Higher risk scores correspond to worse predicted health.

Figure A.4. Joint Distribution of Household Types



Notes: The figure shows the joint distribution of household types implied by the estimates in column 3 of Table 3. The diagonals show the one-way distributions of each parameter across households, and the off-diagonals show bivariate distributions. Households are assigned to a particular type according to the procedure described in Section B.4. Because expected health state can vary over years within a household, for the purposes of this figure we use the first year a household appears in the data. Expected health state is equal to a household's expected total spending absent moral hazard.

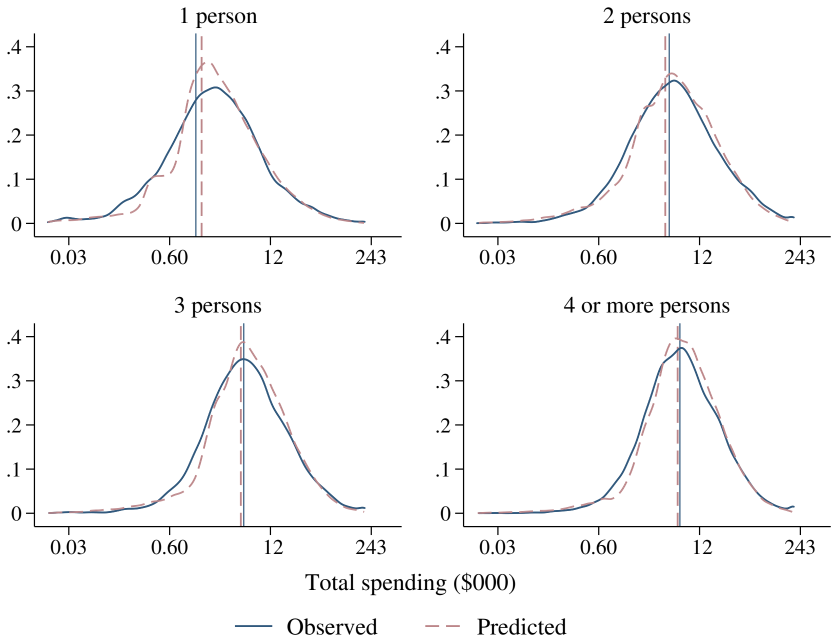
Figure A.5. Model Fit: Plan Choices Year by Year



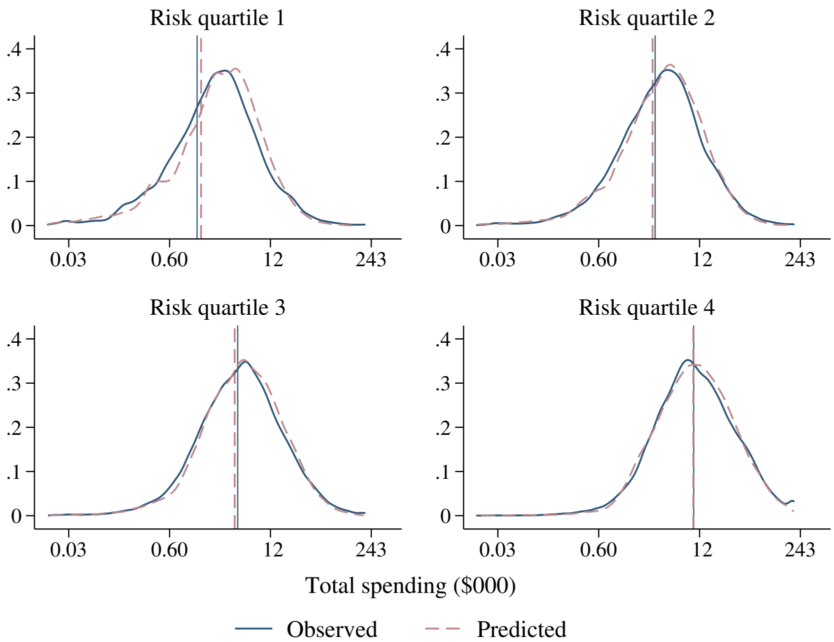
Notes: The figures show predicted and observed market shares at the plan level. In each year, the level of observation is the household. Predicted shares are estimated using the parameters in column 3 of Table 3.

Figure A.6. Model Fit in Subpopulations

(a) By Number of Family Members



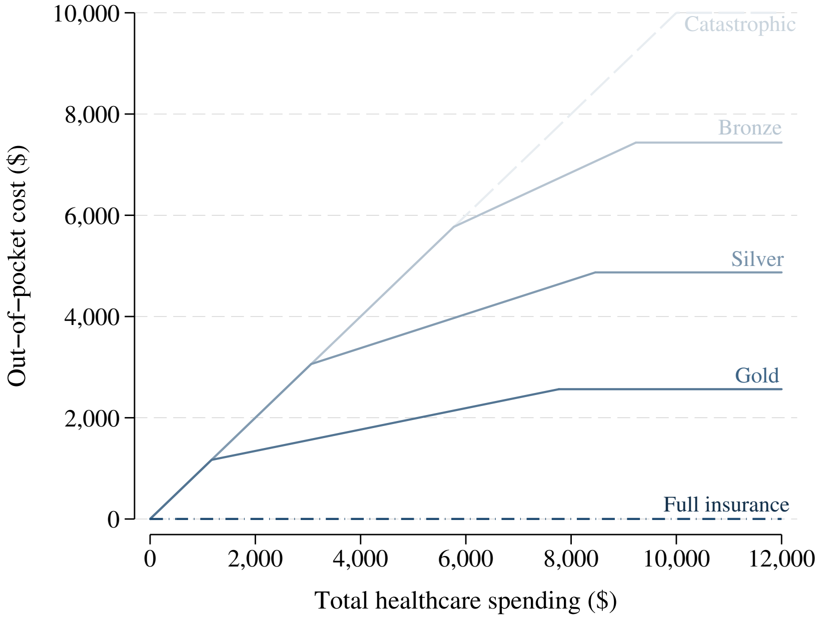
(b) By Household Health Risk



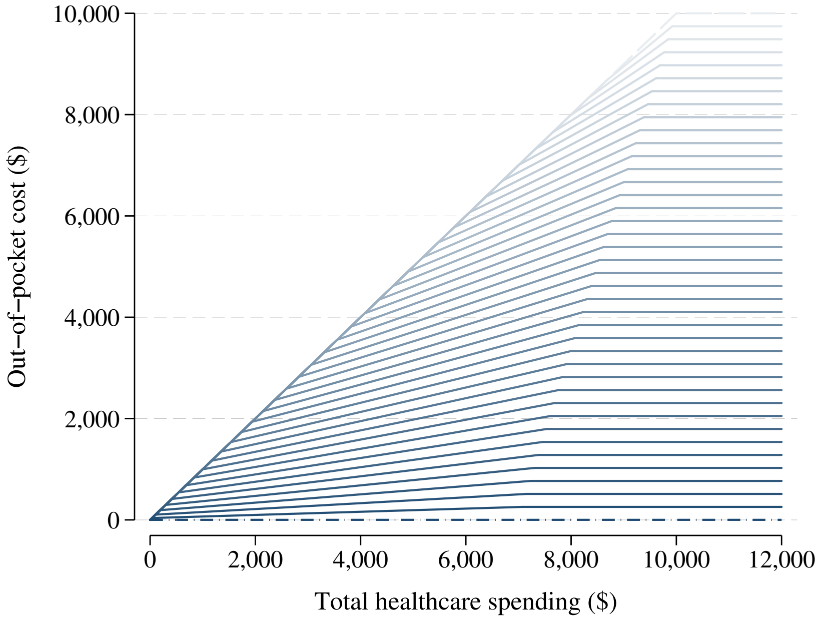
Notes: The figure shows kernel density plots of the predicted and observed distribution of total healthcare spending on a log scale. All years are pooled together, so an observation is a household-year. Vertical lines represent the mean of the respective distribution. Predicted distributions are estimated using the parameters in column 3 of Table 3. Household health risk is measured as the mean risk score of individuals in the household. Quartile 4 is the sickest households.

Figure A.7. Counterfactual Potential Contracts: Out-of-pocket Cost Functions

(a) Set of 5 Potential Contracts

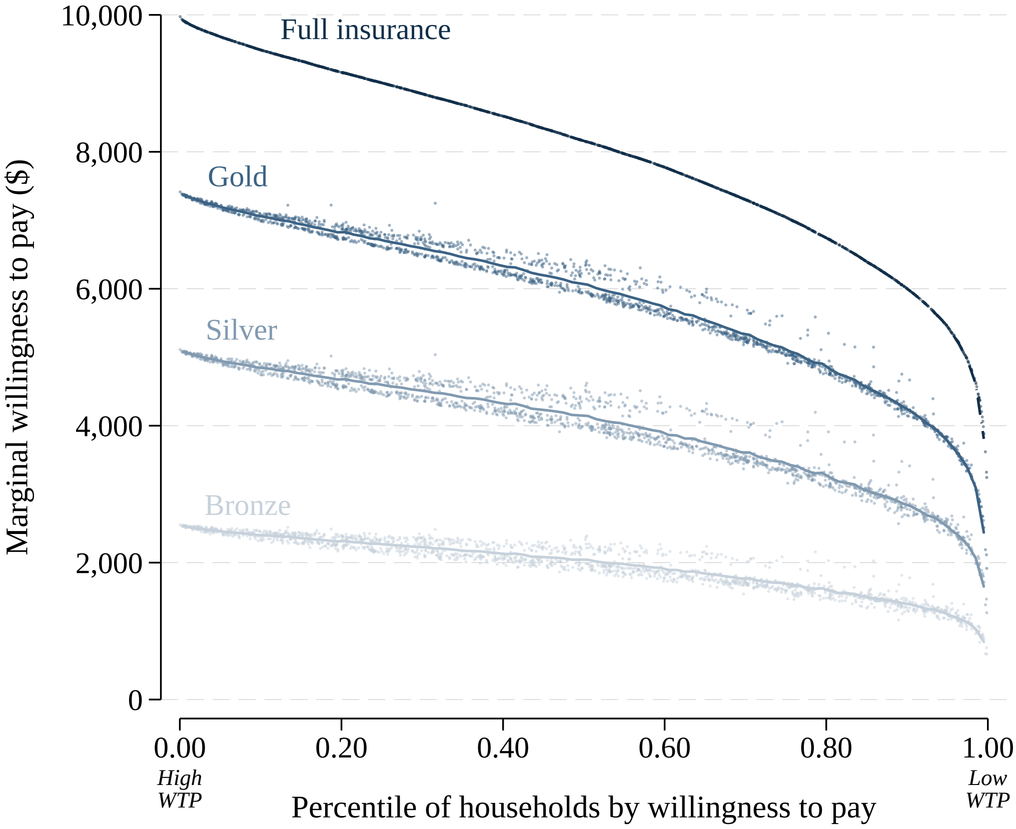


(b) Set of 40 Potential Contracts



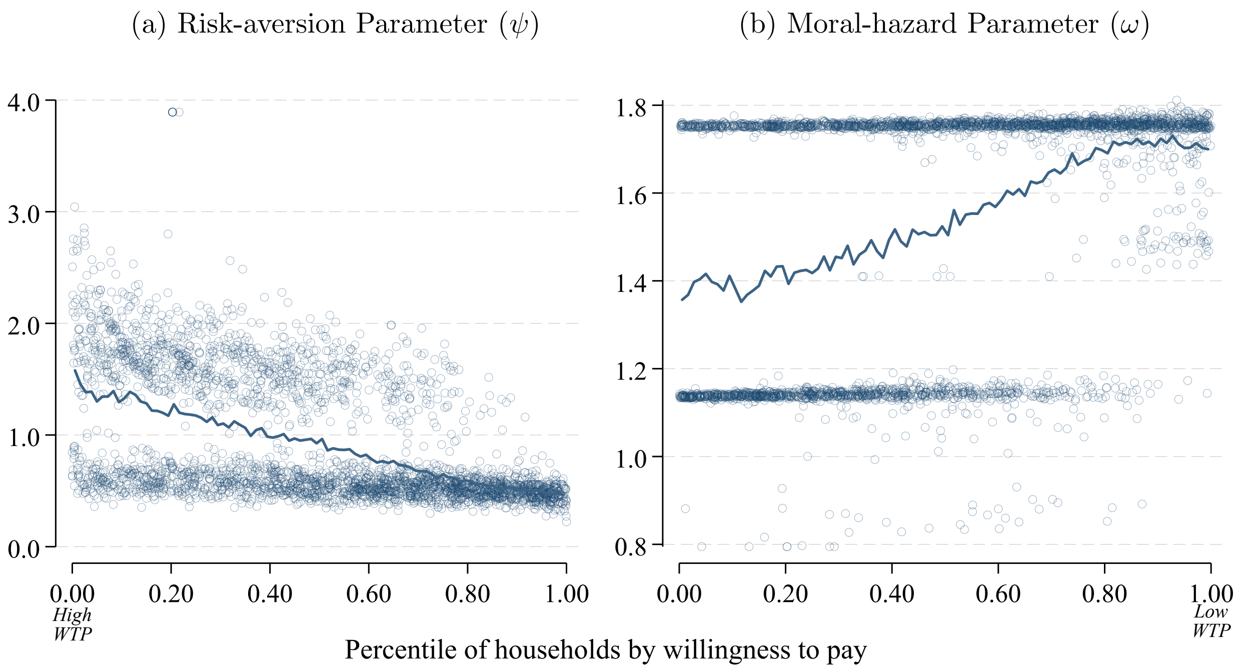
Notes: The figure shows the out-of-pocket cost functions for two sets of potential contracts we consider. Panel (a) shows the subset of five contracts we use for our graphical analysis and in our discussion of counterfactuals. The exact deductibles, coinsurance rates, and out-of-pocket maximums are $1,169, 21%, $2,564 for Gold; $3,060, 34%, $4,872 for Silver; and $5,771, 48%, $7,436 for Bronze; and $10,000, –, $10,000 for Catastrophic. Panel (b) shows the set of 40 contracts we consider in Section C.1. Note than for vertical differentiation, the contracts must be ordered in the sense of second-order stochastic dominance. Graphically, this means that a weakly better contract must be everywhere weakly below and weakly flatter than a worse contract.

Figure A.8. Distribution of Willingness to Pay Across Households



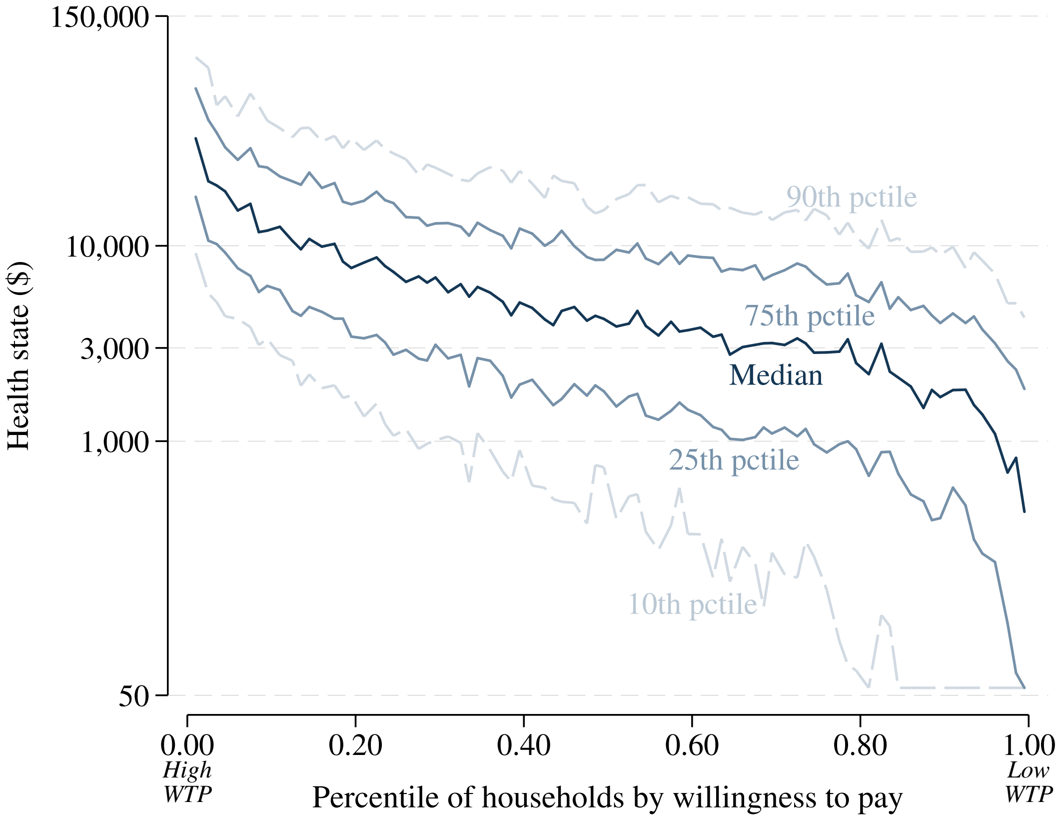
Notes: The figure shows the distribution of willingness to pay across households. Households are ordered on the horizontal axis according to their willingness to pay for full insurance. The line for each plan is a connected binned scatter plot with respect to 100 bins of households; these correspond exactly to Figure 5a. Dots show the underlying data for a 5 percent random sample of households.

Figure A.9. Risk Aversion and Moral Hazard Parameters by Willingness to Pay



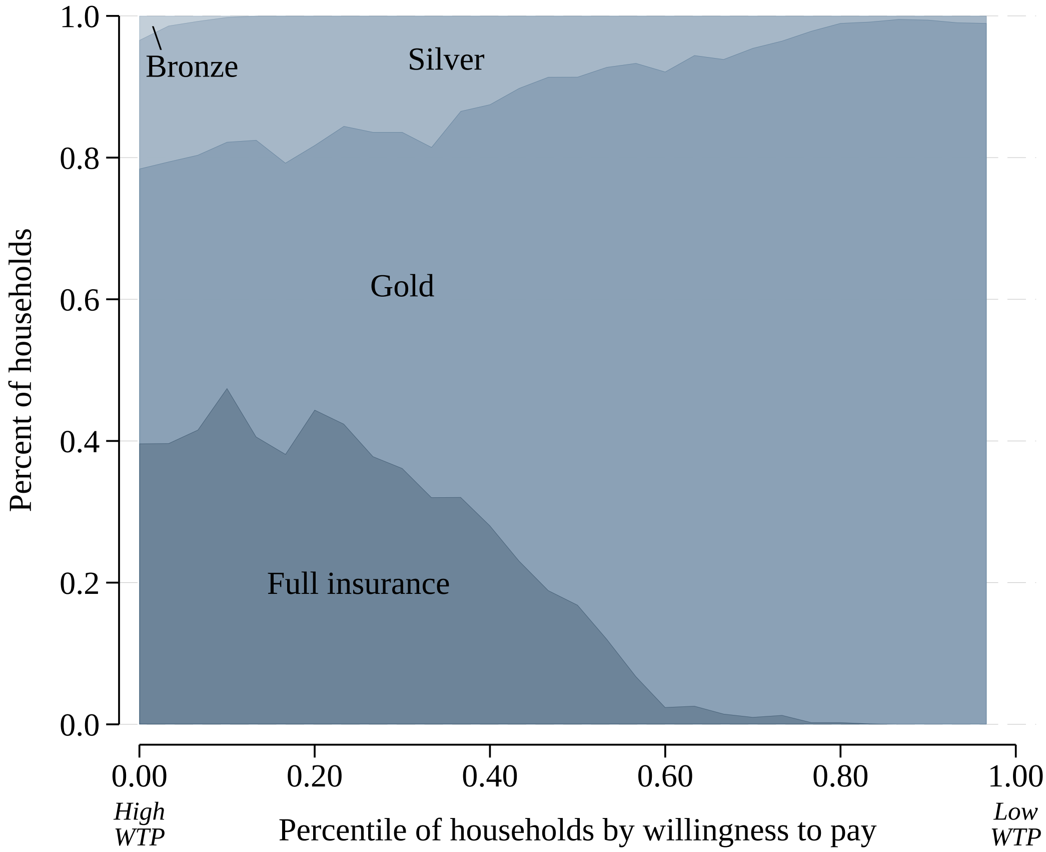
Notes: The figure shows the distribution across households of (a) the risk aversion parameter and (b) the moral hazard parameter. Each dot represents a household, for a 10 percent random sample of households. The dark line is a connected binned scatter plot for all family households; it represents the mean value of the vertical axis variable at each of the 100 percentiles of households by willingness to pay. The clumping at certain parameter values is driven by the intercepts (children versus no children), coupled with the normality assumption on unobserved heterogeneity.

Figure A.10. Household Health State Distributions by Willingness to Pay



Notes: The figure shows the distribution of health states faced by the set of households at each percentile of willingness to pay. Health state distributions are represented by their 10th, 25th, 50th, 75th, and 90th percentiles. A health state realization is equal to total healthcare spending absent moral hazard. The vertical axis is on a log scale in order to show more clearly the relationship between health state distributions and relevant values of the out-of-pocket cost function of the contracts we consider in Section V.B.

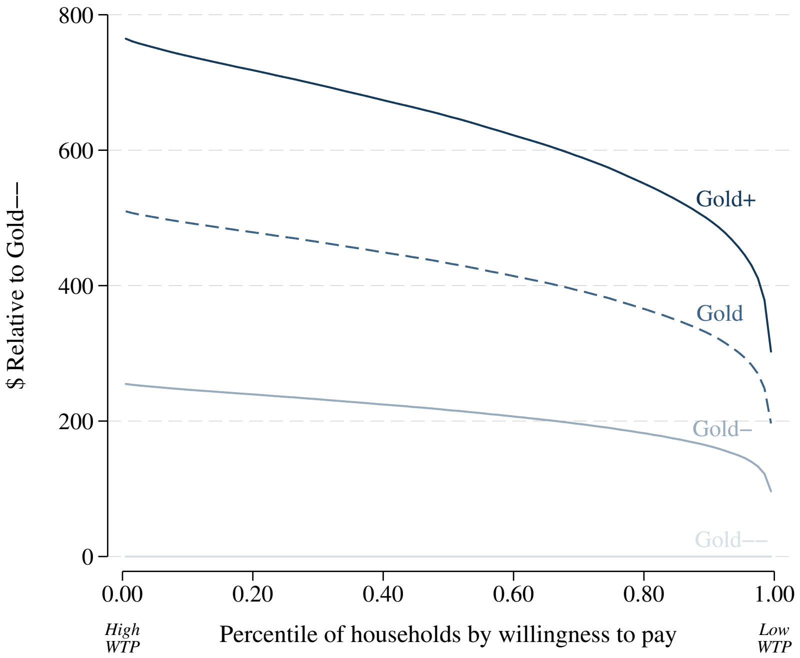
Figure A.11. Efficient Coverage Level by Willingness to Pay



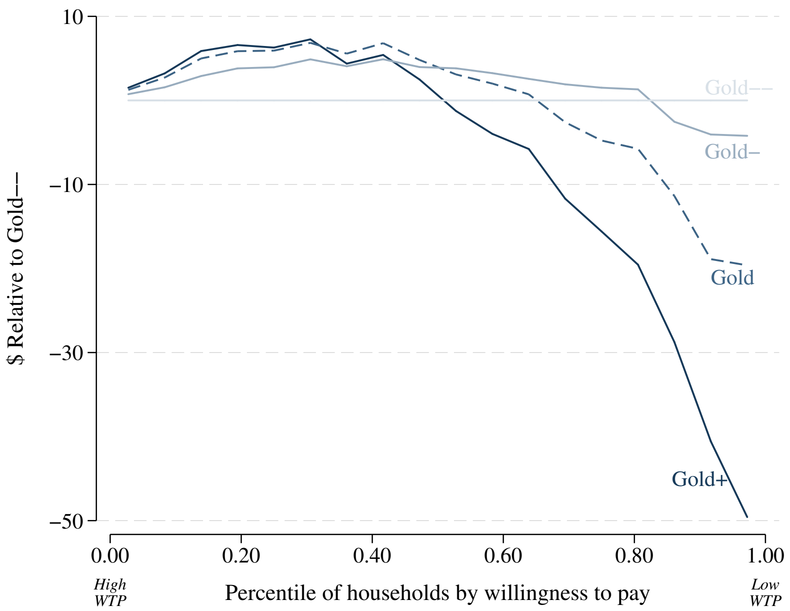
Notes: The figure shows the percentage of households at each percentile of willingness to pay for which each level of coverage is optimal. Households are ordered on the horizontal axis according to their willingness to pay. Overall, full insurance is efficient for 19.6 percent of households, Gold for 70.3 percent of households, Silver for 10.0 percent of households, and Bronze for 0.2 percent of households.

Figure A.12. WTP and Social Surplus in The Neighborhood of The Gold Contract

(a) Willingness to Pay



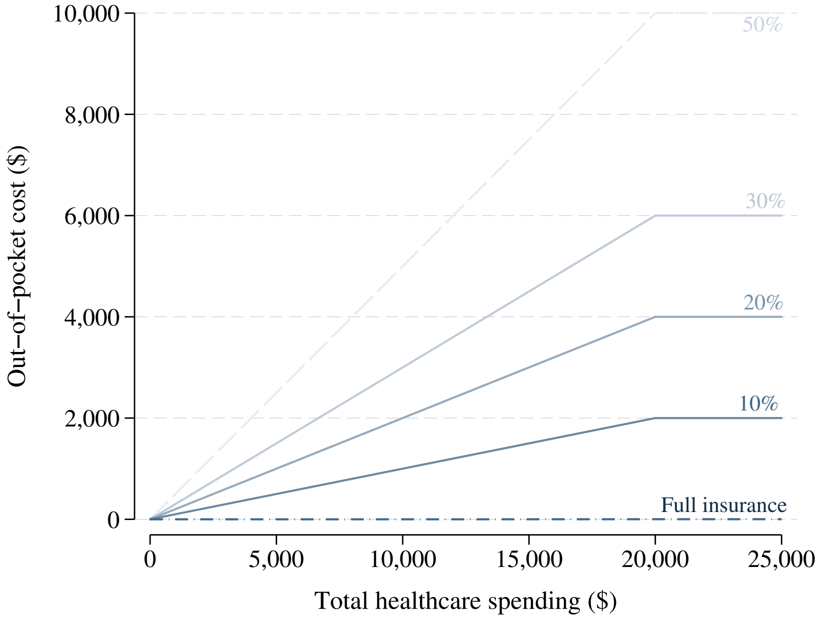
(b) Social Surplus



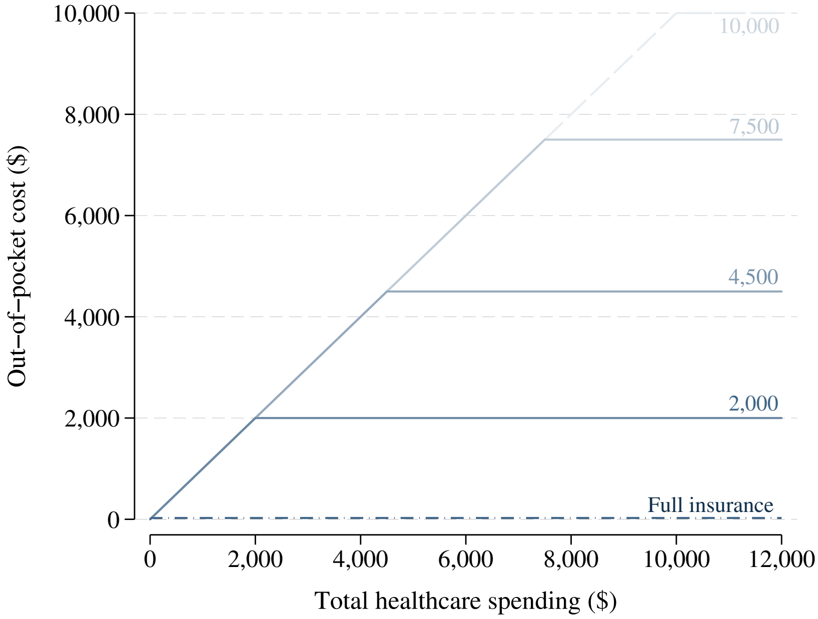
Notes: The figure shows the distribution across family households of (a) willingness to pay and (b) social surplus for the four contracts the regulator would want to offer (among the dense set of 40 contracts). Each panel consists of four connected binned scatter plots, with respect to bins of households ordered by willingness to pay for Gold+ relative to Gold−−. Panel (a) uses 100 bins, and panel (b) uses 20 bins (for readability). Both willingness to pay and social surplus are measured relative to the Gold−− contract.

Figure A.13. More Counterfactual Potential Contracts: Out-of-pocket Cost Functions

(a) Capped Linear Contracts



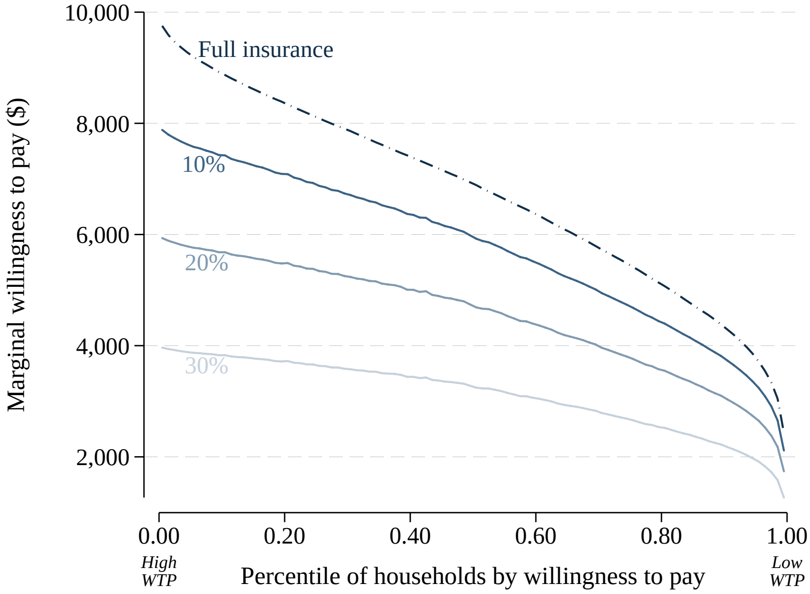
(b) Stop-loss Contracts



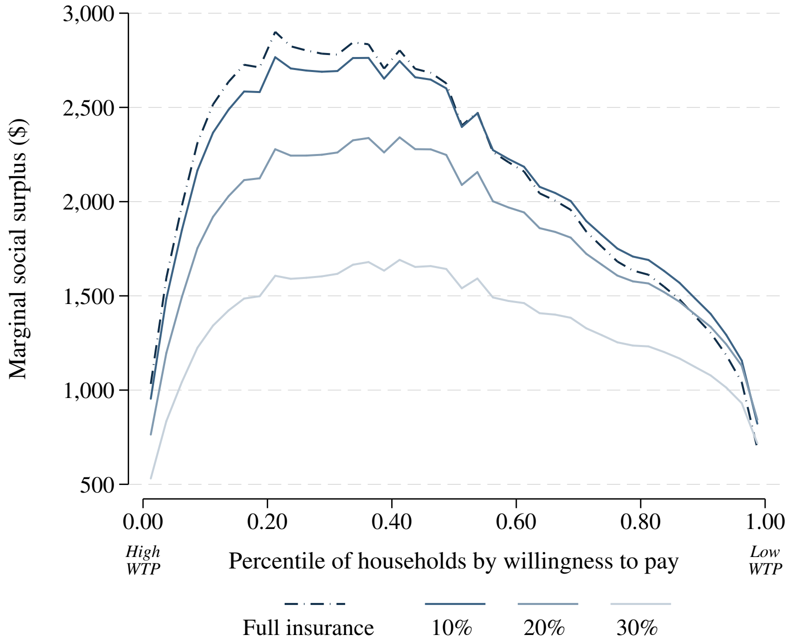
Notes: The figure shows the out-of-pocket cost functions for two additional sets of potential contracts we consider in Section C.1. Panel (a) shows capped linear contracts: The deductibles, coinsurance rates, and out-of-pocket maximums are \$0, 10%, \$2,000 for the "10%" contract; \$0, 20%, \$4,000 for the "20%" contract; \$0, 30%, \$6,000 for the "30%" contract; and \$0, 50%, \$10,000 for the "50%" contract. Panel (b) shows stop-loss contract that provide no coverage until the out-of-pocket maximum: The out-of-pocket maximums are \$2,000, \$4,500, \$7,500, and \$10,000.

Figure A.14. Willingness to pay and Social Surplus for Capped Linear Contracts

(a) Willingness to Pay



(b) Social Surplus

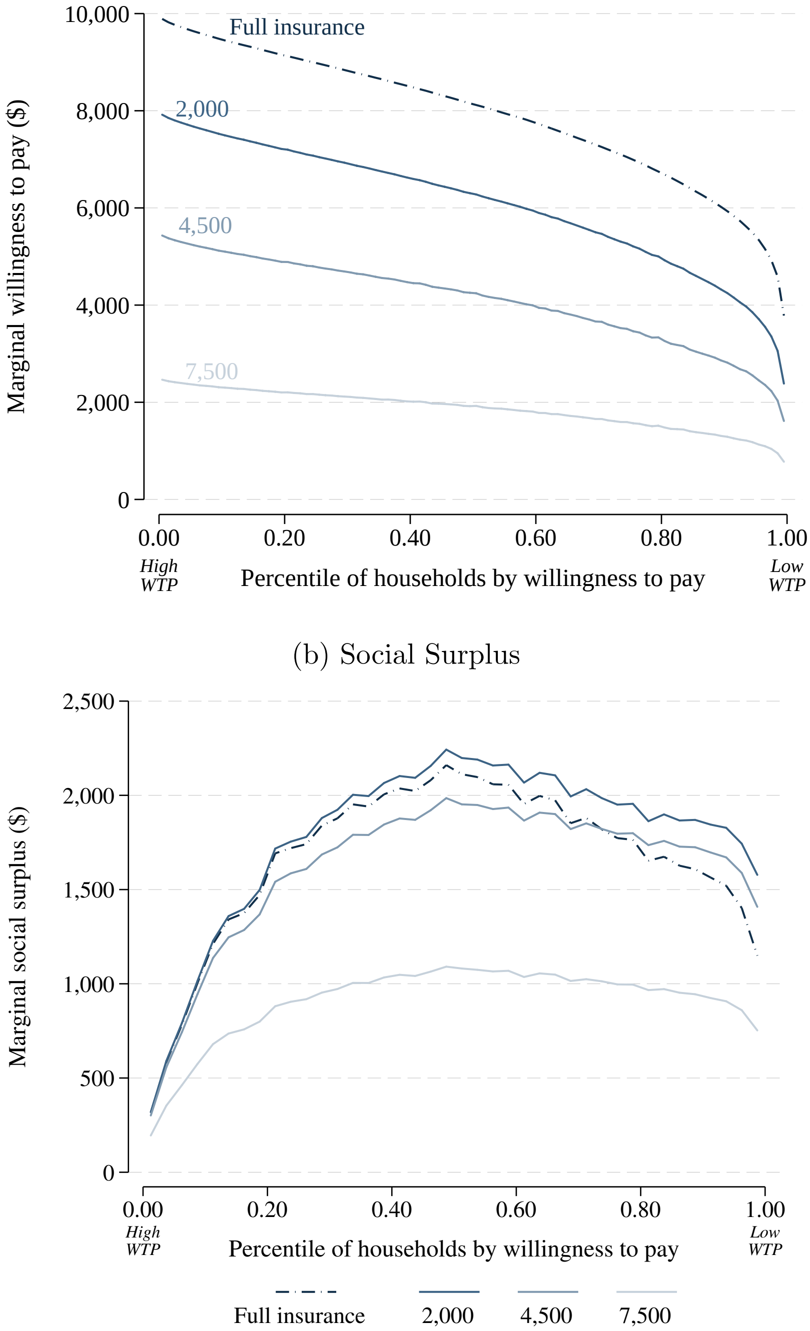


Notes: The figure shows the distribution across family households of (a) willingness to pay and (b) social surplus for the set of capped linear contracts considered in Section C.1 and depicted in Figure A.13a. Each panel consists of four connected binned scatter plots, with respect to bins of households ordered by willingness to pay for full insurance relative to the 50% contract. Panel (a) uses 100 bins, and panel (b) uses 40 bins (for readability). Both willingness to pay and social surplus are measured relative to the 50% contract.

Figure A.15. Willingness to pay and Social Surplus for Stop-loss Contracts

(a) Willingness to Pay

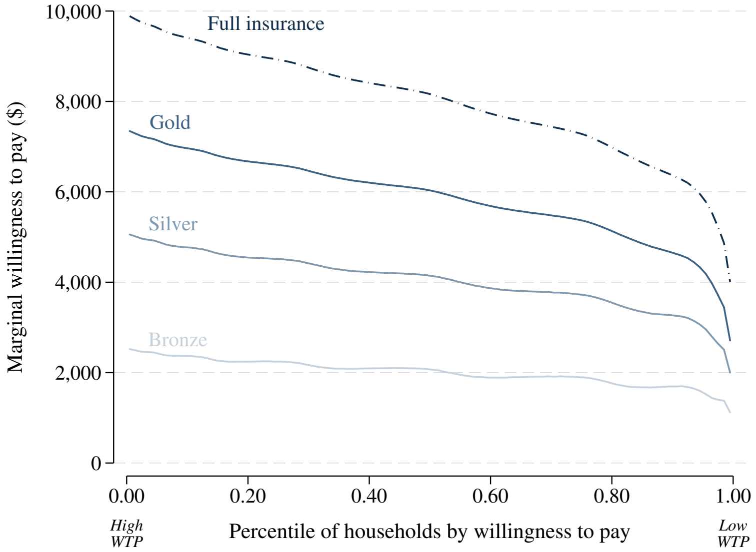
(b) Social Surplus



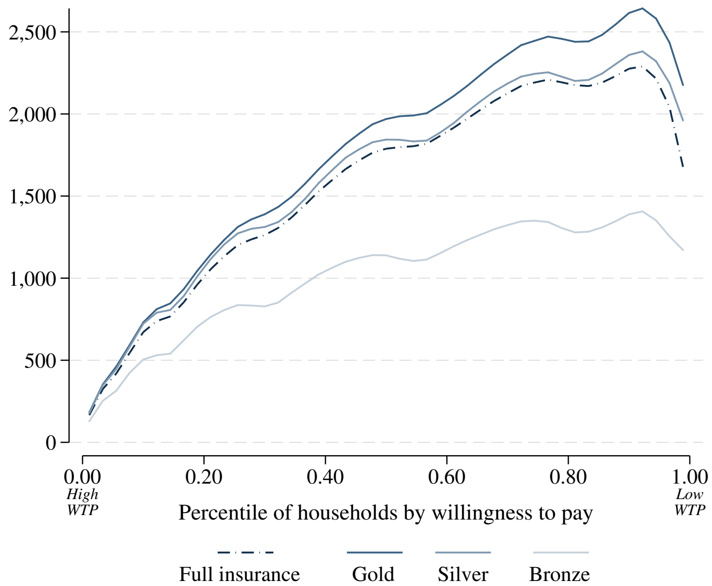
Notes: The figure shows the distribution across family households of (a) willingness to pay and (b) social surplus for the set of stop-loss contracts considered in Section C.1 and depicted in Figure A.13b. Each panel consists of four connected binned scatter plots, with respect to bins of households ordered by willingness to pay for full insurance relative to the $10,000 out-of-pocket maximum contract. Panel (a) uses 100 bins, and panel (b) uses 40 bins (for readability). Both willingness to pay and social surplus are measured relative to the $10,000 out-of-pocket maximum contract (i.e., the Catastrophic contract).

Figure A.16. Willingness to Pay and Social Surplus Under Fixed Preferences

(a) Willingness to Pay



(b) Social Surplus



Notes: The figure shows the distribution across family households of (a) willingness to pay and (b) social surplus under a counterfactual distribution of household types where preferences do not vary, as discussed in Appendix C.2. Risk aversion (ψ) is fixed at 0.599 and the moral hazard parameter (ω) is fixed at 1.751. Each panel consists of four connected binned scatter plots, with respect to bins of households ordered by willingness to pay for full insurance relative to the $10,000 out-of-pocket maximum contract. Panel (a) uses 100 bins, and panel (b) uses 40 bins (for readability). Both willingness to pay and social surplus are measured relative to the Catastrophic contract.