When Should There Be Vertical Choice in Health Insurance Markets?*

Victoria R. Marone§ Adrienne Sabety†

August 2021

Abstract

We study the welfare effects of offering choice over coverage levels—"vertical choice"—in regulated health insurance markets. We emphasize that heterogeneity in the efficient level of coverage is not sufficient to motivate choice. When premiums do not reflect individuals' costs, it may not be in consumers' best interest to select their efficient coverage level. We show that vertical choice is efficient only if consumers with higher willingness to pay for insurance have a higher efficient level of coverage. We investigate this condition empirically and find that as long as a minimum coverage level can be enforced, the welfare gains from vertical choice are either zero or economically small.

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

*We are grateful to Vivek Bhattacharya, David Cutler, Leemore Dafny, David Dranove, Amy Finkelstein, Tal Gross, Igal Hendel, Gaston Illanes, Matthew Leisten, Matt Notowidigdo, Chris Ody, Rob Porter, Elena Prager, Mar Reguant, Bill Rogerson, Mark Shepard, Amanda Starc, Bob Town, Tom Wiseman, and Gabriel Ziegler, as well as to the co-editor, Liran Einav, and three anonymous referees for advice and suggestions that greatly benefited this research. We also thank many participants in seminars at Northwestern, Kellogg, BI Norwegian Business School, University of Chicago, Princeton, MIT, Washington University in St. Louis, Yale SOM, Rochester Simon, NYU, MIT Sloan, Chicago Booth, Wisconsin, UT Austin, NBER Summer Institute, and the 2018 ASHEcon Conference 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, University of Texas at Austin. Email: marone@utexas.edu. †Department of Economics, University of Notre Dame. Email: asabety@nd.edu.

I Introduction

Choice over vertically differentiated financial coverage levels—which we term "vertical choice"—is widely available in U.S. health insurance markets.[1] A notable example is the tiered plans (e.g., Gold, Silver, Bronze) offered on Affordable Care Act exchanges. In contrast, national health insurance schemes typically offer only a single level of coverage. In both contexts, regulation plays a central role in determining the extent of vertical choice, but to date, the economics literature has provided limited guidance to regulators on this topic. This paper develops a theoretical and empirical framework for evaluating whether, and when, it is optimal to offer a vertical choice.

The basic argument in favor of vertical choice is the standard value 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, however, relies critically 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, like health insurance markets, this alignment may not be possible. In these markets, costs are inextricably related to private valuations, and asymmetric information (or regulation) prevents prices from reflecting marginal costs (Akerlof, 1970; Rothschild and Stiglitz, 1976). We show that even if health insurance markets are competitive, regulated, and populated by informed consumers, whether choice can increase welfare 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 how this central tradeoff between the "value of risk protection" and the "social cost of moral hazard" plays out on a consumer-by-consumer basis. The social aim is to design a plan menu that induces consumers to self-select into their efficient level of coverage. In doing so, the designer must contend with the fact that consumers with higher willingness to pay for insurance will select higher coverage. The key challenge is that consumers with higher willingness to pay are not necessarily the consumers with a higher efficient level of coverage. It is precisely this statement that captures the theoretical ambiguity of whether it is optimal to offer a vertical choice.

[1] We use the term *financial coverage level*, or just *coverage level*, to summarize the set of plan features, such as deductibles or out-of-pocket maximums, that determine insurer liability.

We consider the menu design problem facing a market regulator that can offer vertically differentiated plans and can set premiums.[2] The regulator's objective is to maximize allocative efficiency with respect to consumers and plans. As is standard in employer-sponsored insurance and national health insurance schemes, the regulator need not break even plan by plan, nor in aggregate. If more than one plan is demanded at the regulator's chosen premiums, 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 the optimal menu features vertical choice 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 (the expected reduction in out-of-pocket spending). Because a portion of a consumer's private valuation of insurance is a transfer, higher willingness to pay does not necessarily imply higher social surplus. For example, allocating higher coverage to a risk-neutral consumer delivers her a private benefit, but generates no social benefit; 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 over the period 2008 to 2013. Crucially for identification, we observe plausibly exogenous variation in the plan menus offered to employees. The variation is driven by the fact that plan menus are set independently by each of 187 school districts, which in turn select plans from a common superset determined at the state level. In addition, we observe several coverage levels offered by the same insurer with the same provider network, providing isolated variation along our focal dimension.

Our empirical model incorporates observed and unobserved heterogeneity across households

[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, it is the Centers for Medicare and Medicaid Services; under a national health insurance scheme, it is the government.

along three key dimensions: health status, propensity for moral hazard, and risk aversion. We use the model to recover the joint distribution of household types in our population. Given these estimates, 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. A household's efficient level of coverage is that which generates the highest social surplus.

Our estimates imply that all households have an efficient level of coverage in the range between a high-deductible contract (with a \$10,000 deductible and full coverage thereafter) and full insurance. The optimal menu will therefore only feature coverage levels in this range. Within it, we find that households with higher willingness to pay are primarily motivated by a greater expected reduction in out-of-pocket spending, rather than by a greater value of risk protection. Because these households are highly likely to spend past \$10,000, they face little out-of-pocket cost uncertainty under any relevant contract. This negative relationship between willingness to pay and "relevant risk" can be explained by the following pair of factors: (i) variation in willingness to pay is primarily driven by consumers' information about their health, and (ii) the lowest relevant level of coverage is reasonably high. The first factor implies that the highest willingness-to-pay consumers are the sickest, and the second implies that they would face little out-of-pocket cost uncertainty even in the lowest relevant coverage level. These relationships are important drivers of our results, and, in our view, are likely not unique to our setting.

We then solve for the optimal menu of contracts. Before doing so, a final design dimension is how "closely spaced" to permit contracts to be.[3] At baseline, we consider a fairly "sparse" potential contract space, where contracts may be no closer than \$2,500 out-of-pocket maximum intervals. In this case, we find that the optimal menu consists of a single contract, which has a deductible of \$1,125 and out-of-pocket maximum of \$2,500.[4] Introducing any other contract, at any price, leads to over- or under-insurance (on average) among households that would choose the alternative. We then increase the density of allowable contracts by a factor of 10 (to \$250 out-of-pocket maximum intervals). Here, we find that it is efficient to offer a vertical choice: the optimal menu features four contracts, clustered around the original optimal contract. However, because social surplus is quite flat across coverage levels near the optimum, the welfare gains

[3]We make minimum contract differentiation an explicit design dimension in order to sidestep the need to quantify relevant primitives, such as a fixed cost of offering contracts.

[4]Allocating all households to this contract raises welfare by \$1,514 per household per year relative to allocating all households to the high-deductible contract, and by \$104 relative to allocating all households to full insurance.

are economically small. Offering choice increases welfare by only \$5 per household per year relative to what is achieved by a single contract. Such small gains may quickly be outweighed by factors we do not model, such as administrative costs of offering each contract, or of offering a choice at all.

We investigate the robustness of our findings with respect to both the consumers and the types of contracts under consideration. Beginning with consumers, we evaluate various perturbations of the parameter estimates that define our distribution of household types. We find that in a broad neighborhood of our estimates, the welfare gains from vertical choice are either zero or economically small (at most \$16 per household per year). Asking consumers to make a choice consistently reveals more about their health information than it reveals about their preferences. That said, choice becomes efficient when risk protection is responsible for a larger part of the variation in marginal willingness to pay for higher coverage (e.g., when doubling average risk aversion). We then consider alternative contract designs, including removing the deductible, removing the coinsurance region, and extending the coinsurance region. We consistently find that a relatively low out-of-pocket maximum is efficient for all consumers. And as before, the minimal uncertainty facing the sickest households limits the extent to which it is efficient to allocate high willingness-to-pay households higher coverage. Overall, our results suggest that if it is possible to impose a modest minimum coverage level in a market, offering the choice of a higher coverage option is unlikely to deliver meaningful welfare gains.

Finally, we compare welfare outcomes and distributional implications under various pricing policies, including competitive pricing and full vertical choice. Under competitive pricing, all contracts must break even, and we find that the market unravels 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. Relative to this benchmark, the optimal menu (the single contract) increases welfare by \$315 per household per year. These gains are not shared evenly in the population: sicker and larger households fare best under the single contract, while healthier and smaller households fare best under full vertical choice. Our results suggest that one reason for the persistence of vertical choice in settings such as employer-sponsored insurance could be to limit redistribution across these groups.

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.[5] Our paper focuses directly on why this is the case, and brings to bear an empirical approach that permits substantially more flexibility in the distribution of consumer types.

Our paper also closely relates to work that evaluates 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), and more specifically to the growing literature on menu design in these markets.[6] In the context of insurer choice, Bundorf, Levin and Mahoney (2012) investigate the optimal allocation of consumers to insurers, and find that it cannot be achieved by uniform pricing. Our paper is similar in spirit (and in findings), but focuses instead on the financial dimension of insurance. In this context, Ericson and Sydnor (2017) also consider the question of whether choice is welfare-improving. A key difference of our work is that we consider a setting in which contract characteristics are endogenous and premiums are exogenous, as opposed to the reverse. In similar and concurrent work, Ho and Lee (2021) study optimal menu design from the perspective of an employer. Like us, they find that the gains from offering a choice over coverage levels are small. Our contribution relative to these papers is to provide a conceptual characterization of when choice over financial coverage levels is and is not valuable. We view this characterization as a tool that can be directly used to reexamine existing policies through a new lens.[7] Our empirical analysis demonstrates the relevance of the prediction that vertical choice may not be valuable, and links it to the distribution of fundamentals—risk aversion, propensity for moral hazard, and distributions of health outcomes—in a population.

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 et al., 2019; Bundorf et al., 2019; Samek and Sydnor, 2020). Importantly, if privately and socially optimal allocations do not align, more

---

[5]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).

[6]We also note the close relationship between our paper and recent work by Landais et al. (2021)on unemployment insurance and Hendren, Landais and Spinnewijn (2021) on social insurance more broadly. Like us, these papers consider the value of offering a choice from the perspective of a social planner that can set prices.

[7]For example, vertical choice is currently the status quo in the Affordable Care Act exchanges, in Medicare (through the availability of Medigap policies), and in some national health insurance systems (for example, Switzerland and the Netherlands).

diligent consumers may just as well lead to less desirable outcomes (as is found by Handel, 2013). A central aim of the present paper is to inform the design of health insurance markets in such a way that better-informed consumers always lead to better allocations.

The paper proceeds as follows. Section 2 presents our theoretical model and derives the objects relevant to describe private and social incentives. 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 main results. Section 6 evaluates welfare and distributional outcomes. Section 7 concludes.

II Theoretical Framework

II.A Model

We consider a model of a health insurance market in which consumers are heterogeneous along multiple dimensions and the set of traded contracts is endogenous. We assume that premiums may not vary with consumer characteristics, claims may be contingent only on healthcare utilization, and 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 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 governs consumer preferences over healthcare utilization (and ultimately captures the degree of moral hazard). A population is defined 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. Given their health state, consumers decide the money amount $m \in \mathbb{R}_+$ of healthcare utilization ("spending") to consume, a decision which in part depends on their insurance contract. Contracts are characterized by an increasing and concave out-of-pocket cost schedule $c_x : \mathbb{R}_+ \to \mathbb{R}_+$, where $c_x(m) \leq m \; \forall \; m$.

Consumers value healthcare spending m and residual income y. Preferences are represented

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

by $u_\psi(y + b(m; l, \omega))$, where b is a money-metric valuation of healthcare utilization, and u_ψ and $b(\cdot; l, \omega)$ are each strictly increasing and concave. Upon realizing their health state, consumers choose their healthcare utilization by trading off its benefit with its out-of-pocket cost: $m^*(l, \omega, x) = \text{argmax}_m \, (b(m; l, \omega) - c_x(m))$. Privately optimal utilization implies indirect benefit $b^*(l, \omega, x) = b(m^*(l, \omega, x); l, \omega)$ and indirect out-of-pocket cost $c_x^*(l, \omega, x) = c_x(m^*(l, \omega, x))$. Before the health state is realized, expected utility is given by

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

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 utilization, m. Socially optimal healthcare utilization therefore coincides with privately optimal utilization absent insurance.[9] The difference between privately optimal spending $m^*(l, \omega, x)$ and socially optimal spending $m^*(l, \omega, x_0)$ is central to calculating the social cost of insurance. 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."[10] A consumer's net payoff from moral hazard spending is given by

$$v(l, \omega, x) = \underbrace{b^*(l, \omega, x) - b^*(l, \omega, x_0)}_{\substack{\textit{Benefit of moral} \\ \textit{hazard spending}}} - \underbrace{\big(c_x^*(l, \omega, x) - c_x^*(l, \omega, x_0)\big)}_{\substack{\textit{Out-of-pocket cost of} \\ \textit{moral hazard spending}}},$$

where $b^*(l, \omega, x_0)$ is the indirect benefit of uninsured behavior, and $c_x^*(l, \omega, x_0)$ is the out-of-pocket cost of uninsured behavior at insured prices. Note that since any change in behavior is voluntary, $v(l, \omega, x)$ is weakly positive.

Calculations in Appendix A.1 show that if u_ψ features constant absolute risk aversion, will-

[9] Importantly, this is true only if m represents the true cost of healthcare provision and there are not externalities associated with healthcare utilization, as we assume here.

[10] Following convention, we use the term "moral hazard" to describe the scenario at hand, in which there is elastic demand for the insured good and a state that 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 the dialogue between Pauly (1968) and Arrow (1968).

ingness to pay for contract x relative to the null contract x_0 can be expressed as[11]

$$WTP(x,\theta) = \underbrace{\mathbb{E}_l[c^*_{x_0}(l,\omega,x_0) - c^*_x(l,\omega,x_0)]}_{\substack{\text{Expected reduction in out-of-pocket} \\ \text{cost holding behavior fixed}}} + \underbrace{\mathbb{E}_l[v(l,\omega,x)]}_{\substack{\text{Expected payoff from} \\ \text{moral hazard spending}}} + \underbrace{\Psi(x,\theta)}_{\substack{\text{Value of risk} \\ \text{protection}}} . \quad (2)$$

Willingness to pay is composed of three terms: the expected reduction in out-of-pocket cost holding behavior fixed (at uninsured behavior), the expected payoff from moral hazard spending, and the value of risk protection.[12] The first term captures the transfer from the consumer to the insurer of the expected healthcare spending liability that exists even absent moral hazard. It represents an equal and opposite cost to the insurer. The second and third terms, in contrast, are what are relevant to social welfare. Consumers partially value the additional healthcare they consume when they have higher 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 $k_x(m)$, where $m = k_x(m) + c_x(m)$. A reduction in out-of-pocket cost is an increase in insurer cost, so $c^*_{x_0}(l,\omega,x_0) - c^*_x(l,\omega,x_0) = k^*_x(l,\omega,x_0)$.[13] The social surplus generated by allocating a consumer of type θ to contract x (relative to allocating the same consumer to the null contract) is the difference between $WTP(x,\theta)$ and expected insured cost $\mathbb{E}_l[k^*_x(l,\omega,x)]$, which after simplying is:

$$SS(x,\theta) = \underbrace{\Psi(x,\theta)}_{\substack{\text{Value of risk} \\ \text{protection}}} - \underbrace{\mathbb{E}_l[\, k^*_x(l,\omega,x) - k^*_x(l,\omega,x_0) - v(l,\omega,x)\,]}_{\substack{\text{Social cost} \\ \text{of moral hazard}}} . \quad (3)$$

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. The welfare loss equals the portion of the expected increase in healthcare spending that is not valued.[14,15]

[11] The single role of constant absolute risk aversion is to ensure that the value of risk protection, and thereby social surplus, is invariant to the contract premium.

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

[13] To see this, note that $c^*_{x_0}(l,\omega,x_0) = m^*(l,\omega,x_0)$ and $c^*_x(l,\omega,x_0) = c_x(m^*(l,\omega,x_0))$.

[14] The social cost of moral hazard can also be expressed as $\mathbb{E}_l[m^*(l,\omega,x) - m^*(l,\omega,x_0) - (b^*(l,\omega,x) - b^*(l,\omega,x_0))]$.

[15] Note that $v(l,x,\omega)$ must be weakly lower than the insured cost of moral hazard spending $k^*_x(l,\omega,x) - k^*_x(l,\omega,x_0)$, or else that level of spending would have been chosen even absent insurance. The social cost of

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

Supply and Regulation. We suppose contracts are supplied by a regulator, which can observe the distribution of consumer types and can set premiums on all contracts except x_0, which has zero premium. The regulator need not break even on any given contract, nor in aggregate. It can effectively remove any non-null contract from the set of contracts on offer by setting a premium of infinity. It can effectively remove x_0 from offer by setting the premium of any non-null contract to zero. This simple model of supply is isomorphic to a more complicated model involving perfect competition among private insurers and a regulator that can strategically tax or subsidize contracts. Precisely such a model is formalized in Section 6 of Azevedo and Gottlieb (2017).

The regulator sets premiums $\boldsymbol{p}$ in order to maximize social welfare, 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. That is, we ask whether the optimal feasible allocation features enrollment in more than one contract.[16]

II.B Graphical Analysis

We characterize the answer graphically for the case of a market with only 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, risk aversion, and consumer heterogeneity are necessary conditions for vertical choice to be efficient. If there were not moral hazard, the highest coverage contract would be socially optimal for all consumers, and the optimal menu would involve only this contract. Absent risk aversion, the same would be true with the lowest coverage contract. If there were not consumer heterogeneity, all consumers would again have the same socially optimal contract, and the optimal menu would again feature only a single

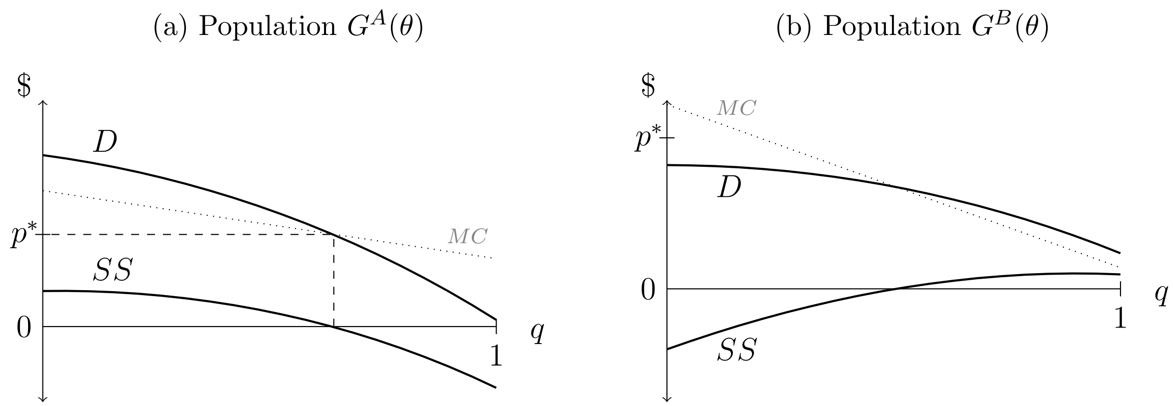
moral hazard is therefore weakly positive and at most the expected insured cost of moral hazard spending.

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

contract. In the following, we explore the more interesting (and more realistic) cases in which consumers do not all have the same socially optimal contract.

Two Contract Example. Suppose there are two potential contracts, x_H and x_L, where x_H provides higher coverage than x_L.[17] Figure 1 depicts the market for x_H in two populations. If a consumer does not choose x_H, they receive x_L; x_0 is excluded by setting p_L to zero. As x_H provides higher coverage, $WTP(x_H, \theta) \geq WTP(x_L, \theta)$ 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 premium $p = p_H - p_L$ at which the contracts are offered. The horizontal axis plots the fraction q of consumers that choose x_H.

Figure 1. Examples in which Vertical Choice (a) *Is* and (b) *Is Not* Efficient



Notes: The figure shows two health insurance markets in which there are two contracts available: x_H and x_L, where x_H provides higher coverage than x_L. Each panel shows the demand curve D, the marginal cost curve MC, and the social surplus curve SS for x_H relative to 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.

Each panel also shows the marginal cost curve MC and the marginal social surplus curve SS. The marginal cost curve measures the expected cost of insuring consumers under x_H relative to x_L: $\mathbb{E}_l[k^*_{x_H}(l, \omega, x_H) - k^*_{x_L}(l, \omega, x_L)]$. 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 level of marginal willingness to pay). The social surplus curve SS plots the vertical difference between D and MC, or equivalently, the average value of $SS(x_H, \theta) - SS(x_L, \theta)$ among all consumers at a particular point on the horizontal axis.

Though vertical differentiation implies D and MC must be weakly positive, the presence of moral hazard means that SS need not be. It is possible for consumers to be over-insured.

[17]See Appendix A.2 for a formal definition of coverage level ordering.

Moreover, our precondition that all consumers do not have the same socially optimal contract guarantees that in both populations, marginal social surplus will be positive for some consumers and negative for others.[18] 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 population $G^A(\theta)$, marginal social surplus is increasing in marginal willingness to pay. The optimal marginal premium p^* can therefore sort consumers with on-average positive SS into x_H, and on-average negative SS into x_L. In population $G^B(\theta)$, meanwhile, such a premium does not exist.

In population $G^B(\theta)$, any interior allocation results in some amount of "backward sorting," meaning that there is a group of consumers enrolled in x_H who would be more efficiently enrolled in x_L, and vice versa. Consequently, any allocation with enrollment in both contracts is dominated by an allocation with enrollment in only one. No sorting dominates backward sorting because it is always possible to prevent "one side" of the backward sort.[19] In the example shown, the integral of SS is negative, meaning that the population would on average be over-insured in x_H. p^* is therefore anything high enough to induce all consumers to choose x_L.

Remarks. The limitation of choice as a screening mechanism is directly related to the idea that a single (community-rated) price may not be able to efficiently sort consumers that vary in cost (Einav, Finkelstein and Levin, 2010; Glazer and McGuire, 2011; Bundorf, Levin and Mahoney, 2012; Geruso, 2017). Consumers select a contract based on the available consumer surplus, $CS = WTP - p$, while efficiency relies on a comparison with cost, $SS = WTP - MC$. When CS and SS diverge (when $p \neq MC$), the efficiency of choice turns on whether they are at least positively related. If they are not, choice can only result in some degree of "backward sorting."[20]

In the simple case of two contracts and a social surplus curve that crosses zero at most once, vertical choice is efficient if and only if it crosses from above. In a more general case with

[18]Note that since SS represents an average, this condition does not itself guarantee that the social surplus curve will cross zero. Since it is necessary for SS to cross zero for vertical choice to be optimal, we focus our two examples on cases in which that occurs. If SS did not cross zero, a single plan would be on-average optimal at every level of willingness to pay, and the optimal menu would feature a single contract.

[19]To see this, consider the (worst possible) allocation $\tilde{q}$ at the point where SS intersects zero. A slightly higher allocation $\tilde{q}'$ strictly dominates, as more consumers with positive marginal social surplus now enroll in 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.

[20]Though not by this name, the idea of backward sorting has appeared previously in related literature, in the setting of insurer choice (Bundorf, Levin and Mahoney, 2012) and provider network choice (Shepard, 2016).

multiple potential contracts and arbitrary social surplus curves, this necessary and sufficient condition is still directly informative. If consumers all have the same socially optimal contract (or more plausibly, if the same contract is socially optimal at all levels of willingness to pay), there will be no crossing in the upper envelope of social surplus curves, and the optimal menu will feature this single contract. If instead there *is* crossing in the upper envelope of social surplus curves, one must assess whether the higher-coverage contracts cross from above, or in other words, whether or not choice would lead to backward sorting.

Taken together, the procedure for evaluating the efficiency of vertical choice can be summarized by a test for the condition of whether consumers with higher willingness to pay have a higher efficient coverage level, where we emphasize that *higher*, in both instances, is to be evaluated strictly. This condition itself is complex. It is both theoretically ambiguous and, by our own assessment, not obvious. 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 determine 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 around the world and present in the empirical setting we study. Ultimately, whether consumers with higher private valuations of higher coverage also generate a higher social value from higher coverage is an open empirical question.

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.[21]

The data contain employees' plan menus, 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 a 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 variation in both prices and choice sets. While all plan menus we observe are quite generous, in that the plans are generally high-coverage and 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 vary 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

[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).

[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.

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.

Plan Characteristics. During our sample period, OEBB contracted with three insurers: Kaiser, Moda, and Providence. 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. Each insurer used a single provider network and offered multiple plans. Within insurer, plans were differentiated only by financial coverage level.

Table 1 summarizes the state-level master list of plans made available by OEBB in 2009. The average employee premium represents the average annual premium employees would have had to pay for each plan. The full premium reflects the per-employee premium paid to the insurer.[26] The difference 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.[27]

As a way to summarize and compare plan coverage levels, we construct each plan's actuarial value. This measure reflects the share of total population spending that would be insured under a given plan.[28] Full insurance would have an actuarial value of one; less generous plans have lower actuarial values. In later years, the distribution of coverage levels looks qualitatively similar, with the notable exception that Providence was no longer available in 2012 and 2013.[29]

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 our sample begins in 2009. These restrictions leave us with 44,562 households, representing 117,934 individuals.[30]

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

[27]Many other cost-sharing details determine plan coverage level. For the purposes of our empirical model, we estimate the 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 B.1.

[28]We evaluate out-of-pocket spending for each household in each plan, and then divide average insured spending by average total spending across all households for each plan. We evaluate counterfactual out-of-pocket spending using the "claims calculator" developed for this setting by Abaluck and Gruber (2016).

[29]Corresponding information for the plans offered between 2010 and 2013 are provided in Table A.2.

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

Table 1. 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	<0.01
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: The table shows the state-level master list of plans available in 2009. Actuarial value is the ratio of the sum of insured spending across all households to the sum of total spending across all households. 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 out-of-pocket maximum shown are for in-network services for a family household.

There is a clear bifurcation of our sample between Kaiser and non-Kaiser households. That is, 78 percent of households always chose either Moda or Providence, 19 percent always chose Kaiser, and only 3 percent at some point switched between. This pattern is not necessarily surprising. Kaiser offers a substantially different type of insurance product, and persistent consumer preference heterogeneity along this dimension would be a reasonable expectation. That said, modeling the choice over insurer type somewhat distracts from our focus on choice over financial coverage level. We therefore take advantage of this division in the data and conduct our primary analysis on the set of households that never enrolled with Kaiser. We consider the full sample in a robustness analysis in Section V.C.[31]

Table 2 provides summary statistics on our panel of households. The first column describes the full sample, while the second column describes the subset of households that never enrolled in a Kaiser plan. Focusing on the non-Kaiser sample, 49 percent of households have children, and 74 percent of households are "families" (anything other than the employee alone). The average employee is age 47.9, and the average enrollee (employees and their covered dependents) is age 40.4. Households on average have 2.6 enrollees.

Employees received large subsidies toward the purchase of health insurance. The average

[31] As discussed in that section, using the full sample leaves our results are qualitatively unchanged.

Table 2. Household Summary Statistics

Sample demographics	Full Sample	Excluding Kaiser
Number of households	44,562	34,606
Number of enrollees	117,934	92,244
Pct. of households with children	0.49	0.49
Enrollees per household, mean (med.)	2.57 (2)	2.60 (2)
Enrollee age, mean (med.)	39.8 (37.8)	40.4 (38.7)
Premiums		
Employee premium ($), mean (med.)	880 (0)	843 (0)
Full premium ($), mean (med.)	11,500 (11,801)	11,582 (11,801)
Household healthcare spending		
Total spending ($), mean (med.)	10,754 (4,620)	11,689 (5,173)
Out-of-pocket ($), mean (med.)	1,694 (1,093)	2,054 (1,540)
Switching (pct. of household-years)		
Forced to switch plan	0.20	0.21
insurer	0.01	0.02
Unforced, switched plan	0.17	0.20
insurer	0.03	0.03

Notes: Enrollees are employees plus their covered dependents. Sample statistics are calculated across all years, 2009–2013. Premiums statistics are for households' chosen plans, as opposed to for all possible plans. Sample medians are shown in parentheses.

household paid only $843 per year for their chosen plan; the median household paid nothing. Meanwhile, the average full premium paid to insurers was $11,582, meaning that the average household received an employer contribution of $10,739. Households had average out-of-pocket spending of $2,054 and average total healthcare spending of $11,689.

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 21 percent of household-years to switch plans, and 2 percent to switch insurers. When the prior choice *was* available, 20 percent of household-years voluntarily switched plans and only 3 percent voluntarily switched insurers. The presence of both forced and unforced switching is important in our empirical model for identifying the extent of "inertia" in households' choice of plan and insurer.

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 55 percent of households); the Eugene and Medford HRRs in southwest Oregon (32 percent of households); and the Bend, Spokane, and

Boise HRRs in eastern Oregon (13 percent of households).[32]

III.B Variation in Plan Menus

For the purposes of the present 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 that 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 actuarial value that an average household would be most likely to choose if offered a given plan menu, against the average spending of the households presented with that menu. This measure of plan menu generosity captures both the facts that a level of coverage can only be chosen if it is offered, and is more likely to be chosen if it is cheaper.[33] Each point on the plot represents the set of plan menus that share the same predicted actuarial value. Points are then weighted by the number of households represented.[34] The resulting pattern indicates that households that were offered a more generous plan menu had higher spending. The patterns in both panels persist when we control for observables, suggesting the presence of adverse selection on unobservables, and of moral hazard.

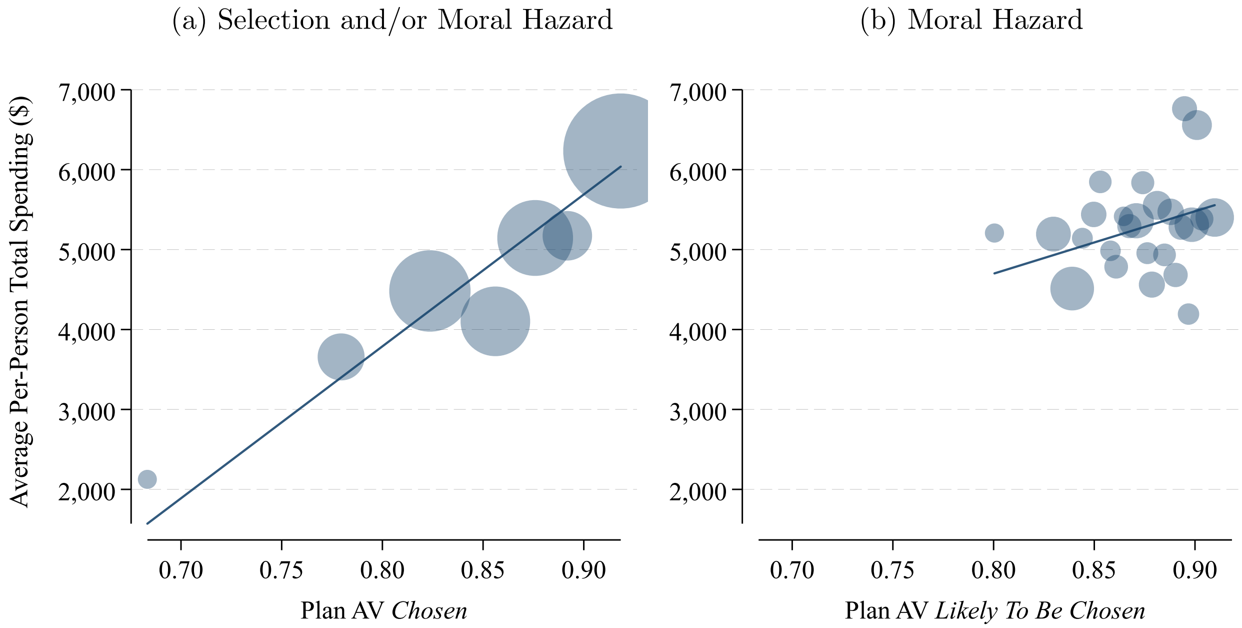
Identification of our structural model proceeds in much the same way as the above arguments. A key identifying assumption is that plan menus are independent of household unobservables, conditional on household observables. An important threat to identification is that school districts chose plan menu generosity in response to unobservable information about employees that would also drive healthcare spending. To the extent that districts with unobservably

[32] For more information and HRR maps, see http://www.dartmouthatlas.org/data/region.

[33] We construct this measure using a conditional logit model of household plan choice. This model and the resulting measure of plan menu generosity are described in detail in Appendix B.2.

[34] To improve readability, very close values of predicted actuarial value are bucketed together.

Figure 2. Average Spending by Plan Coverage Level



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, households are grouped by their chosen plan, and each dot represents one of the seven Moda plans. In the right panel, households are grouped by the plan menu they were offered, and each dot represents a set of plan menus with the same predicted actuarial value. Predicted actuarial value is the AV most likely to be chosen if an average household was presented with that plan menu. The size of each dot indicates the number of households represented. Lines of best fit are weighted accordingly.

sicker households provided more generous health benefits, this would lead us to overstate the extent of moral hazard.[35]

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 did school districts 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 B.2 describes these results in greater detail. It also presents additional tests for what *does* explain variation in plan menus. We find that, among other

[35]The relationship could also run the other way: households could move across school districts, or select a district initially, based on the available health benefits. Such selection could again result in unobservably sicker households obtaining more generous health benefits. To the extent that observable health factors are correlated with unobservable factors that would drive this relationship, the analysis that follows is also relevant to this concern.

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 within our structural model, but can also use it in a more isolated way to produce reduced-form estimates of moral hazard. Appendix B.3 presents an instrumental variables analysis using two-stage least squares. The estimates yield a moral hazard "elasticity" that can be directly compared with others in the literature. We estimate that the elasticity of demand for healthcare spending with respect to its average end-of-year out-of-pocket cost 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 structural model and of our research question.

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, a dollar in premiums may be valued differently than a dollar in out-of-pocket spending, and consumers make repeated plan choices over time.

Household Utility. Following Cardon and Hendel (2001) and Einav et al. (2013), we parameterize the value of healthcare spending to be quadratic in its distance 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 spending and, ultimately, the degree to which optimal spending varies across coverage levels. Given out-of-pocket cost function $c_{jt}(m)$ for plan

j in year t, privately optimal healthcare spending is $m^*_{jt}(l, \omega_k) = \text{argmax}_m \left(b(m; l, \omega_k) - c_{jt}(m)\right)$.[36]

This parameterization is attractive because it produces reasonable predicted behavior under nonlinear insurance contracts, and it is tractable enough to be used inside an optimization routine.[37] Additionally, ω_k can be usefully interpreted as the incremental spending induced by moving a household from no insurance to full insurance. 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 in payoffs only through uncertainty in $b^*_{jt}(l, \omega_k) - c^*_{jt}(l, \omega_k)$.

Household k in year t derives the following expected utility from plan choice j:

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

where ψ_k is a coefficient of absolute risk aversion, z_{kjt} is the payoff associated with realization of health state l, and F_{kft} is the distribution of health states faced if the plan belongs to insurer $f(j)$. The payoff associated with health state realization l is given by

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

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.[38] The payoff z_{kjt} is measured in units of premium dollars. Out-of-pocket costs may be valued differently from premiums through parameter α^{OOP}. Finally, ϵ_{kjt} represents a household-plan-year idiosyncratic shock, with magnitude σ_ϵ to be estimated. We assume these shocks are independently and identically distributed Type 1 Extreme Value, and that households chose the plan that maximized expected utility from among the set of plans $\mathcal{J}_{kt}$ available to them:

[36] Note that c_{jt} is indexed by t because cost-sharing parameters vary within a plan across years. It also varies by household type (individual versus family), but we omit an additional index to save on notation. With a linear out-of-pocket cost function with coinsurance rate c and nonnegative health states: $m^* = \omega(1-c) + l$ and $b^* = \frac{\omega}{2}(1-c^2)$. Appendix C.2 provides solutions when contracts are piecewise linear and negative health states are permitted.

[37] The model predicts, for example, that if a consumer realizes a health state just under the deductible, she will take advantage of the proximity to cheaper healthcare and consume a bit more (putting her into the coinsurance region). Figure A.2 provides a depiction of optimal spending behavior predicted by this model.

[38] $\mathbf{X}_{kjt}$ 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. The associated parameters for health account and vision/dental contributions are α^{HA} and α^{VD}, respectively.

$j^*_{kt} = \text{argmax}_{j \in \mathcal{J}_{kt}} U_{kjt}$.

Distribution of Health States. We assume that individuals face a lognormal distribution of health states, and that households face the sum of health state draws across all individuals 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. 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. 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. Our method is to estimate individuals' health state distributions, allowing parameters to vary with individual-level demographics. Additional details can be found in Appendix C.1.

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 a mass of individuals with zero spending. If κ_{it} is positive, negative health states are permitted, which may imply zero spending.[39] Parameters μ_{it}, σ_{it}, and κ_{it} are in turn projected on individual demographics (such as 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} are a function of 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 individuals and each individual's demographics. In addition to this observable heterogeneity, we incorporate unobserved heterogeneity in household health though parameter μ_{kt}. Households can in this way hold private information about their health that can drive both plan choices and spending outcomes.

Finally, 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 insur-

[39] A negative health state implies zero spending as long as ω_k is not so large that $m^*_{jt}(l, \omega_k)$ becomes positive.

ers, conditional on health state.[40] For example, the same physician office visit might lead to different amounts of total spending across insurers simply because each insurer paid the physician a different price. We do not want such variation to be attributed to differences in underlying health. Our approach is to estimate insurer-level parameters that multiply realized health states, transforming them from underlying "quantities" of healthcare utilization into the monetary spending amounts we observe in the claims data. We model a household's money-metric health state l as the product of an insurer-level "price" 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

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

IV.B Identification

Our aim is 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 coverage. Variation in these objects arises from variation in either household preferences (the risk-aversion and moral-hazard parameters) or in households' distributions 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 provide informal identification arguments addressing each concern.

We first explain how ω, which captures moral hazard, is distinguished from unobserved variation in μ_{kt}, which captures adverse selection on unobservables. 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

[40] Provider prices are a well-documented source of heterogeneity in total healthcare spending across insurers (Cooper et al., 2018), and these differences are often modeled to be linear in utilization (Gowrisankaran, Nevo and Town, 2015; Ghili, 2016; Ho and Lee, 2017; Liebman, 2018). Of course, ϕ_f may also capture other differences across insurers, such as care management protocols or provider practice patterns.

the variation in plan menus.

Both within and across school districts, we observe similar households facing different menus of plans. 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. 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 rely on 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 spending fixed. α^{OOP} is identified by the rate at which households trade off premiums with expected out-of-pocket spending, holding uncertainty in out-of-pocket spending 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 arguments could be applied household by household, and we could nonparametrically identify the distribution of ψ and ω in the population. In reality, we have at most five observations for each household. We ask less of this data by assuming that the unobserved heterogeneity is normally distributed. 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 manifest as unobserved heterogeneity in risk-aversion.

IV.C Estimation

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

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

Covariate vectors $\mathbf{X}^{\mu}_{it}$, $\mathbf{X}^{\sigma}_{it}$, and $\mathbf{X}^{\kappa}_{it}$ contain indicators for the 0–30th, 30–60th, 60–90th, and 90–100th percentiles of individual health risk scores each year.[41] $\mathbf{X}^{\mu}_{it}$ and $\mathbf{X}^{\kappa}_{it}$ also contain a linear term in risk score, separately for each percentile group. $\mathbf{X}^{\mu}_{it}$ also contains an indicator for whether the individual is a female between the ages of 18 and 35 and for whether the individual is under 18 years old.

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

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

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

We assume that μ_{kt}, ω_k, and $\log(\psi_k)$ 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

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

whether the household has children and a constant.[42]

We model inertia at both the plan and insurer level: $\gamma_{kjt}^{inertia} = \gamma_k^{plan}\mathbf{1}_{kt,j=j(t-1)} + \gamma_k^{ins}\mathbf{1}_{kt,f=f(t-1)}$. We allow γ_k^{plan} to vary by whether a household has children. To capture whether sicker households face higher barriers to switching insurers (and therefore provider networks), we allow γ_k^{ins} to vary linearly with household risk score.[43] Insurer fixed effects $\delta_k^{f(j)}$ can 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).[44] We normalize the fixed effect for Moda to be zero. As the parameters of 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 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 approximate the jointly normal distribution of unobserved heterogeneity, as well as to approximate the lognormal distributions of household health states. Additional details on the estimation procedure are provided in Appendix C.2.

V Results

V.A Model Estimates

Table 3 presents our parameter estimates. Column 3 presents our primary specification, as described in the previous section. 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.10.

Column 1 presents a version of the model in which there is no moral hazard and no observable heterogeneity in individuals' health. That is, ω is fixed at zero, and we do not allow μ_{it}, σ_{it},

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

[43] Household risk score is calculated as the mean risk score of all individuals in a household across all years.

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

Table 3. Parameter Estimates

Variable	(1) Parameter	(1) Std. Err.	(2) Parameter	(2) Std. Err.	(3) Parameter	(3) Std. Err.
Employee Premium ($000s)	$-1.000^{\dagger}$		$-1.000^{\dagger}$		$-1.000^{\dagger}$	
Out-of-pocket spending, $-\alpha^{OOP}$	−1.628	0.023	−1.661	0.024	−1.469	0.019
HRA/HSA contributions, α^{HA}	0.255	0.021	0.259	0.020	0.259	0.020
Vision/dental contributions, α^{VD}	1.341	0.024	1.302	0.022	1.209	0.021
Plan inertia intercept, γ^{plan}	4.763	0.060	4.431	0.056	4.630	0.063
Plan inertia * $\mathbf{1}$[Children], γ^{plan}	−0.129	0.039	−0.102	0.037	−0.138	0.038
Insurer inertia intercept,γ^{ins}	2.605	0.107	2.509	0.102	2.413	0.097
Insurer inertia * Risk score, γ^{ins}	−0.074	0.083	−0.120	0.078	−0.037	0.080
Narrow net. plan, $\nu^{NarrowNet}$	−2.440	0.155	−2.286	0.145	−2.334	0.151
Providence utiliz. multiplier, ϕ_P	1.022	0.018	1.072	0.017	1.063	0.002
Risk aversion intercept, $\boldsymbol{\beta}^{\psi}$	−0.706	0.046	−1.018	0.059	−0.251	0.052
Risk aversion * $\mathbf{1}$[Children], $\boldsymbol{\beta}^{\psi}$	0.005	0.031	−0.367	0.083	−0.361	0.050
Moral hazard intercept, $\boldsymbol{\beta}^{\omega}$					1.028	0.038
Moral hazard * $\mathbf{1}$[Children], $\boldsymbol{\beta}^{\omega}$					0.671	0.008
Std. dev. of private health info., σ_μ	0.683	0.002	0.331	0.064	0.225	0.005
Std. dev. of log risk aversion, σ_ψ	0.701	0.062	1.140	0.012	0.833	0.021
Std. dev. of moral hazard, σ_ω					0.281	0.013
Corr(μ, ψ), $\rho_{\mu,\psi}$	0.130	0.018	−0.365	0.049	0.227	0.005
Corr(ψ, ω), $\rho_{\psi,\omega}$					−0.137	0.042
Corr(μ, ω), $\rho_{\mu,\omega}$					0.062	0.017
Scale of idiosyncratic shock, σ_ϵ	2.313	0.025	2.160	0.023	2.116	0.024
Insurer * {Region, Age, $\mathbf{1}$[Child.]}	Yes		Yes		Yes	
Observable heterogeneity in health			Yes		Yes	
Number of observations	451,268		451,268		451,268	

Notes: The table presents estimates for selected parameters; Table A.10 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 34,606 households, 11 plans, and 5 years. The utilization multiplier for Moda is normalized to one. †By normalization.

or κ_{it} to vary with individual demographics. 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. In column 3, we introduced moral hazard. Here, an important difference is the increase in the estimated amount of risk aversion. With moral hazard as an available explanation, the model can explain a larger part of the dispersion in spending for observably similar households. This implies that households are facing less uncertainty in their health state 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}) falls.

Using column 3, we estimate an average moral hazard parameter ω of \$1,001 among individual households and \$1,478 among families.[45] Recall that ω represents the additional total spending induced by lowering marginal out-of-pocket cost from one to zero. Our estimates imply that moving a household from a plan where their health state was below the deductible to a plan where their health state would put them past the out-of-pocket maximum would increase total spending by 15.8 percent of mean spending for individuals and 11.4 percent for families.

Our estimates imply a mean (median) coefficient of absolute risk aversion of 0.92 (0.85).[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 \$91.7 (\$92.1).[47] We note, however, that our estimates of risk aversion are with respect to both financial risk *and* risk in the value derived 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 \$1,152 on average across household-plan-years. The average standard deviation of out-of-pocket costs alone (c^*_{jt}) is \$1,280. To avoid a normally distributed lottery with mean zero and standard deviation \$1,152 (\$1,280), the median household would be willing to pay \$489 (\$544).

The importance of unobserved heterogeneity varies for health, risk aversion, and moral hazard.[48] Once we account for the full set of household observables and moral hazard, the estimated amount of private information about health is fairly small: Unobserved heterogeneity in μ_{kt} accounts for only 11 percent of the total variation in μ_{kt} across household-years. On the other hand, unobserved heterogeneity in risk aversion accounts for 93 percent of its total variation across households. Unobserved heterogeneity in the moral hazard parameter accounts

[45]For comparison, the average ω estimated by Einav et al. (2013) is \$1,330, in a sample of households with average total healthcare spending of \$5,283. In our sample, average total spending is \$6,339 for individuals and \$12,954 for families.

[46]Note that 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.

[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 C.3.

for 18 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 tend to 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, our estimates imply that households' expected health state $\mathbb{E}[\tilde{l}]$ has a correlation of 0.22 with risk aversion, and a correlation of 0.25 with the moral hazard parameter. The risk aversion and moral hazard parameters have a correlation of only 0.01. Figure A.3 plots the full joint distribution of these three key dimensions of household type.

Our estimates imply substantial disutility from switching insurer or plan. The average disutility from switching insurer is $2,408, and from switching plans (but not insurers) is $4,562. We estimate that insurer inertia is decreasing in household risk score, and that plan inertia is on average $138 lower for households with children.[49] The exceptionally large magnitudes of our inertia coefficients reflect, in large part, the infrequency with which households switch plan or insurer, as shown in Table 2. Only 3 percent of household-years ever voluntarily switch insurer, and only 20 percent of household-years ever voluntarily switch plan.

Finally, the estimates in column 3 indicate that households weight out-of-pocket expenditures 46.9 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.[50] 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).

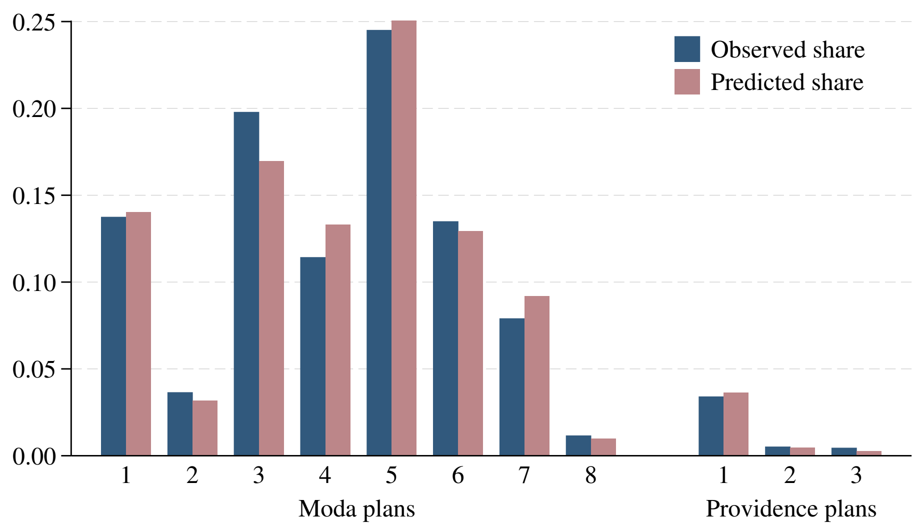
Model Fit. We conduct two procedures to evaluate model fit, corresponding to the two stages of the model. First, we compare households' predicted plan choices with those observed in the data. Figure 3 displays the predicted and observed market shares for each plan, pooled across

[49] 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.

[50] 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. At this tax rate, a dollar of out-of-pocket spending (after tax) would be equivalent to 1.41 dollars of premiums (pre-tax).

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 value of healthcare utilization. Given the relative inflexibility of the model with respect to household plan choice within an insurer, the fit is quite good.

Figure 3. Model Fit: Plan Choices



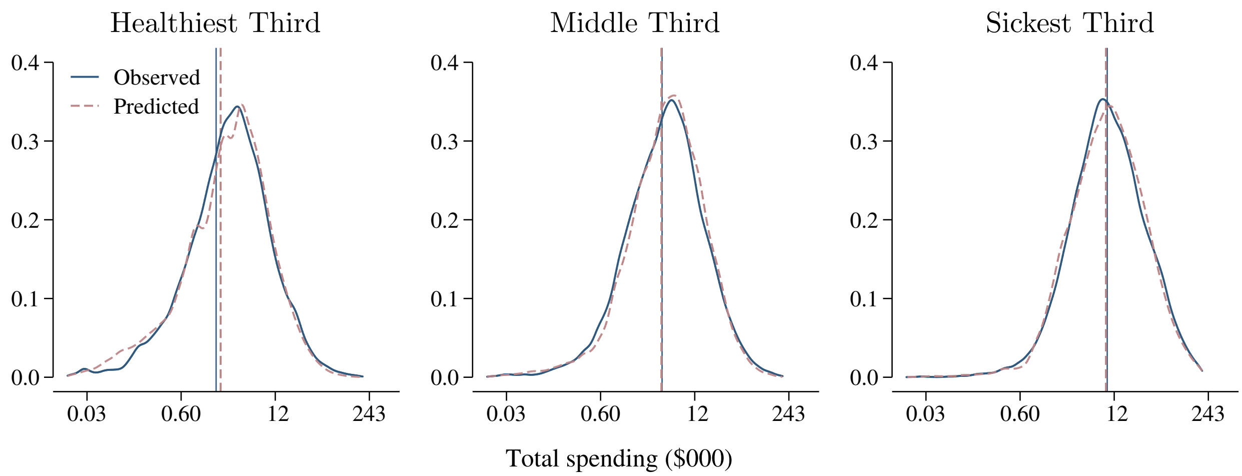
Notes: The figure shows predicted and observed market shares at the plan level. All years are pooled, so an observation is a household-year. Predicted shares are calculated using the estimates in column 3 of Tables 3 and A.10.

Second, we compare the predicted distributions of households' total healthcare spending to the distributions of total healthcare spending observed in the data. In a given year, each household faces a predicted distribution over health states and, due to moral hazard, a corresponding plan-specific distribution of total healthcare spending. To construct the predicted distribution of total spending in a population of households, we take a random draw from each household's predicted spending distribution corresponding to their chosen plan. Figure 4 presents kernel density plots of the predicted and observed distributions of total healthcare spending. We assess fit separately by tertile of household risk score (the average risk score of individuals in the household). Vertical lines in each plot represent the mean of the respective distribution. Overall, average total healthcare spending is observed to be \$11,689 and predicted to be \$11,632. The standard deviation of total healthcare spending is observed to be

[51]The model predicts 67.5 percent of household plan choices correctly (i.e., assigns the highest predicted probability to the correct plan). If households were modeled to be choosing randomly, 25.2 percent of plan choices would be predicted correctly.

$20,803 and predicted to be $20,174. The spending distributions fit well both overall and in subsamples of households, reflecting our flexible approach to modeling household health state distributions.

Figure 4. Model Fit: Healthcare Spending, by Tertile of Households by Risk Score



Notes: The figure shows kernel density plots of the predicted and observed distributions of total healthcare spending on a log scale, separately by tertile of household health risk score, conditional on predicted/observed spending greater than $10. All years are pooled, so an observation is a household-year. Vertical lines represent the mean of the respective distribution. Predicted distributions are based on estimates from column 3 in Tables 3 and A.10. Overall, the observed probability of household spending less than $10 is 2.9 percent, and the predicted probability is 2.8 percent.

V.B Evaluating Vertical Choice

We now construct the ingredients needed to evaluate the optimal plan menu: 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 first specify the contracts under consideration.

Potential Contracts. We consider contracts that are vertically differentiated and have a deductible, coinsurance rate, and out-of-pocket maximum design.[52] While our numerical simulations consider all coverage levels between the null contract and full insurance, we limit attention in our graphical analysis to the range of coverage levels that are ultimately relevant given our parameter estimates. The lowest level of coverage we consider graphically is a contract with a deductible and out-of-pocket maximum of $10,000. The highest level of coverage

[52]In addition to being vertically differentiated, the contracts we consider are also well-ordered in the amount of risk protection provided. Appendix A.2 provides the conditions on contracts that imply this ordering.

remains full insurance. We begin by considering five contracts spanning this range, and refer to them as Catastrophic, Bronze, Silver, Gold, and full insurance. The contracts' actuarial values are 0.53, 0.61, 0.72, 0.86, and 1.00. Their out-of-pocket cost functions are depicted in Figure A.4a.[53] We revisit the specification of potential contracts in Section V.C.

Willingness to Pay. We make several simplifications to our empirical model in order to map it from the setting in Oregon back to our theoretical model, maintaining parameterizations and the estimated distribution of consumer types. To start, we put aside intertemporal variation in household health 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 $\theta_k = \{F_k, \psi_k, \omega_k\}$, where F_k is a shifted lognormal distribution described by parameters $\{\mu_k, \sigma_k, \kappa_k\}$.[54] With respect to payoffs (equation 6), we (i) hold all non-financial features fixed, so any insurer fixed effects cancel; (ii) suppose households choose from the new menu of contracts for the first time, removing any effects of inertia; (iii) assume the idiosyncratic shock is not welfare-relevant;[55] and (iv) set α^{OOP} to one so that premiums and out-of-pocket costs are valued one-for-one.[56]

With attention restricted to the dimension of coverage level, we can use equation 2 to express willingness to pay under our parameterization:[57]

$$WTP(x,\theta_k) = \underbrace{\mathbb{E}_l[c_{x_0}(l) - c_x(l)]}_{\substack{\text{Expected reduction in out-of-pocket} \\ \text{cost holding behavior fixed}}} + \underbrace{\mathbb{E}_l[\tfrac{\omega_k}{2}\left(1 - c_x'(m^*(l,\omega_k,x))\right)^2]}_{\substack{\text{Expected payoff from moral} \\ \text{hazard spending}}} + \underbrace{\Psi(x,\theta_k)}_{\substack{\text{Value of risk} \\ \text{protection}}} .$$

As before, willingness to pay is composed of three parts: the "transfer" of expected out-of-pocket costs holding behavior fixed (at uninsured behavior), the expected payoff from moral

[53]The deductibles, coinsurance rates, and out-of-pocket maximums are \$10,000, –, \$10,000 for Catastrophic; \$5,846, 40%, \$7,500 for Bronze; \$3,182, 27%, \$5,000 for Silver; and \$1,125, 15%, \$2,500 for Gold.

[54]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 C.3.

[55]As our model allows for rich heterogeneity in preferences over financially differentiated contracts, we are comfortable with the interpretation that any remaining determinants of plan choice contained in ϵ can be considered "mistake-making" (e.g., Handel and Kolstad, 2015) or "monkey-on-the-shoulder tastes" (Akerlof and Shiller, 2015), and so can be omitted from the social welfare calculation. In our counterfactuals, we suppose consumers have access to a tool that perfectly aids them in expressing their true preferences. Our question is whether, for this dimension of choice, such a tool is needed.

[56]Otherwise, welfare could be created by moving a dollar of spending between premiums and out-of-pocket cost, which we find undesirable. If α^{OOP} is left as estimated, efficient levels of coverage increase.

[57]See Appendix A.2 for the expression of the value of risk protection.

hazard spending, and the value of risk protection. Recall that only the latter two components are relevant to social welfare.

Figure 5 presents the distribution of willingness to pay among family households.[58] Whereas our point of reference in the two-contract example was x_L, our reference contract now is the Catastrophic contract. We hereinafter refer to "willingness to pay" for a given contract, but emphasize that this is *marginal* willingness to pay with respect to this particular reference. Figure 5, as well as the figures that follow, is composed of connected binned scatter plots: Households are ordered on the horizontal axis according to their willingness to pay, those at each percentile are binned together, and the average value of the vertical axis variable is plotted for each bin.[59] These 100 points are then connected with a line. The left panel shows the willingness to pay curves for our candidate contracts. As contracts 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 the Catastrophic contract.[60]

The right panel shows, for one contract, the decomposition of willingness to pay. We find that the transfer represents the majority of willingness to pay for most households, but that this varies across the distribution of willingness to pay. For households with low willingness to pay, about half is made up by the transfer. For households with high willingness to pay, nearly all is made up by the transfer. The highest willingness-to-pay households are willing to pay $7,500 more for Gold than for Catastrophic only in order to avoid paying an additional $7,500 in out-of-pocket costs. Importantly, this means that allocating them to higher coverage generates almost no additional social surplus.

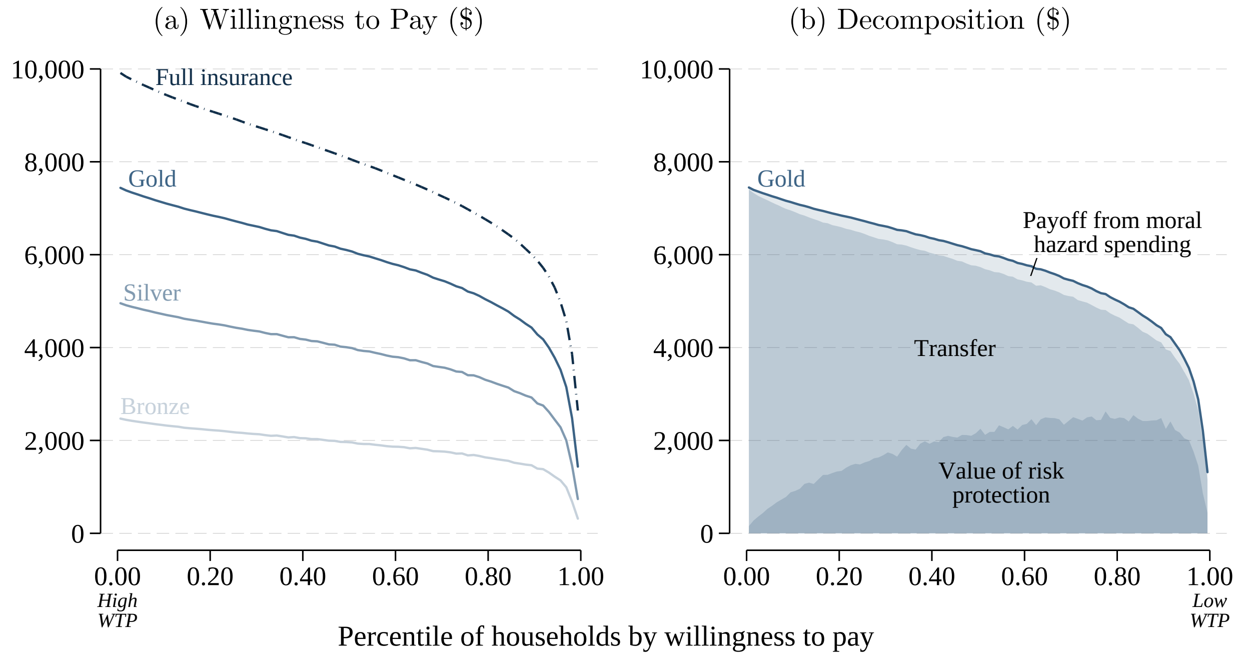
Social Surplus. As in equation 3, the social surplus generated by allocating a household to a given contract is the difference between willingness to pay and expected insurer cost, which

[58]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 typically offered to families. Our results among individual households are qualitatively unchanged.

[59]Households are in fact ordered by willingness to pay for full insurance, but the ordering is nearly identical across contracts. The consistent willingness-to-pay ordering of households across contracts is what permits a graphical analysis of multiple contracts analogous to the two-contract example in Figure 1. See Geruso et al. (2019) for a detailed discussion of this point.

[60]Figure A.5 provides demographic information about households across the distribution of willingness to pay. Higher willingness-to-pay households tend to be older, have more family members, be more risk averse, and most strikingly, have higher expected healthcare spending.

Figure 5. Willingness to Pay



Notes: The figure shows the distribution across households of (a) willingness to pay for each contract and (b) the decomposition of willingness to pay for the Gold contract. 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 relative to the Catastrophic contract.

under our parameterization is equal to:

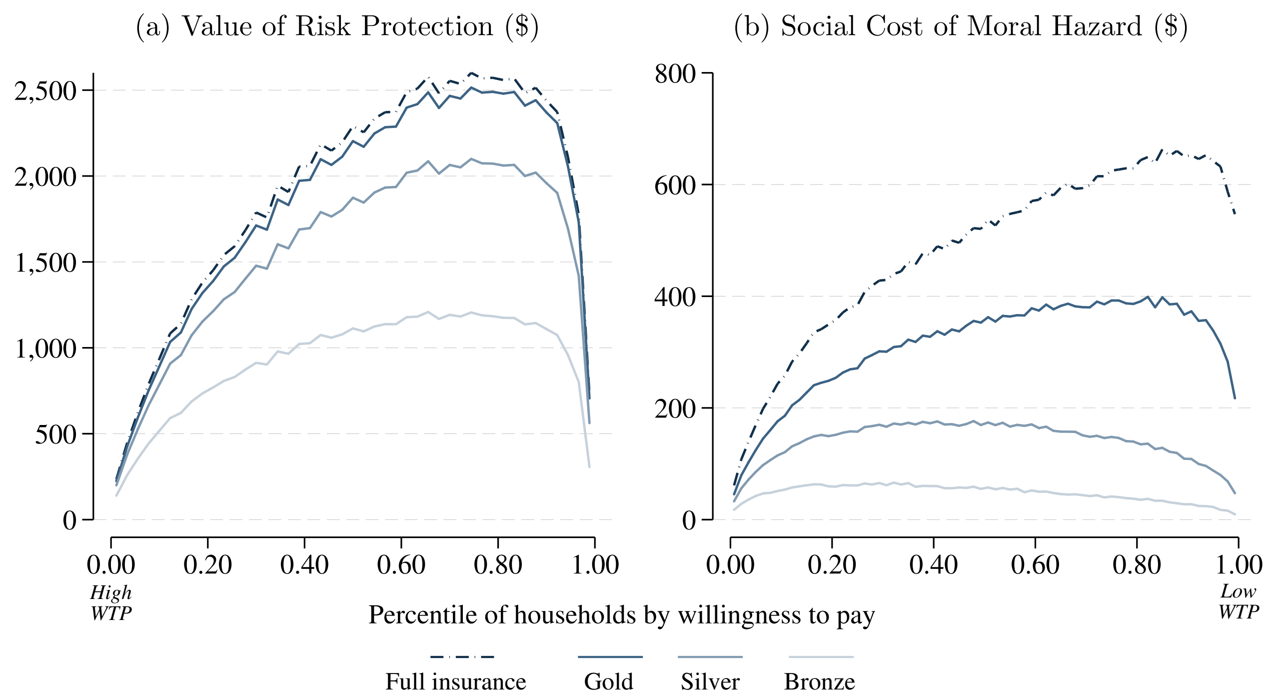
$$SS(x,\theta_k) = \underbrace{\Psi(x,\theta_k)}_{\substack{\text{Value of risk}\\ \text{protection}}} - \underbrace{\mathbb{E}_l\left[\tfrac{\omega_k}{2}\left(1-c_x'(m^*(l,\omega_k,x))\right)^2\right]}_{\substack{\text{Social cost}\\ \text{of moral hazard}}}. \tag{10}$$

The value of risk protection varies in the population to the extent there is variation in risk aversion and in the amount of uncertainty about out-of-pocket costs. The social cost of moral hazard varies in the population to the extent there is variation in the moral hazard parameter and in consumers' expected marginal out-of-pocket cost.

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 marginal value of risk protection generated by each contract, relative to the Catastrophic contract. We find that the majority of the social welfare gains from more generous coverage are driven by households with intermediate levels of willingness to pay. This key pattern is driven by the concavity of the contracts we consider.

High willingness-to-pay households are likely to realize health states that put them above the out-of-pocket maximum of every contract, leaving them little uncertainty about out-of-pocket costs. Among the top fifth of households by willingness to pay, the probability of spending more than \$10,000, even without moral hazard, is 65 percent. Variation in uncertainty only becomes meaningful for households for whom much of the density of their spending distribution lies in the range \$0–\$10,000, within which marginal out-of-pocket cost varies across contracts. Figure A.6 shows the full spending distributions faced by households at different levels of willingness to pay.

Figure 6. Components of Social Surplus



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

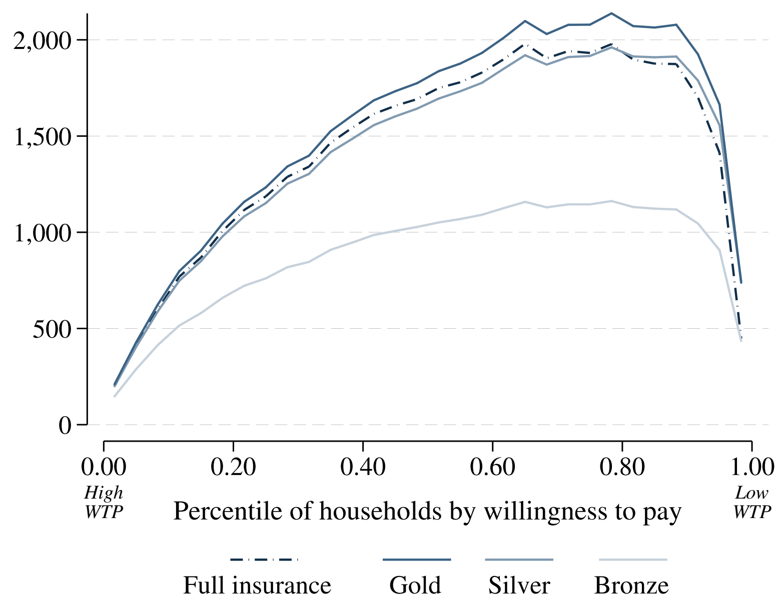
Figure 6b then shows the distribution of the marginal social cost of moral hazard. It provides two important insights. First, high willingness-to-pay households on average barely change their behavior across this range of coverage levels.[61] For similar reasons as with risk protection,

[61] We note that this finding is closely related to 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; **?**), here we do not find evidence of moral hazard among high-risk households (see Table A.7). If the data did suggest 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 strong treatment effect.

the majority of the social cost of more generous coverage comes from households with lower willingness to pay. The second insight is that the Gold contract recovers about half of the social cost of moral hazard induced by full insurance. The \$1,125 deductible is high enough to deter excess spending, but low enough to sacrifice only a small amount of risk protection.

Finally, Figure 7 shows the resulting social surplus curves, equal to the vertical differences between the curves in Figures 6a and 6b. The social surplus curves represent the average social surplus acheived by allocating all households at a given percentile of willingness to pay to a given contract, relative to allocating them to the Catastrophic contract. Since households can be screened only by their willingness to pay, the figure permits a direct assessment of the optimal menu.

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 (to reduce noise) quantiles of households by willingness to pay.

First, note that all curves lay everywhere above zero, meaning the Catastrophic contract (and any lower level of coverage) should be unambiguously excluded from the optimal menu.[62] Among the remaining contracts, the social surplus curves of Bronze, Silver, and full insurance similarly lay everywhere below that of the Gold contract, which delivers higher average social

[62] Any lower level of coverage can be ruled out because its social surplus curve will lay everywhere below that of the Catastrophic contract (c.f. Proposition 3 in Appendix A.2).

surplus at *every* level of willingness to pay.[63] Households with higher willingness to pay should therefore not have a higher level of coverage than households with lower willingness to pay: they should, on average, have the same level of coverage. It follows that offering choice over these contracts is not efficient in this population. Numerical optimization confirm this result. The optimal menu consists of only the Gold contract, and this allocation achieves social surplus (relative to allocating all households to Catastrophic) equal to the integral of the Gold social surplus curve: $1,514 per household.

V.C Robustness

More Contracts. A natural question is how the optimal menu would change if more contracts were available. Figure 7 indicates that the Silver contract is everywhere too little coverage, and that full insurance is everywhere too much coverage, but it says nothing about the potential gains of offering additional contracts within this range. We explore this question by expanding the number of contracts in the Silver-to-full insurance range from one (the Gold contract) to 20. The out-of-pocket cost functions for this denser set of potential contracts are depicted in Figure A.4b.

We find that when efficient coverage level can be measured more finely, high willingness-to-pay households do have a slightly higher efficient level of coverage. In a small neighborhood of the Gold contract, it is therefore efficient to offer a choice. The optimal menu features four contracts.[64] This allocation achieves social surplus, relative to allocating all households to the Catastrophic contract, of $1,528 per household. This represents a gain of $14 over what is achieved by the Gold contract alone, and of only $5 over what can be achieved by a single contract in the denser set.[65] Because social surplus is so flat in this range of coverage levels, the welfare gains from a denser contract space are economically small. Other considerations, such as any fixed costs of offering each contract, or of offering a choice at all, could quickly become first order.

Different Contracts. We next explore whether our results are robust to alternative contract

[63] Although Gold is the efficient contract at every level of willingness to pay, it is not the efficient contract for every household. Figure A.7 shows the heterogeneity in households' efficient contracts.

[64] The four contracts are the Gold contract (actuarial value 0.86) and the three next-less-generous contracts (actuarial values 0.84, 0.83, and 0.81). At the optimal feasible allocation, 28 percent of households choose Gold, and 34 percent, 37 percent, and 1 percent of households choose the next three contracts respectively.

[65] The optimal single contract in the dense set is the 0.83 actuarial value contract.

designs. We have so far considered one particular design, as depicted in panels (a) and (b) of Figure A.4. We now consider three alternatives, as depicted in panels (c)–(e). These are: (c) removing deductibles, (d) removing the coinsurance region, and (e) extending the coinsurance region. Within each alternate contract design, we consider a set of five vertically differentiated contracts.

We solve for the optimal menu within each new set of contracts. These results are presented in Table A.11. We find that among contracts without a deductible and without a coinsurance region, the optimal menu again features a single contract. For much the same reasons as this was true among the original contracts, higher willingness-to-pay consumers do not have a higher efficient level of coverage. We also find that among contracts with an extended coinsurance region, the optimal menu does feature vertical choice. Because it takes longer to reach the out-of-pocket maximum, households are less likely to hit it, and high willingness-to-pay households face much more variation in risk across contracts.

Taken together, we find that the contract dimension most relevant for the optimality of vertical choice is the stop-loss point, or the level of total spending at which the out-of-pocket maximum is reached. Our results suggest that if a regulator wanted consumers to pay on the margin for only a short time (a low stop-loss point), vertical choice may not offer meaningful welfare gains. If instead a regulator wanted consumers to pay on the margin for a long time (a high stop-loss point), our results suggest it may be useful to offer consumers a choice.[66]

Different Consumers. We next explore the robustness of our findings to different populations of consumers. We do this in two ways: (i) by re-estimating our model in the full sample of households that includes Kaiser enrollees, and (ii) by adjusting individual parameter estimates to establish their isolated effects on results.

As discussed in Section III.A, we exclude Kaiser enrollees from our primary analysis sample in order to focus on the vertical choice across coverage levels, as opposed to the horizontal choice across plan types (HMO vs. PPO). Kaiser enrollees are on average slightly younger and healthier on average, and 3 percent of households did at one point switch between a Kaiser and non-Kaiser plan. We investigate how these factors impact our results by re-estimating our

---

[66]Evaluating a regulator's choice between these options is no longer a question of vertical choice. Though our estimates suggest that a lower stop-loss point is more efficient, we acknowledge that there are important considerations our model may not capture. For example, consumers may inefficiently restrict utilization in response to even moderate marginal out-of-pocket costs (Aron-Dine et al., 2015; Dalton, Gowrisankaran and Town, 2020), which may pull in favor of a lower stop-loss point, or they may benefit from smoothing out-of-pocket spending within a year (Ericson and Sydnor, 2018; Hong and Mommaerts, 2021), which may pull in favor of a higher stop-loss point.

model using the full sample of households. Table A.12 presents these parameter estimates.[67] Figure A.8 presents the corresponding willingness to pay and social surplus curves. Though the shapes and levels of the resulting social surplus curves are slightly different than under our original estimates, our qualitative results and the underlying mechanisms are unchanged. The optimal menu remains the Gold contract alone.

Second, we explore how specific perturbations of our parameter estimates affect our results. We explore nine cases, including raising and lowering the moral hazard parameter, raising and lowering risk aversion, and increasing heterogeneity in risk aversion and the moral hazard parameter.[68] We also present three cases in which households vary only in their preferences: risk aversion and/or the moral hazard parameter. Our findings are summarized in Table A.13. For each case, the table shows the percent of households enrolled in each contract under the optimal menu. Intuitively, we find that the optimal menu is more likely to feature a choice when risk aversion plays a larger role in driving variation in willingness to pay. At the extreme, when households vary only in their risk aversion, nearly perfect screening is possible as private and social incentives are directly aligned. The table also reports the welfare gains available from a denser contract space, and whether or not a choice would be efficient in that context. We find that while choice is almost always efficient in the denser contract space, the welfare gains available are consistently small, never exceeding \$16 per household per year. In the extreme case in which households vary only in their moral hazard parameter, private and social incentives are directly misaligned, and choice is not efficient even among the dense set of contracts. Across all nine cases, the welfare gains from choice relative to what can be achieved by a single contract do not exceed \$10 per household. In a broad neighborhood of our parameter estimates, the efficiency loss from forgoing vertical choice is therefore either zero or economically small.

[67] We make two changes from our preferred specification. First, we estimate insurer inertia terms separately for Kaiser and for Moda/Providence. Second, we estimate the moral hazard parameter ω only among Moda/Providence plans, as opposed to among all three insurers. Though it would be interesting to also consider a Kaiser-specific ω, limited variation in coverage level among Kaiser plans prevents us from estimating it. Any Kaiser-specific effects of coverage level on utilization are absorbed into the utilization multiplier ϕ_{Kaiser}.

[68] We present fairly large perturbations, changing our estimates by a factor of 2, in order to show cases in which our results do vary. Smaller changes to our parameter estimates, e.g., raising and lowering mean risk aversion by up to 30 percent, do not affect our results

VI Counterfactual Pricing Policies

Returning to our focal set of metal-tier contracts and the estimated distribution of consumer types, 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 and a mandate, (iv) competitive pricing with type-specific prices and a mandate, and (v) premiums to support vertical choice. Under regulated pricing, premiums are set to maximize social surplus. Under competitive pricing, premiums are endogenous and must equal average costs on a plan-by-plan basis. A mandate enforces a minimum level of coverage at the Catastrophic contract. Under premiums to support vertical choice, premiums are set to support the availability of (read: enrollment in) every contract.

We consider two pricing policies, (ii) and (iv), in which premiums can vary by consumer attributes. If observable dimensions of household type are predictive of efficient coverage level, tailoring plan menus to observables 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.[69] 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.

VI.A Welfare Outcomes

Table 4 summarizes outcomes under each of these five pricing policies. It shows the percent of households Q enrolled in each contract at the optimal feasible allocation, the percent of first-best social surplus achieved, and the expected per-household insurer cost AC among households in each contract (in thousands of dollars). We benchmark outcomes against the first best allocation of households to contracts (as depicted in Figure A.7), which cannot be supported by prices unless premiums can vary by all aspects of consumer type. The first best allocation generates \$1,542 in social surplus per household relative to the counterfactual of allocating all households to the Catastrophic contract. Expected total healthcare spending per household at the first best allocation is \$12,400, and expected insurer cost per household is \$10,351.

Policy (i) is the baseline policy considered in this paper, in which the regulator can set

[69] Among family households, 6 percent are childless and under age 45, 27 percent are childless and over age 45, 52 percent have children and are under age 40, and 15 percent have children and are over age 45.

Table 4. Outcomes Under Alternative Pricing Policies

	Policy	% of First Best Surplus		Potential Contracts				
				Full	Gold	Silver	Bronze	Ctstr.
*	First best	1.000	*Q:*	0.06	0.75	0.19	<0.01	–
			AC:	18.35	9.43	11.48	39.18	–
(i)	Regulated pricing with community rating	0.982	*Q:*	–	1.00	–	–	–
			AC:	–	10.62	–	–	–
(ii)	Regulated pricing with type-specific prices	0.989	*Q:*	–	0.98	0.02	–	–
			AC:	–	10.71	0.75	–	–
(iii)	Competitive pricing with community rating	0.000	*Q:*	–	–	–	–	1.00
			AC:	–	–	–	–	6.30
(iv)	Competitive pricing with type-specific prices	0.075	*Q:*	–	–	0.05	–	0.95
			AC:	–	–	4.95	–	6.41
(v)	Premiums to support vertical choice	0.796	*Q:*	0.01	0.07	0.63	0.28	0.01
			AC:	61.04	31.91	8.47	1.75	0.28

Notes: The table summarizes outcomes under five pricing policies as well as the first best allocation, among the 25,636 family households. Q represents the percent of households enrolled in each contract. AC represents average expected insurer cost (in thousands of dollars) among households enrolled in each contract. Social surplus is measured relative to the Catastrophic contract. At the first best allocation, social surplus is $1,542 per household and expected insurer cost is $10,351 per household.

premiums but is restricted to community rating. As indicated by Figure 7, it is welfare maximizing to offer only the Gold contract.[70] Interestingly, although 25 percent of households are misallocated, this policy is almost equally as efficient as the first best allocation. That is, the ability to perfectly discriminate among consumers would increase welfare by only $28 per household per year.[71] Driving this result is the fact that social welfare is quite flat across the top contracts, and particularly so among the households who are misallocated under policy (i). Among these households, the social surplus at stake between the Silver and Gold contracts is on average only $26; among all households, it is $112.

[70]This allocation is implementable because the regulator need not break even in aggregate. The Gold contract can be provided for free, and the deficit of $10,619 per household can be funded by taxing incomes (here, at zero cost of public funds). We note that if the regulator *did* need to break even in aggregate, vertical choice would likely be efficient. The focus would shift to ensuring low-WTP consumers were not left out of the market entirely, even if that induced some high-WTP consumers to select lower-than-efficient coverage. See Azevedo and Gottlieb (2017) and Geruso et al. (2019) for a full treatment of a setting in which the regulator must break even in aggregate.

[71]In an interesting parallel, Ho and Lee (2021) find that the ability to perfectly discriminate among consumers would increase welfare by only $33 per household per year relative to what can be achieved by a single contract.

Because pricing policy (i) is almost as efficient as the first best outcome, there is little scope for improvement by varying prices by consumer type. Even so, under policy (ii) we do find that allowing the regulator to discriminate can improve allocational efficiency by a small amount. To young households under with children, it is efficient to offer a choice between Gold and Silver. To the other three sets of households, it is still efficient to offer only Gold. It becomes possible to productively offer lower coverage to young households with children because the other households, to whom it is not efficient to provide such low coverage, can now be excluded.

Policy (iii) considers competitive pricing with community rating and a mandated minimum level of coverage at the Catastrophic contract. We calculate the competitive equilibrium proposed by Azevedo and Gottlieb (2017).[72] We find that a separating equilibrium above minimum coverage cannot be supported in this population, and the market unravels. Though choice is permitted, the market cannot deliver it. Policy (iv) considers which allocations could be supported if the market could be segmented. We find that among young households with children, a separating equilibrium between the Silver and Catastrophic contracts can be supported. The other three market segments unravel.

The first four policies are natural benchmarks, but none turn out to feature the same degree of vertical choice that is observed in many U.S. health insurance markets, including the market we study. A major difference between these real markets and our benchmark policies is that the former feature a complex set of taxes and subsidies that affect consumer premiums in ways not replicated by our benchmark 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.[73] Because households with intermediate 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.

[72]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.

[73]Shares are from Kaiser Family Foundation and are available at https://www.kff.org/health-reform/state-indicator/marketplace-plan-selections-by-metal-level. We map Platinum coverage to full insurance. Premiums that can support these shares are $7,059 for full insurance, $4,594 for Gold, $2,173 for Silver, $375 for Bronze, and $0 for Catastrophic, resulting in an aggregate deficit of $6,856 per household.

VI.B Distributional Outcomes

The population faces an unavoidable healthcare spending bill of $11,723 per household. It is unavoidable because it arises even if all households have the minimum allowed coverage (Catastrophic). While full insurance provides the benefit of additional risk protection, it also raises the population's healthcare spending bill by 8 percent due to moral hazard, to $12,695 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 and, to the extent optimal premiums imply an aggregate deficit, by taxes. Different coverage levels imply different divisions between out-of-pocket costs and insured costs.[74] There are therefore large differences across policies in the source of funding for the population healthcare spending bill, and thereby in how evenly the spending bill is shared in the population.

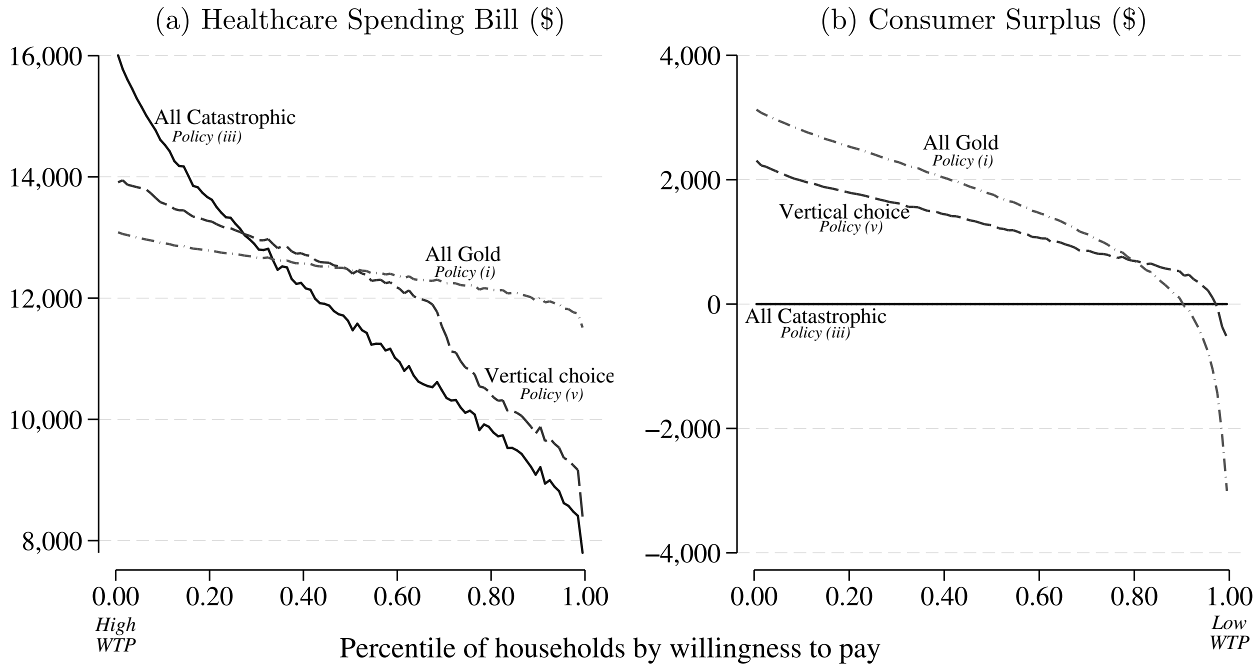
Figure 8 shows distributional outcomes under three of our candidate policies: (i) regulated pricing ("All Gold"), (iii) competitive pricing ("All Catastrophic"), and (v) premiums to support vertical choice ("Vertical Choice"). The left panel shows the distribution across households of the population healthcare spending bill. Each household's spending bill equals the premium plus expected out-of-pocket cost associated with their chosen contract, plus any tax assessed on all consumers. For simplicity (and because we lack information on income), we assess taxes equally across households. Households are again ordered on the horizontal axis according to their willingness to pay.[75] Under "All Catastrophic," there is a premium-plus-tax of $6,298. The highest willingness-to-pay households then also have expected out-of-pocket costs of $9,708, implying a healthcare spending bill of $16,006. The lowest willingness-to-pay households have expected out-of-pocket costs of only $1,500, implying a healthcare spending bill of only $7,798. When the population has higher coverage, as under the other two policies, the healthcare spending bill is shared more evenly in the population.

The right panel evaluates the distribution of consumer surplus, incorporating preferences over risk and healthcare utilization in addition to just spending outcomes. Consumer surplus is typically measured relative to the absence of a product. As we enforce a minimum level of

[74]For example, 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 the spending bill would be insured.

[75]All households at the same level of willingness to pay choose the same contract, and thus pay the same premium plus tax, but there may still be variation in their expected out-of-pocket costs. The plot is therefore composed of connected binned scatter plots, as in the previous figures.

Figure 8. Distributional Outcomes



Notes: The figure shows the distribution across households of (a) expected healthcare spending bill (premium plus tax plus expected out-of-pocket cost), and (b) consumer surplus, under three of the policies considered in Table 4. Households are arranged on the horizontal axis according to their willingness to pay. Consumer surplus equals marginal willingness to pay less marginal premium-plus-tax, relative to the allocation of all households to Catastrophic coverage. The premium-plus-tax that supports the single contract is $6,298 under "All Catastrophic" and $10,619 under "All Gold." Premiums under "Vertical choice" are $7,059 for full insurance, $4,594 for Gold, $2,173 for Silver, $375 for Bronze, $0 for Catastrophic, and a tax of $6,856.

coverage, here it is measured relative to the absence of a better product. Under each policy, consumer surplus is the difference between a household's marginal willingness to pay for their chosen plan, and the marginal premium-plus-tax associated with that choice. The integral of each consumer surplus curve equals the social welfare generated by that policy, relative to the "All Catastrophic" outcome. 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.

Figure 8b depicts a classic feature of insurance markets with adverse selection. The optimal feasible allocation ("All Gold") results in higher coverage and greater social welfare gains, while the competitive outcome ("All Catastrophic") results in lower coverage but a more even distribution of welfare gains. At the competitive outcome, no one is made worse off than they were absent the market. Regulatory intervention can offer substantial efficiency gains, at the cost of making some households worse off.

Dynamic Considerations. These static gains from trade, and the distribution thereof, are evaluated at a point in time at which households are aware of their endowed type, θ. In the spirit of Hendren (2020) and Ghili et al. (2020), we can also consider welfare from the perspective of an "unborn" consumer, who, prior to participating in our spot insurance market, faces a lottery over types.[76] Note that a consumer's type θ uniquely determines their willingness to pay, and thus their position on the horizontal axis in Figure 8. Instead of considering a lottery over types, we can therefore directly consider the lottery over levels of willingness to pay. Under each policy, the "lottery over types" faced by an unborn consumer is equivalent to the uniformly distributed lottery over consumer surplus outcomes shown in Figure 8b.

The question then becomes where to normalize utility across consumers. In Figure 8b, we have assumed consumers are equally well off absent the market. But from "behind the veil of ignorance," it may be more natural to assume they are equally well off when fully insured. Such a renormalization would be reflected in Figure 8b by rotating the consumer surplus curves counter-clockwise, until an "All full insurance" consumer surplus curve were horizontal (as depicted in Figure A.9). In this case, it becomes clear that the "All Gold" policy delivers the least-risky distribution of surplus in the population, consistent with the intuition that higher coverage provides greater dynamic risk protection (Handel, Hendel and Whinston, 2015). Among the three candidate policies, the "All Gold" policy therefore delivers the most efficiency *and* the most equity in the spot market.[77]

VII Conclusion

This paper presents a framework for evaluating optimal menus of coverage levels in regulated health insurance markets. Our framework incorporates consumer heterogeneity along multiple dimensions, endogenous healthcare utilization, and menus of nonlinear insurance contracts among which traded contracts are endogenous. We show how willingness to pay for insurance can be decomposed between a transfer component that is only privately relevant, and the components that gives insurance value beyond as a redistributive tool. We also emphasize

[76] The interpretation that consumers face a lottery over all elements of type, including preferences, is consistent with Harsanyi (1953, 1955). We take this approach because it permits a simple informal analysis, but refer the reader to Eden (2020) for an alternative potential approach.

[77] Given the chosen normalization, maximal equity is achieved by allocating all households to full insurance. Maximal (static) efficiency, meanwhile, is achieved by allocating all households to Gold. A regulator placing some weight on each of these objectives may want to offer a vertical choice between the Gold contract and full insurance.

how the redistributive component plays a central role in determining feasible allocations: If premiums must be uniform, it may not be possible to align the private incentive to maximize one's own transfer with 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 efficiency of vertical choice hinges on whether consumers with higher willingness to pay have higher efficient levels of coverage. In reverse, this condition implies that a lowest-coverage plan should only be offered if the lowest willingness-to-pay consumers are the intended recipients. In the setting we study, we find that lowest willingness-to-pay consumers are sufficiently risk averse, and facing sufficient risk, to warrant coverage as least as high as the Silver contract. Our results also imply that a highest-coverage plan should only be offered if the highest willingness-to-pay consumers are the intended recipients. The highest coverage we consider is full insurance, and we find that it would be more efficient for the highest willingness-to-pay consumers to have lower coverage. Between these bounds, we find that private values for higher coverage are not strongly positively correlated with social values, and thus that offering a choice cannot provide economically meaningful welfare gains. We also find that the welfare stakes of misallocation are low. Relative to what can be achieved by a single contract, the ability to perfectly screen consumers would increase welfare by only \$28 per household per year.[78]

An important limitation of this paper is that we assume the socially optimal level of healthcare utilization is the level a consumer would choose absent insurance. If healthcare providers charge supracompetitive prices, or if there are positive externalities of healthcare utilization, it may well be the case that using insurance to induce additional utilization is desirable. In addition, important considerations our model does not address arise when consumers face liquidity constraints (Ericson and Sydnor, 2018) or are protected from large losses by limited liability in addition to by insurance (Gross and Notowidigdo, 2011; Mahoney, 2015). 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 multiple dimensions of healthcare utilization and insurance as an important direction for future research.

[78] Our estimates imply that allowing prices to vary with certain consumer demographics would close 40 percent of this gap (\$11 out of \$28). Ho and Lee (2021) find that pairing coverage level with a horizontal plan characteristic such as the carrier may also be an effective strategy.

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.

Cooper, Zack, Stuart V Craig, Martin Gaynor, and John Van Reenen. 2018. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured*." *The Quarterly Journal of Economics*, 134(1): 51–107.

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 J. Town. 2020. "Salience, myopia, and complex dynamic incentives: Evidence from Medicare Part D." *Review of Economic Studies*, 87(2): 822–869.

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.

Eden, Maya. 2020. "Welfare Analysis with Heterogeneous Risk Preferences." *Journal of Political Economy*, 128(12): 000–000.

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 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.

Geruso, Michael, Timothy J Layton, Grace McCormack, and Mark Shepard. 2019. "The Two Margin Problem in Insurance Markets." National Bureau of Economic Research Working Paper 26288.

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

Ghili, Soheil, Benjamin R Handel, Igal Hendel, and Michael D Whinston. 2020. "The Welfare Effects of Long-Term Health Insurance Contracts." National Bureau of Economic Research Working Paper 23624.

Glazer, Jacob, and Thomas G. McGuire. 2011. "Gold and Silver health plans: Accommodating demand heterogeneity in managed competition." *Journal of Health Economics*, 30(5): 1011–1019.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town. 2015. "Mergers When Prices Are Negotiated: Evidence from the Hospital Industry." *American Economic Review*, 105(1): 172–203.

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, and 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.

Harsanyi, John. 1953. "Cardinal Utility in Welfare Economics and in the Theory of Risk-taking." *Journal of Political Economy*, 61.

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy*, 63(4): 309–321.

Hendren, Nathaniel. 2020. "Measuring Ex-Ante Welfare in Insurance Markets." *Review of Economic Studies*.

Hendren, Nathaniel, Camille Landais, and Johannes Spinnewijn. 2021. "Choice in Insurance Markets: A Pigouvian Approach to Social Insurance Design." National Bureau of Economic Research Working Paper 27842.

Ho, Kate, and Robin Lee. 2017. "Insurer Competition in Health Care Markets." *Econometrica*, 85(2): 379–417.

Ho, Kate, and Robin Lee. 2021. "Health Insurance Menu Design for Large Employers."

Hong, Long, and Corina Mommaerts. 2021. "Time Aggregation in Health Insurance Deductibles." National Bureau of Economic Research Working Paper 28430.

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.

Landais, Camille, Arash Nekoei, Peter Nilsson, David Seim, and Johannes Spinnewijn. 2021. "Risk-Based Selection in Unemployment Insurance: Evidence and Implications." *American Economic Review*, Forthcoming.

Liebman, Eli. 2018. "Bargaining in Markets with Exclusion: An Analysis of Health Insurance Networks."

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.*

Mahoney, Neale. 2015. "Bankruptcy as Implicit Health Insurance." *American Economic Review*, 105(2): 710–46.

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.

Samek, Anya, and Justin R Sydnor. 2020. "Impact of Consequence Information on Insurance Choice." National Bureau of Economic Research Working Paper 28003.

Shepard, Mark. 2016. "Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange." National Bureau of Economic Research Working Paper 22600.

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.

Appendix A Derivations and Proofs

A.1 Derivation 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:

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

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

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

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,

$$\begin{aligned}
EV(x,\theta) &= \mathbb{E}_l[\ b^*(l,\omega,x) - c_x^*(l,\omega,x)\ ] \\
&= \mathbb{E}_l[\ b^*(l,\omega,x_0) - c_x^*(l,\omega,x_0) + v(l,\omega,x)\ ], \text{ and} \\
RP(x,p,\theta) &= EV(x,\theta) + \hat{y} - p - u_\psi^{-1}(U(x,p,\theta)),
\end{aligned} \tag{11}$$

where as before $v(l,\omega,x) = b^*(l,\omega,x) - b^*(l,\omega,x_0) - \big(c_x^*(l,\omega,x) - c_x^*(l,\omega,x_0)\big)$. A consumer's willingness to pay for contract x relative to the null contract x_0 is equal to $\tilde{p}$ that solves:

$$\begin{aligned}
CE(x,\tilde{p},\theta) &= CE(x_0,0,\theta) \\
EV(x,\theta) + \hat{y} - \tilde{p} - RP(x,\tilde{p},\theta) &= EV(x_0,\theta) + \hat{y} - RP(x_0,0,\theta) \\
\tilde{p} &= EV(x,\theta) - EV(x_0,\theta) + RP(x_0,0,\theta) - RP(x,\tilde{p},\theta).
\end{aligned}$$

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.[79] In this case, marginal willingness to pay for contract x relative to the null contract is given by:

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

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.

A.2 Definitions and Proofs

Assumptions. Consider the model in Section II.A. Suppose contracts $x \in X$ are characterized by increasing, continuous, and concave out-of-pocket cost functions $c_x : \mathbb{R}_+ \to \mathbb{R}_+$, where $c_x(m) \le m\ \forall\ m$ and which are differentiable almost everywhere with $c_x' \in [0,1]$, where c_x' denotes the derivative wherever it exists. Suppose consumers have type $\theta = (F,\omega,\psi) \in \Delta^c(\mathbb{R}) \times \mathbb{R}_{++} \times \mathbb{R}_{++} =: \Theta$.[80] Given health state realization $l \in \mathbb{R}$, contract premium p, and initial income $\hat{y}$, suppose consumers value healthcare spending $m \in \mathbb{R}_+$ according to $u_\psi\left(\hat{y} - p + b(m;l,\omega) - c_x(m)\right)$, where $b(m;l,\omega) = (m-l) - \frac{1}{2\omega}(m-l)^2$ and where $u_\psi(x) = -\exp(-\psi x)$.

[79] In equation 11, $\hat{y}-p$ cancels out completely. This assumption is most reasonable when marginal premiums between relevant plans are small relative to initial income.

[80] $\Delta^c(\mathbb{R})$ denotes the the set of continuous probability measures on the Borel σ-algebra of $\mathbb{R}$.

Under these assumptions, social surplus is given by $SS(x,\theta) = \Psi(x,\theta) - \textsc{scmh}(x,\theta)$, where

$$\Psi(x,\theta) = RP(x_0,\theta) - RP(x,\theta)$$

$$\text{where} \quad RP(x,\theta) = \psi^{-1} \log\Big(\mathop{\mathbb{E}}_{l\sim F}\big[\exp\big(-\psi(z_x(l,\theta) - \bar{z}_x(\theta))\big)\big]\Big),$$

$$\textsc{scmh}(x,\theta) = \mathop{\mathbb{E}}_{l\sim F}\Big[\tfrac{\omega}{2}\big(1 - c'_x(m^*(l,\omega,x))\big)^2\Big],$$

and where $z_x(l,\theta) = \hat{y} - p + b(m^*(l,\omega,x);l,\omega) - c_x(m^*(l,\omega,x))$ and $\bar{z}_x(\theta) = \mathbb{E}_{l\sim F}[z_x(l,\theta)]$. Appendix C.2 solves for privately optimal spending $m^*(l,\omega,x) = \text{argmax}_m\,(b(m;l,\omega) - c_x(m))$ when contracts are piecewise linear with a deductible, coinsurance rate, and out-of-pocket maximum. As m^* never falls on a kink, $c'_x(m^*)$ always exists. The indirect benefit from privately optimal spending is given by $b(m^*(l,\omega,x);l,\omega) = \frac{\omega}{2}\,(1 - c'_x(m^*(l,\omega,x))^2)$. Willingness to pay is given by $WTP(x,\theta) = \bar{z}_x(\theta) - \bar{z}_{x_0}(\theta) + \Psi(x,\theta)$.

Definitions. We say that a given contract is "higher coverage" than another if it provides both a higher certainty equivalent payoff $WTP(x,\theta)$ as well as greater risk protection $\Psi(x,\theta)$. This notion of coverage level is slightly stronger that what is implied by vertical differentiation alone. We use it because it allows our model to have the following desirable properties:

(i) the value of risk protection is increasing in coverage level;

(ii) the social cost of moral hazard is increasing in coverage level;

(iii) efficient coverage level is increasing in risk aversion;

(iv) efficient coverage level is decreasing in the moral hazard parameter.

Definitions 1 and 2 formalize the distinction between vertical differentiation and coverage level ordering. Propositions 1 and 2 provide the conditions on contracts that yield each ordering. Briefly, vertical differentiation requires only a relation on contracts' *level* of out-of-pocket costs, while coverage level ordering (as defined) also requires a relation on contracts' *marginal* out-of-pocket costs. A higher-coverage contract must have an out-of-pocket cost function that is everywhere below and everywhere flatter than a lower-coverage contract.

Implications. The most important reason we use this notion of coverage level is that it allows extrapolation of social surplus across coverage levels. Namely, it implies that social surplus is single peaked in coverage level. Proposition 3 states this formally. Single-peakedness allows one to infer, for example, that if a given contract is less-than-socially-optimal coverage for all households, the same would be true of any lower level of coverage.

Proofs are provided below. Of the four stated properties of the model, property (i) is true by definition, property (ii) is established in the proof of Proposition 3, and properties (iii) and (iv) are proved in Lemmas 2 and 3, respectively.

Definition 1. Contracts $x', x \in X$ are *vertically differentiated* (with x' preferred) if and only if $WTP(x', \theta) \geq WTP(x, \theta) \; \forall \; \theta \in \Theta$.

Definition 2. Given $x', x \in X$, contract x' is *higher coverage* than contract x if and only if $WTP(x', \theta) \geq WTP(x, \theta) \; \forall \; \theta \in \Theta$ and $\Psi(x', \theta) \geq \Psi(x, \theta) \; \forall \; \theta \in \Theta$. We denote this relationship by writing $x' \geq x$.

Proposition 1. *Contracts $x', x \in X$ are vertically differentiated (with x' preferred) if and only if $c_{x'}(m) \leq c_x(m) \; \forall \; m$.*

Proposition 2. *Given $x', x \in X$, contract x' is higher coverage than contract x if and only if $c_{x'}(m) \leq c_x(m) \; \forall \; m$ and $c'_{x'}(m) \leq c'_x(m)$ almost everywhere.*

Proposition 3. *Social surplus is single peaked in coverage level. That is, fixing $\theta \in \Theta$ and $x, x', x'' \in X$ where $x \leq x' \leq x''$: if $SS(x'', \theta) \geq SS(x', \theta)$, then $SS(x', \theta) \geq SS(x, \theta)$.*

Proof of Proposition 1: Contracts $x', x \in X$ are vertically differentiated (with x' preferred) if and only if $c_{x'}(m) \leq c_x(m) \; \forall \; m$.

Fix $\theta \in \Theta$. Let $Z_x =: z_x(l, \theta)$ be the random payoff induced by health state distribution F. At any health state l, lower out-of-pocket costs deliver higher payoffs:

$$\begin{aligned} Z_x = z_x(l, \theta) &= \hat{y} - p + b(m^*(l, \omega, x); l, \omega) - c_x(m^*(l, \omega, x)) \\ &\leq \hat{y} - p + b(m^*(l, \omega, x); l, \omega) - c_{x'}(m^*(l, \omega, x)) \\ &\leq \hat{y} - p + b(m^*(l, \omega, x'); l, \omega) - c_{x'}(m^*(l, \omega, x')) = z_{x'}(l, \theta) = Z_{x'} \; , \end{aligned}$$

where the second inequality holds by the optimality of $m^*(l, \omega, x')$. [$\Leftarrow$] $Z_{x'}$ therefore first-order stochastically dominates Z_x, and the result follows because u_ψ is increasing. [$\Rightarrow$] If $c_{x'}(\tilde{m}) > c_x(\tilde{m})$ for some $\tilde{m}$, the first inequality fails to hold for consumer type $\tilde{\omega}$ at health state realization $\tilde{l}$ at which $m^*(\tilde{l}, \tilde{\omega}, x) = \tilde{m}$. Such a consumer type exists for any $\tilde{m}$ we might choose because as ω approaches zero, privately optimal utilization approaches the health state, meaning any m can be approached arbitrarily closely. As c_x is continuous, if $c_{x'}(\tilde{m}) > c_x(\tilde{m})$, the same will be true in a neighborhood of $\tilde{m}$. A consumer with health state distribution $\tilde{F}$ degenerate on $\tilde{l}$ would strictly prefer contract x. By continuity, a consumer with a health state distribution that is sufficiently concentrated at $\tilde{l}$ would also prefer contract x. □

Proof of Proposition 2: Contract x' is higher coverage than contract x if and only if $c_{x'}(m) \leq c_x(m)\ \forall\ m$ and $c'_{x'}(m) \leq c'_x(m)$ almost everywhere.

By Proposition 1, $c_{x'}(m) \leq c_x(m)\ \forall\ m$ is necessary and sufficient for the contracts to be vertically differentiated. It remains to show that $c'_{x'}(m) \leq c'_x(m)$ almost everywhere is necessary and sufficient for $\Psi(x', \theta) \geq \Psi(x, \theta)$. Fix $\theta \in \Theta$. Let $\dot{Z}_x =: z_x(l, \theta) - \bar{z}_x(\theta)$ be the mean-zero random payoff induced by health state distribution F. Differentiating $\dot{Z}_x$ with respect to the health state realization l:

$$\begin{aligned} \frac{d\dot{Z}_x}{dl} &= \frac{\partial b}{\partial l}(m^*(l, \omega, x); l, \omega) \\ &\leq \frac{\partial b}{\partial l}(m^*(l, \omega, x'); l, \omega) = \frac{d\dot{Z}_{x'}}{dl} \\ &\leq 0. \end{aligned}$$

That is, the payoff is weakly decreasing in the health state, and is doing so faster for contract x than for contract x'. The first equality holds by the envelope theorem. Because $\frac{\partial^2 b}{\partial l \partial m} = \omega^{-1} \geq 0$, the first inequality holds as long as $m^*(l, \omega, x)$ is increasing in x. The second inequality holds because $\frac{\partial b}{\partial l} = \omega^{-1}(m^*(l, \omega, x) - l) - 1 \leq 0$, or in other words, because moral hazard spending never exceeds ω.[81] [$\Leftarrow$] Lemma 1 shows that $m^*(l, \omega, x)$ is increasing in x as long as $c'_{x'}(m) \leq c'_x(m)$. $\dot{Z}_x$ is therefore a mean preserving spread of $\dot{Z}_{x'}$, and would be preferred by any risk-averse expected utility maximizer: $\mathbb{E}_{l \sim F}[u_\psi(\dot{Z}_{x'})] \geq \mathbb{E}_{l \sim F}[u_\psi(\dot{Z}_x)]$. The result follows because $-\psi^{-1} \log(-x)$ is increasing. [$\Rightarrow$] If $c_{x'}(\tilde{m}) > c_x(\tilde{m})$ for some $\tilde{m}$, the first inequality fails to hold for consumer type $\tilde{\omega}$ at health state realization $\tilde{l}$ at which $m^*(\tilde{l}, \tilde{\omega}, x) = \tilde{m}$. Such a consumer type exists for any $\tilde{m}$ we might choose because as ω approaches zero, privately optimal utilization approaches the health state, meaning any m can be approached arbitrarily closely. As c_x is continuous, if $c_{x'}(\tilde{m}) > c_x(\tilde{m})$, the same will be true in a neighborhood of $\tilde{m}$. At $\tilde{l}$, the payoff would therefore be decreasing faster in the health state under contract x' than under contract x, and x would provide strictly more risk protection to a consumer with health state distribution $\tilde{F}$ sufficiently concentrated around $\tilde{l}$. □

Proof of Proposition 3. Social surplus is single peaked in coverage level. That is, fixing $\theta \in \Theta$ and $x, x', x'' \in X$ where $x \leq x' \leq x''$: if $SS(x'', \theta) \geq SS(x', \theta)$, then $SS(x', \theta) \geq SS(x, \theta)$.

Let $\tilde{c}_x(l) = c_x(m^*(l, \omega, x))$ be the indirect out-of-pocket cost function for consumer type

[81] Note that this statement would not be true under the "multiplicative" specification of preferences proposed by Einav et al. (2013) and used in Ho and Lee (2021). In that case, $\frac{\partial b}{\partial l}$ becomes positive at a certain health state level, and the payoff $z_x(l, \theta)$ begins increasing in the health state. The conditions given in Proposition 2 would therefore not be sufficient to guarantee coverage level ordering in that context.

ω under contract x.[82] As θ is fixed throughout the proof, we omit ω as an argument in $\tilde{c}_x(l)$. Similarly, let $\tilde{c}'_x(l) = c'_x(m^*(l, \omega, x))$ be the indirect marginal out-of-pocket cost function. Note that because $m^*(l, \omega, x)$ is increasing in x (see Lemma 1) and contracts are concave, $\tilde{c}'_{x''}(l) \leq \tilde{c}'_{x'}(l) \leq \tilde{c}'_x(l)$ wherever these derivatives exist.

Next, for each contract $x \in \{x, x', x''\}$, calculate the cutoff values of the health state l that determine which segment of the piecewise linear out-of-pocket cost function the consumer of type θ will choose. Appendix C.2 describes this procedure and provides formulas for the cutoffs. As the contracts we consider have at most three segments, each contract has at most three cutoffs: one at which positive healthcare utilization begins and two separating the segments of the out-of-pocket cost function.[83] Considering the three cutoff values of our three candidate contracts simultaneously, the space of health states (the real line) is divided into at most 10 regions. Denote these regions by $\{R_r\}_{r=1}^{10}$, where $R_r = (l_r^{lb}, l_r^{ub})$ and $l_r^{ub} = l_{r+1}^{lb}$.[84] The lower bound of the first region is $-\infty$ and the upper bound of the final region is ∞. For each contract x in each region R_r, out-of-pocket costs are linear in the health state, and so can be written $\tilde{c}_{x,r}(l) = \gamma_{x,r} + l\,\tilde{c}'_{x,r}$, with intercept $\gamma_{x,r}$ and slope $\tilde{c}'_{x,r}$. As before, higher coverage contracts are flatter: $c'_{x'',r} \leq c'_{x',r} \leq c'_{x,r}\ \forall\ r$.

Extend this notation to the consumer's payoff $z_x(l, \theta)$. Omitting θ, the payoff in region r under contract x can now be written:

$$\begin{aligned} z_x(l) &= \hat{y} - p_x + \tfrac{\omega}{2}(1 - \tilde{c}'_x(l)^2) - \tilde{c}_x(l) \\ &= \hat{y} - p_x + \tfrac{\omega}{2}(1 - \tilde{c}'^{\,2}_{x,r}) - \gamma_{x,r} - \tilde{c}'_{x,r}\,l, \quad l \in R_r. \end{aligned}$$

The payoff is linear in the health state with slope and intercept determined by the relevant segment of the indirect out-of-pocket cost function. To isolate the effects of level from the effects of slope, it is useful to express the payoff in terms of differences from its mean in a given region. To this end, write:

$$z_x(l) = \bar{z}_{x,r} - \tilde{c}'_{x,r}(l - \bar{l}_r), \quad l \in R_r$$

where $\bar{l}_r = \mathbb{E}_{l|R_r}[l]$ is the conditional expectation of the health state in region r with respect to the consumer's health state distribution F, and $\bar{z}_{x,r} = z_x(\bar{l}_r)$ is the conditional expectation of the payoff. Note that because higher coverage contracts deliver everywhere higher payoffs (see proof of Proposition 1): $\bar{z}_{x'',r} \geq \bar{z}_{x',r} \geq \bar{z}_{x,r}\ \forall\ r$. Each contract is now fully characterized by the payoff function it generates, which in turn is fully described by its mean and slope in

[82]The line labelled c^* in Figure A.2 represents the function $\tilde{c}_x(l)$ in that example.

[83]The proof extends trivially to piece-wise linear out-of-pocket functions with a different number of segments.

[84]As we have assumed F is continuously distributed, there is zero mass on region boundaries.

each region: $\{\bar{z}_{x,r}, \tilde{c}'_{x,r}\}_{r=1}^{10}$. Higher coverage contracts generate both higher and flatter payoffs in every region. Expressing the payoff function in this way allows us think about changing a contract's slope while holding its expected payoff fixed, and vice versa.

We now proceed in two steps. We first show that the social cost of moral hazard $\textsc{scmh}(x, \theta)$ is increasing and "convex" in coverage level. As coverage level itself has no cardinal interpretation, the idea of convexity is applicable with respect to the slope of contracts' indirect out-of-pocket cost functions $\tilde{c}'_{x,r}$. We then show that the value of risk protection $\Psi(x, \theta)$ is increasing and "concave" in coverage level, where the idea of concavity is again applicable with respect to $\tilde{c}'_{x,r}$. Note that the tradeoff between risk protection and moral hazard operates entirely through the slope of the out-of-pocket cost function. The level of out-of-pocket costs impacts only the value of risk protection, and does so monotonically. As $SS(x, \theta) = \Psi(x, \theta) - \textsc{scmh}(x, \theta)$, these two steps imply $SS(x, \theta)$ is itself concave in the slope of the out-of-pocket function. Single-peakedness in coverage level follows from the fact that this slope is monotonic in coverage level.

1. $\textsc{scmh}(x, \theta)$ is increasing and "convex" in coverage level.

First, split the expectation between the defined regions, omitting θ as an argument:

$$\begin{aligned} \textsc{scmh}(x) &= \mathop{\mathbb{E}}_{l \sim F}\left[\tfrac{\omega}{2}(1 - \tilde{c}'_x(l))^2\right] \\ &= \textstyle\sum_{r=1}^{10} \pi_r \left[\tfrac{\omega}{2}(1 - \tilde{c}'_{x,r})^2\right], \end{aligned}$$

where $\pi_r = Pr(l \in R_r | l \sim F)$ is the probability of realizing a health state in region R_r. Taking the derivative with respect to the slope of the indirect out-of-pocket cost function in a given region:

$$\frac{d\textsc{scmh}(x)}{d\tilde{c}'_{x,r}} = -\pi_r \omega(1 - \tilde{c}'_{x,r}) \leq 0.$$

As $\textsc{scmh}(x)$ is decreasing in $\tilde{c}'_{x,r}$ in any region, it is increasing in coverage level. Taking the second derivative:

$$\frac{d^2\textsc{scmh}(x)}{d\tilde{c}'^{\,2}_{x,r}} = \pi_r \omega \geq 0.$$

The social cost of moral hazard is therefore increasing in the slope of the indirect out-of-pocket cost function $\tilde{c}'_{x,r}$ at an increasing rate. It is unaffected by changes in $\bar{z}_{x,r}$.

2. $\Psi(x, \theta)$ is increasing and "concave" in coverage level.

First, split the expectation between the defined regions, omitting θ as an argument:

$$\begin{aligned}\Psi(x) &= RP(x_0) - \psi^{-1}\log\left(\mathbb{E}_l\left[\exp\left(-\psi(z_x(l)-\bar{z}_x)\right)\right]\right)\\ &= RP(x_0) - \psi^{-1}\log\left(\textstyle\sum_{r=1}^{10}\pi_r\mathbb{E}_{l|R_r}\left[\exp\left(-\psi(z_x(l)-\bar{z}_x)\right)\right]\right),\end{aligned}$$

where $\pi_r = Pr(l \in R_r | l \sim F)$ is the probability of realizing a health state in region R_r. Taking the derivative with respect to the slope of the indirect out-of-pocket cost function in a given region:

$$\frac{d\Psi(x)}{d\tilde{c}'_{x,r}} = \left(\mathbb{E}_l[\exp(-\psi z_x(l))]\right)^{-1}\pi_r\mathbb{E}_{l|R_r}[\exp(-\psi z_x(l))(\bar{l}_r - l)] \leq 0.$$

Because the function $\exp(-\psi x)$ is convex and the payoffs $z_x(l)$ are decreasing in the health state, worse-than-average health states ($l \geq \bar{l}_r$) receive more weight than better-than-average health states ($l \leq \bar{l}_r$), and the expression is nonpositive. Taking the second derivative:

$$\frac{d^2\Psi(x)}{d\tilde{c}'_{x,r}{}^2} = \psi\left[\left(\frac{\pi_r\,\mathbb{E}_{l|R_r}[\exp(-\psi z_x(l))(\bar{l}_r - l)]}{\mathbb{E}_l[\exp(-\psi z_x(l))]}\right)^2 - \left(\frac{\pi_r\,\mathbb{E}_{l|R_r}[\exp(-\psi z_x(l))(\bar{l}_r - l)^2]}{\mathbb{E}_l[\exp(-\psi z_x(l))]}\right)\right] \leq 0.$$

The first term is the squared conditional expectation of $(\bar{l}_r - l)$. The second term is the conditional expectation of $(\bar{l}_r - l)^2$. Because x^2 is convex, the expression is nonpositive by Jensen's inequality.

□

Lemma 1. *Healthcare utilization is increasing in coverage level.*

Proof. Fix $l \in \mathbb{R}$, $\omega \in \mathbb{R}_{++}$, and $x, x' \in X$ where $x \leq x'$. Optimal utilization $m^*(l,\omega,x) = \text{argmax}_m\,(b(m;l,\omega) - c_x(m))$. Consider $m, m' \in \mathbb{R}_+$ where $m \leq m'$:

$$\begin{aligned}b(m';l,\omega) - c_{x'}(m') - [b(m';l,\omega) - c_x(m')] &= c_x(m') - c_{x'}(m')\\ &\geq c_x(m) - c_{x'}(m)\\ &= b(m;l,\omega) - c_{x'}(m) - [b(m;l,\omega) - c_x(m)]\,,\end{aligned}$$

where the inequality holds because $c_{x'}(m) \leq c_x(m)$ and $c'_{x'}(m) \leq c'_x(m)$ guarantees c is submodular in m and x. The objective $b(m;l,\omega) - c_x(m)$ is therefore supermodular and standard monotone comparative statics imply $m^*(l,\omega,x)$ is increasing in x.

□

Lemma 2. *Efficient coverage level is increasing in risk aversion.*

Proof. Fix $\theta \in \Theta$. Efficient coverage level $x^{eff} = \text{argmax}_x(RP(x_0,F,\omega,\psi) - RP(x,F,\omega,\psi) - \textsc{scmh}(x,F,\omega))$. As the insurer is risk-neutral, the social cost of moral hazard is unaffected by

ψ. Differentiating $RP(x, F, \omega, \psi)$ with respect to ψ:

$$\frac{dRP(x,\theta)}{d\psi} = -\psi^{-1}\left[RP(x,\theta) + (\mathop{\mathbb{E}}_{l\sim F}[\exp(-\psi\dot{Z}_x)])^{-1}\mathop{\mathbb{E}}_{l\sim F}[\exp(-\psi\dot{Z}_x)\dot{Z}_x]\right],$$

where $\dot{Z}_x =: z_x(l,\theta) - \bar{z}_x(\theta)$. The first term in the brackets, $RP(x,\theta)$, is shown to be decreasing in x in Proposition 2. The second term represents a weighted average of deviations from mean payoffs, where the weights correspond to the utility weight at that realization. As $\dot{Z}_x$ becomes less risky as x increases (see proof of Proposition 2), this term is also decreasing in x. $\frac{dSS(x,\theta)}{d\psi}$ is therefore increasing in x, and standard monotone comparative statics imply x^{eff} is increasing in ψ.

□

Lemma 3. *Efficient coverage level is decreasing in the moral hazard parameter.*

Proof. Fix $\theta \in \Theta$. Efficient coverage level $x^{eff} = \text{argmax}_x\,(\Psi(x,\theta) - \textsc{scmh}(x,\theta))$, where $\textsc{scmh}(x,\theta) = \mathbb{E}_{l\sim F}[\frac{\omega}{2}(1 - c'_x(m^*(l,\omega,x)))^2]$. Differentiating $\textsc{scmh}(x,\theta)$ with respect to ω:

$$\frac{d\textsc{scmh}(x,\theta)}{d\omega} = \mathop{\mathbb{E}}_{l\sim F}[\tfrac{1}{2}(1 - c'_x(m^*(l,\omega,x)))^2] \leq 0.$$

Note that contracts are piecewise linear and $c'_x \in [0,1]$. Because $m^*(l,\omega,x)$ is increasing in x (see Lemma 1) and contracts are concave, $c'_x(m^*(l,\omega,x))$ is decreasing in x and $\frac{dSCMH(x,\theta)}{d\omega}$ is increasing in x. $\frac{dSS(x,\theta)}{d\omega}$ is therefore decreasing in x, and standard monotone comparative statics imply x^{eff} is decreasing in ω.

□

Appendix B Additional Analysis

B.1 Estimation of Plan Cost-sharing Features

A crucial input to our empirical model is the cost-sharing function of each plan. While Table 1 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 huge simplification that healthcare is a homogenous good over which the consumer chooses only the quantity to consume. We then 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 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.[85] We carry out this procedure separately for each plan, year, and family status (individual or family).[86] Figure A.1 shows an example of the data and estimates for a particular plan: Moda - 3 for individual households in 2012. Table A.3 presents the estimated cost-sharing features for all plans in all years.

B.2 Variation in Plan Menu Generosity

Measuring Plan Menu Generosity. We want to measure the likelihood that a household would choose generous health insurance coverage when presented with a particular plan menu. We refer to this measure as "plan menu generosity." At a simple level, if plan menus consisted of only a single plan, 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

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

[86] 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.

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:

$$plan_{jk} = \underset{j \in \mathcal{J}_d}{\operatorname{argmax}}(\alpha p_{jd} + \alpha^{VD} p_{jd}^{VD} + \alpha^{HA} p_{jd}^{HA} + \nu_j + \epsilon_{jk}), \tag{12}$$

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.6, 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 12 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 plan j is chosen as ρ_{jd}.[87] 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 given by:

$$\begin{aligned} \rho_{fd} &= \sum_{j \in \mathcal{J}_d^f} \rho_{jd}, \\ \widehat{AV}_{fd} &= \sum_{j \in \mathcal{J}_d^f} (\frac{\rho_{jd}}{\rho_{fd}}) AV_j, \end{aligned} \tag{13}$$

where $\mathcal{J}_d^f$ is the set of plans in $\mathbf{menu}_d$ offered by insurer f.

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

Explaining Plan Menu Generosity. Because the majority of the variation in coverage level lies within Moda, we focus on explaining plan menu generosity using the predicted actuarial value among Moda plans. We first compare plan menu generosity to observed household health (see 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 find that plan menus are consistently most generous for single employee coverage and least generous for employee plus family coverage. This pattern is consistent with our understanding of OEBB's benefit structure, and is common in employer-sponsored health insurance.

We further explore which covariates, in addition to family structure, can 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."[88] 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.[89] *Pct. Republican* measures the percent of households in a school district that are registered as Republicans as of 2016.[90]

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 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.

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

[89] 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.

[90] 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.

B.3 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.[91] 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 (14) and (15):

$$plan_k = f(\mathbf{menu}_d, \mathbf{X}_k, \xi_k), \tag{14}$$

$$y_k = g(plan_k, \mathbf{X}_k, \xi_k). \tag{15}$$

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 15 according to

$$\log(y_k) = \delta_f \mathbf{1}_{f(k)=f} + \gamma \log(1 - AV_{j(k)})\mathbf{1}_{f(k)=Moda} + \beta \mathbf{X}_k + \xi_k, \tag{16}$$

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.[92] Our parameter of interest is γ, which represents the responsiveness

[91] 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.

[92] 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.

of total spending to plan generosity, holding the insurer fixed (at Moda).[93] 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.[94]

We estimate equation 16 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 B.2. 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.7 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.7 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 12 for only that subsample of households.[95] 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 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

[93] 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.

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

[95] Estimates for each subsample are presented in Table A.8.

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.9 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.9 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.7 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 most closely spans the range over which there would be marginal out-of-pocket price variation across plans. Likewise for family households, those in the fourth quartile are nearly all above their 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 C Estimation Details

C.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.[96] 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:

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

$$Var(\tilde{l} + \kappa_k) = \sum_{i \in \mathcal{I}_k} Var(\tilde{l}^i + \kappa_i), \tag{18}$$

$$E(\tilde{l}) = \sum_{i \in \mathcal{I}_k} E(\tilde{l}^i), \tag{19}$$

where $\mathcal{I}_k$ is the set of individuals in household k. Equation 17 sets the mean of the household's distribution equal to the sum of the means of each individual's distribution. Equation 18 matches the variance. Because we have a third parameter to estimate (the shift, κ_k), we use a third moment-matching condition to match the first moment of the unshifted distribution, shown in equation 19.

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:

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

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

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.

C.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.[97,98] 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

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

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.

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

[98]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).

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].[99]

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 privately optimal healthcare spending choice m_{kjtsd} associated with each potential health state realization.

Plans in our empirical setting are characterized by a deductible D, a coinsurance rate C, and an out-of-pocket maximum O. Marginal out-of-pocket costs $c'(m)$ equal 1 in the deductible region, c in the coinsurance region, and 0 in the out-of-pocket maximum region. Denote the boundary between the coinsurance region and the out-of-pocket maximum region (the "stop loss" level of total spending) by $A = C^{-1}(O - D(1 - C))$. Privately optimal spending falls into one of these three regions depending on the realization of the health state l and the moral hazard parameter ω. The relevant cutoff values for the health state are

[99] 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.

$$Z_1 = D - \omega(1-C)/2,$$
$$Z_2 = O - \omega/2,$$
$$Z_3 = A - \omega(1 - C/2),$$

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:

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

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

A graphical example (of the case in which the coinsurance region is sufficiently large) is shown in Figure A.2b. 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:

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

where the monetary payoff z 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}:

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

where $\bar{z}_{kjts} = \mathbb{E}_d[z_{kjts}(l_{kjtsd})]$. 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," in the sense 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

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

C.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

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

Joint Distribution of Household Types. We investigate the distribution implied by our primary estimates in column 3 of Tables 3 and A.10. 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.3 presents the joint distribution of household types along the dimensions of risk aversion, moral hazard parameter, and expected health state. We measure the expected health state on a log scale for readability.

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: The 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. This table is referenced at footnote 30.

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	<0.01
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.

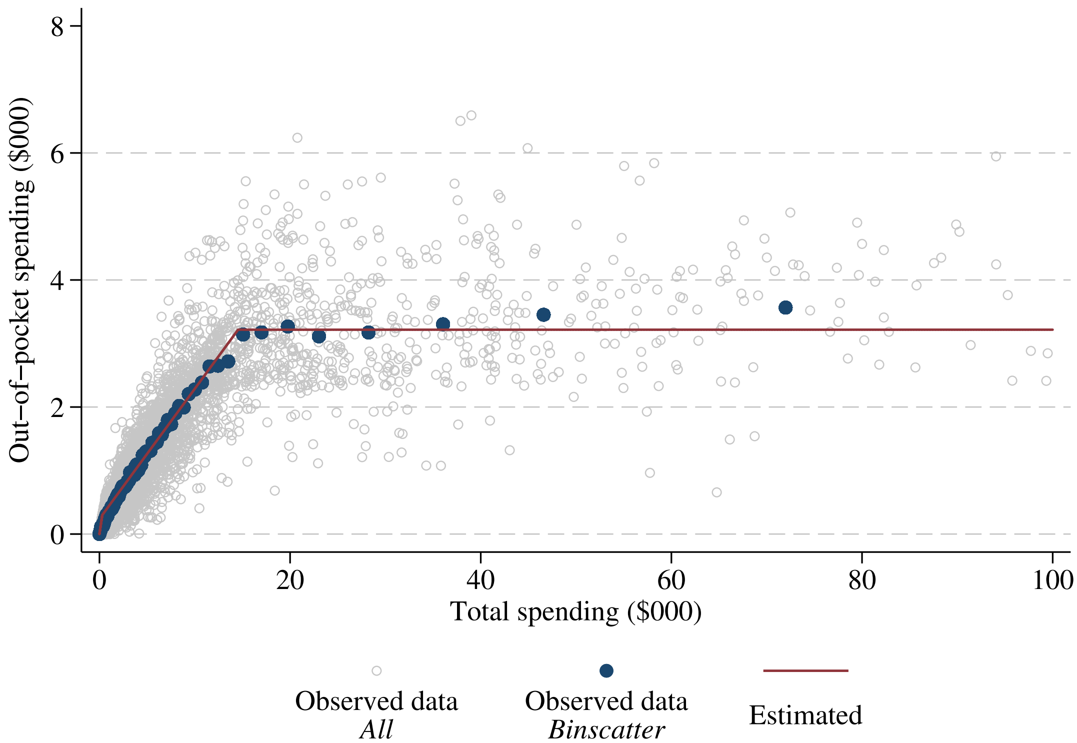
2011 Plan	Actuarial Value	Avg. Employee Premium ($)	Full Premium ($)	Deductible ($)	OOP Max. ($)	Market Share
Kaiser - 1	0.95	520	11,051	0	2,400	0.16
Kaiser - 2	0.92	348	10,126	300	4,000	0.04
Moda - 1	0.86	3,414	15,622	600	4,500	0.06
Moda - 2	0.84	1,009	12,391	900	6,000	<0.01
Moda - 3	0.84	1,208	12,688	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	<0.01

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: The table shows the state-level master lists of plans available in 2008 and 2010–2013. Actuarial value is the ratio of the sum of insured spending across all households to the sum of total spending across all households. 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 out-of-pocket maximum shown are for in-network services for a family household. This table is referenced at footnote 29.

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. Total healthcare spending is on the horizontal axis and out-of-pocket cost is on the vertical axis. Each gray dot represents a household, for a 20 percent random sample of households. 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. We estimate a best-fit cost-sharing function by finding the coinsurance rate and out-of-pocket maximum that 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. This figure is referenced in Section B.1.

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

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 B.1. This table is referenced in Section B.1.

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 table shows the relationship between plan menu generosity and household health. The unit of observation is the household. The dependent variable is plan menu generosity, as measured by predicted actuarial value conditional on choosing Moda, $\widehat{AV}_{d,Moda}$. This measure is calculated according to equation 13, and it is multiplied by 100 to increase the magnitude of estimated coefficients on household risk score. Household risk score is the mean risk score among all individuals in a household, and it 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. †By normalization. This table is referenced in Section B.2.

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	$0.000^\dagger$	$0.000^\dagger$	$0.000^\dagger$	$0.000^\dagger$
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				
log(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 table shows the relationship between plan menu generosity and household/employee characteristics. The unit of observation is the household. The dependent variable is plan menu generosity, as measured by predicted actuarial value conditional on choosing Moda, $\widehat{AV}_{d,Moda}$. This measure is calculated according to equation 13, and it is multiplied by 100 to increase the magnitude of estimated coefficients on household risk score. Household risk score is the mean risk score among all individuals in a household, and it 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. †By normalization. This table is referenced in Section B.2.

Table A.6. 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: The table presents parameter estimates from the conditional logit model described by equation 12, presented separately for each year. The unit of observation is a household-plan. Moda - 1 (the highest coverage Moda plan) is the omitted plan. †By normalization. This table is referenced in Section B.2.

Table A.7. 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
Number of observations	35,146	35,146	8,962	26,184

Notes: The table shows the OLS and IV estimates of equation 16, 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, where 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. This table is referenced in Section B.3.

Table A.8. 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 observations	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 12 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. Moda - 1 (the highest coverage Moda plan) is the omitted plan. The coefficient on employee premium (measured in thousands of dollars) is normalized to -1. This table is referenced in Section B.3. †By normalization.

Table A.9. Spending Distributions and Moda Plan Characteristics, 2008

(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

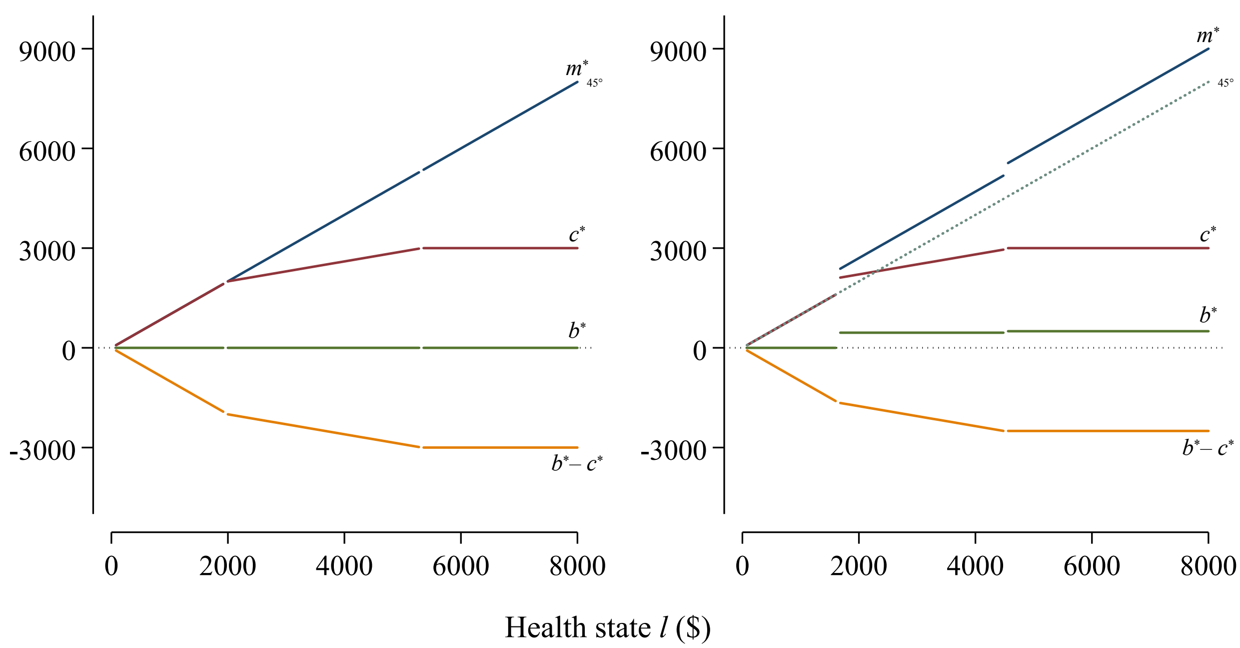
(b) Plan Characteristics

	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. This table is referenced in Section B.3.

Figure A.2. Healthcare Spending Choice Example

(a) No Moral Hazard ($\omega \approx \$0$) (b) Some Moral Hazard ($\omega = \$1{,}000$)



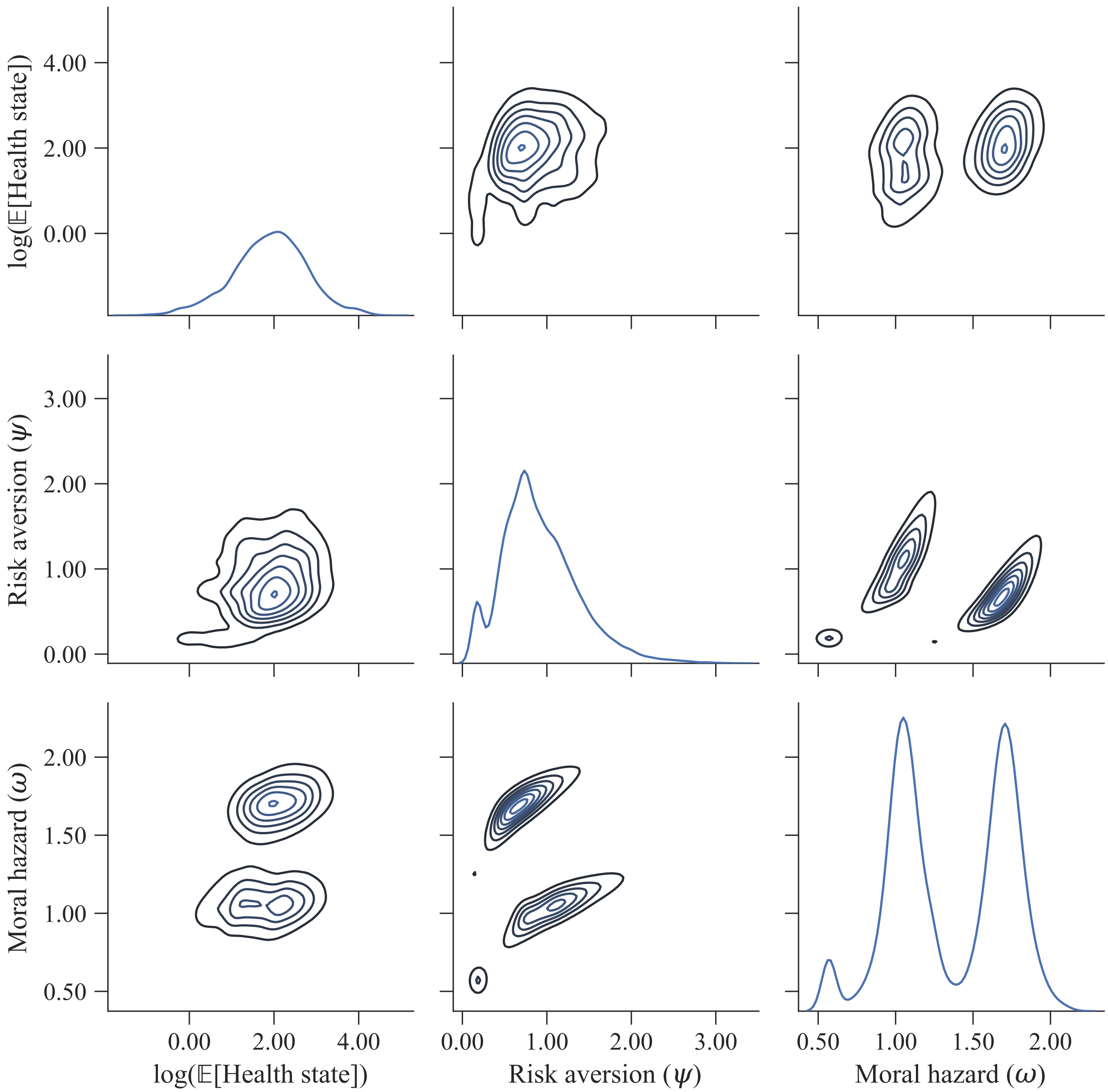
Notes: The figure shows optimal healthcare spending m^*, indirect benefit of optimal healthcare spending b^*, and the corresponding out-of-pocket cost c^* predicted by our parameterization of consumer preferences (equation 4). The examples consider a contract with a deductible of \$2,000, a coinsurance rate of 30 percent, 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 l. Absent moral hazard (left panel), optimal healthcare spending is equal to the health state. The vertical axis also shows the net payoff from optimal healthcare utilization, $b^* - c^*$; this is the outcome over which households face a lottery. This figure is referenced at footnotes 37 and 82.

Table A.10. Additional Parameter Estimates

Variable	(1) Parameter	(1) Std. Err.	(2) Parameter	(2) Std. Err.	(3) Parameter	(3) Std. Err.
Insurer fixed effects						
Providence * (Age−40)	−0.024	0.007	−0.023	0.007	−0.028	0.007
Providence * $\mathbf{1}$[Children]	−0.681	0.151	−0.501	0.146	−0.595	0.147
Providence * Region 1	−2.114	0.144	−2.071	0.137	−1.649	0.138
Providence * Region 2	−2.658	0.185	−2.635	0.176	−2.179	0.176
Providence * Region 3	−1.877	0.207	−2.036	0.200	−1.409	0.193
Health state distributions						
κ	0.155	0.002				
κ * Risk Q_1			0.096	0.002	0.127	0.000
κ * Risk Q_2			0.224	0.002	0.155	0.001
κ * Risk Q_3			0.218	0.002	0.228	0.000
κ * Risk Q_4			0.128	0.042	0.418	0.041
κ * Risk Q_1 * Risk score			0.187	0.004	0.225	0.001
κ * Risk Q_2 * Risk score			0.140	0.002	0.019	0.002
κ * Risk Q_3 * Risk score			−0.060	0.001	0.002	0.001
κ * Risk Q_4 * Risk score			0.155	0.026	0.177	0.027
μ	0.590	0.005				
μ * $\mathbf{1}$[Female 18–35]			0.125	0.017	0.088	0.018
μ * $\mathbf{1}$[Age $<$ 18]			−0.113	0.017	−0.104	0.019
μ * Risk Q_1			1.405	0.137	1.872	0.154
μ * Risk Q_2			0.894	0.025	0.457	0.030
μ * Risk Q_3			0.815	0.008	0.504	0.009
μ * Risk Q_4			1.379	0.017	1.303	0.017
μ * Risk Q_1 * Risk score			3.590	0.185	4.875	0.210
μ * Risk Q_2 * Risk score			1.978	0.067	1.946	0.081
μ * Risk Q_3 * Risk score			0.894	0.019	1.053	0.022
μ * Risk Q_4 * Risk score			0.310	0.005	0.329	0.005
σ	1.174	0.002				
σ * Risk Q_1			1.626	0.006	1.748	0.007
σ * Risk Q_2			1.173	0.005	1.403	0.006
σ * Risk Q_3			1.060	0.003	1.215	0.004
σ * Risk Q_4			0.988	0.006	1.016	0.006

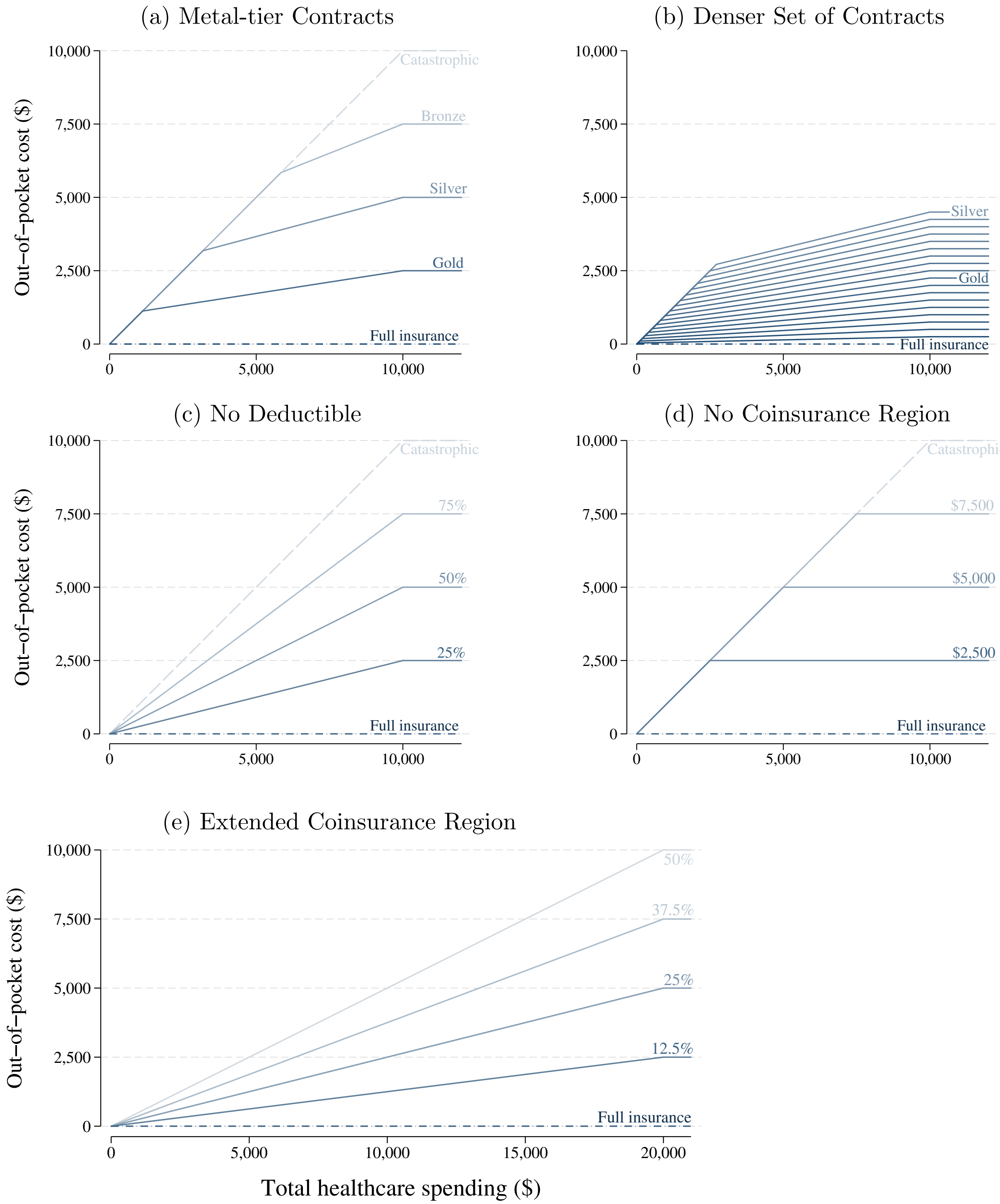
Notes: The table presents the parameter estimates that were not presented in Table 3. "Risk Q_n" is an indicator for an individual's risk quartile, where Q_4 is the sickest individuals. Higher risk scores correspond to worse predicted health. All parameters are measured in thousands of dollars. The insurer fixed effect of Moda is normalized to zero. This table is referenced in Section V.A.

Figure A.3. Joint Distribution of Household Types



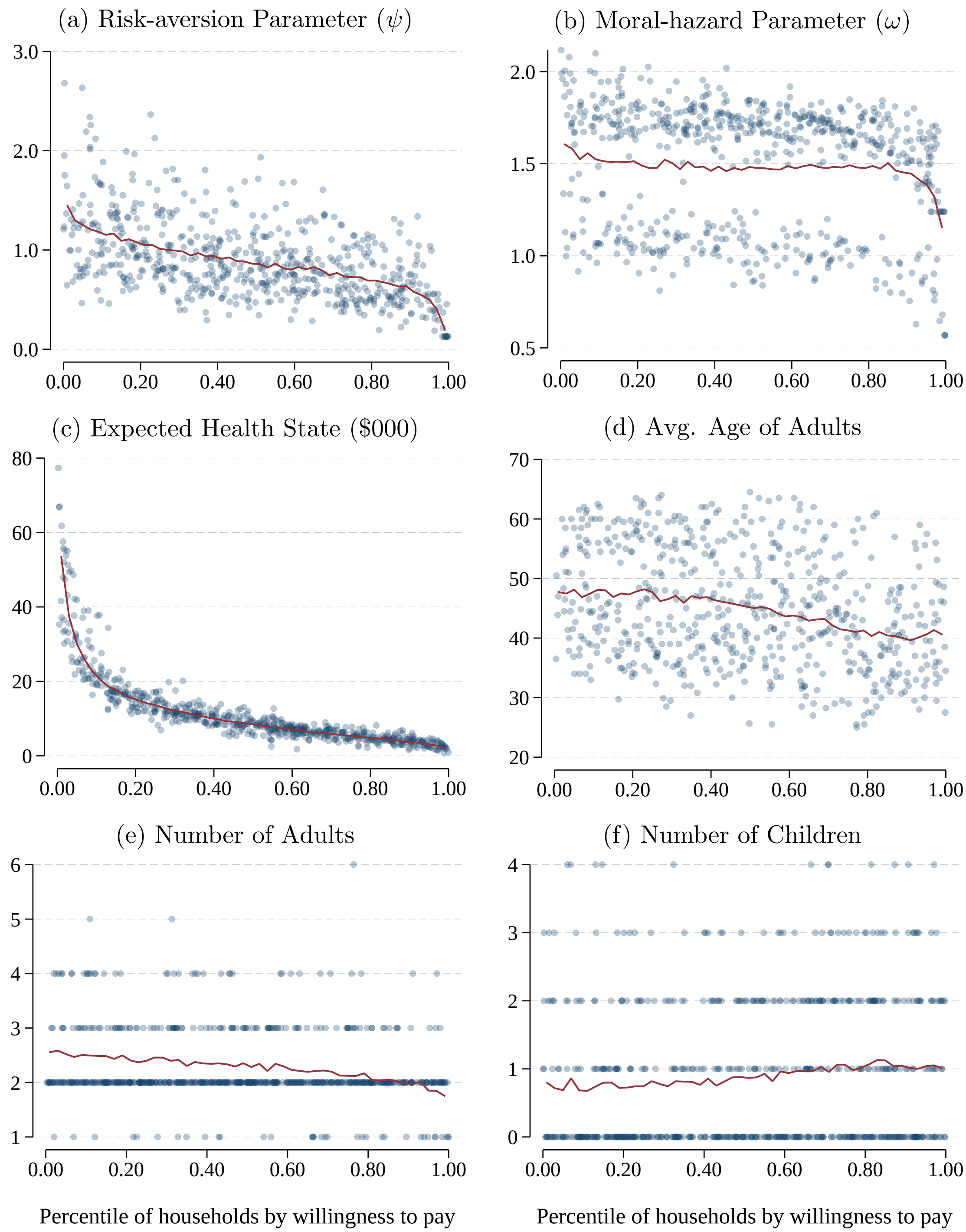
Notes: The figure shows the joint distribution of household types implied by parameter estimates in column 3 of Tables 3 and A.10. The diagonals show one-way distributions across households, and the off-diagonals show bivariate distributions. Households are ex post assigned a single type according to the procedure described in Section C.3. Because expected health state can vary over years within a household, this figure uses the first year a household appears in the sample. Expected health state is equivalent to a household's expected total spending absent moral hazard. This figure is referenced in Section V.A.

Figure A.4. Sets of Potential Contracts: Out-of-pocket Cost Functions



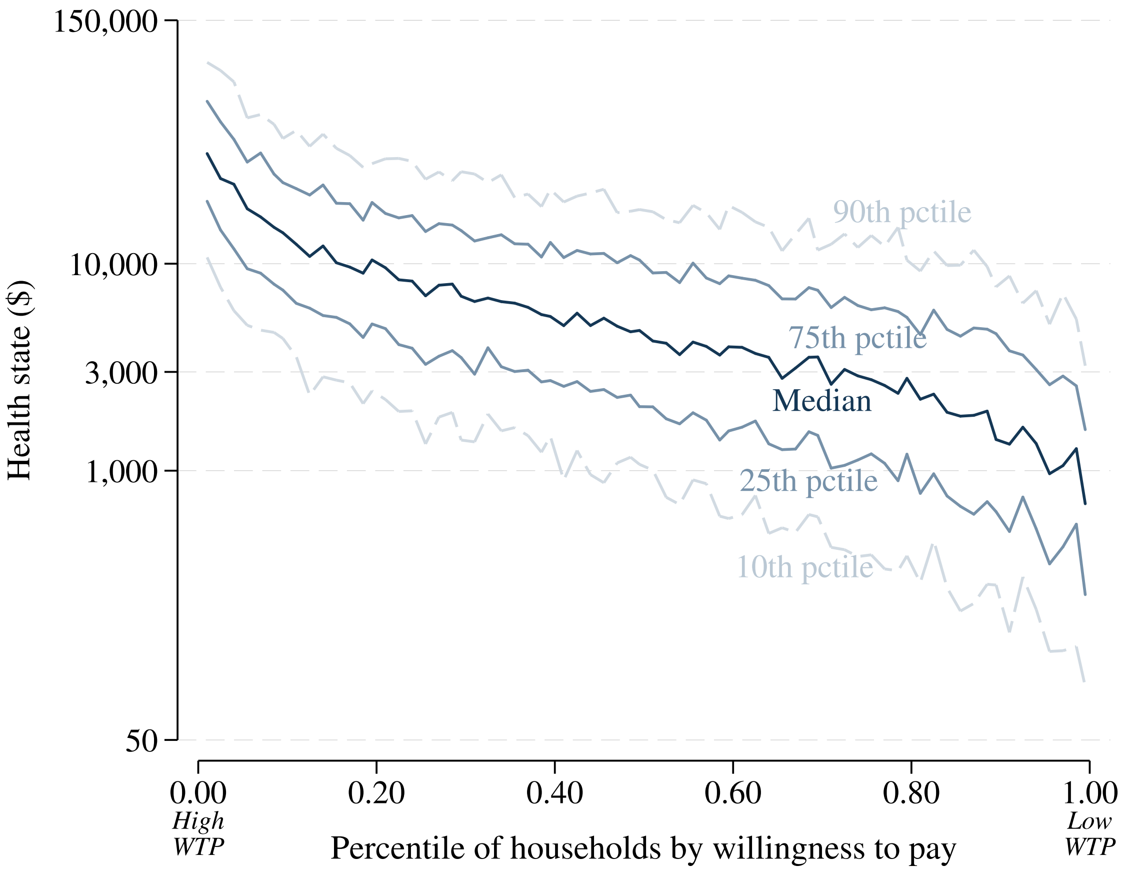
Notes: The figure shows out-of-pocket cost functions for five sets of potential contracts. Horizontal axes shows total healthcare spending, and vertical axes shows out-of-pocket cost. Panel (a) depicts our focal set of metal-tier contracts; panel (b) depicts a denser set of contracts with the same design. Panels (c)–(e) show alternative sets of potential contracts. Contract labels represent the varying feature: the coinsurance rate in panels (c) and (e) and the deductible in panel (d). Contracts are vertically differentiated and well-ordered by coverage level within each panel, but not necessarily across panels. See Appendix A.2 for these definitions. This figure is referenced in Sections V.B and V.C.

Figure A.5. Household Demographics by Willingness to Pay



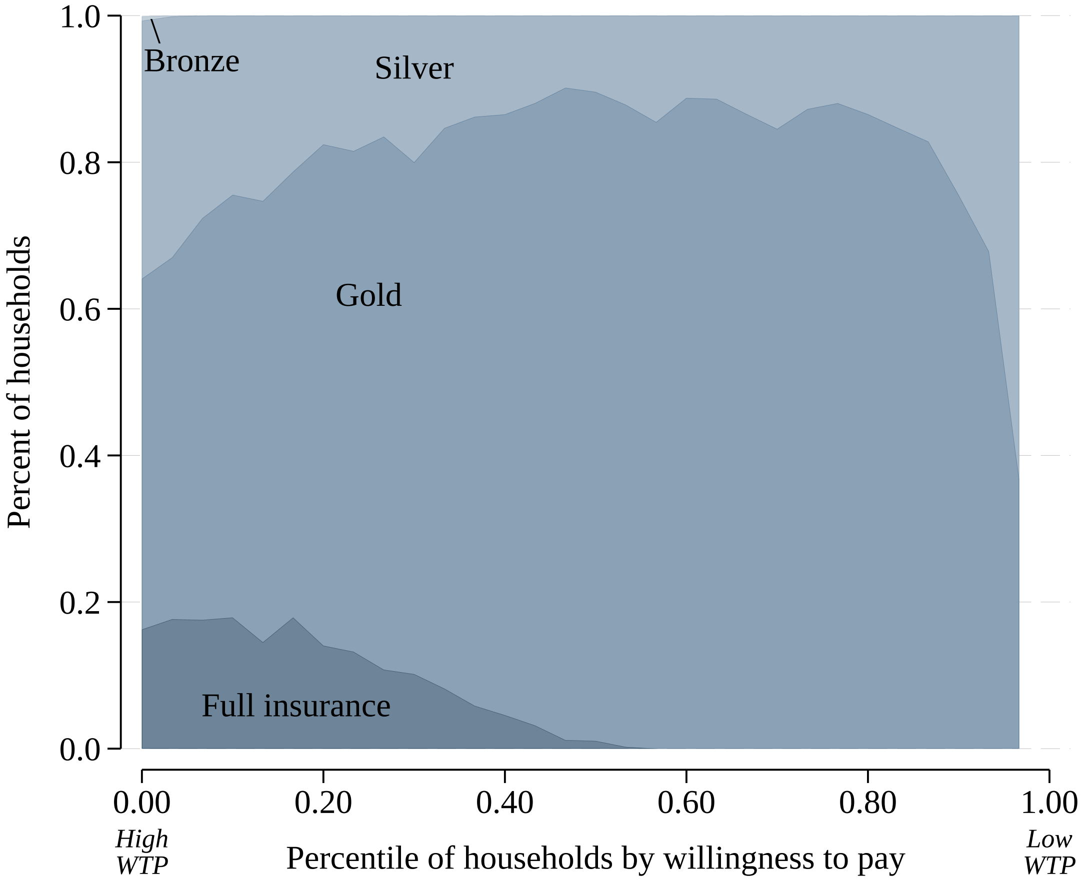
Notes: The figure shows the distribution across family households of (a) the risk aversion parameter, (b) the moral hazard parameter, (c) the expectation of the health state distribution, (d) the average age of adults in the household, (e) the number of adults in the household, and (f) the number of children in the household. An adult is defined as anyone 18 and older. Each dot represents a household, for a 2.5 percent random sample of households. The line in each panel is a connected binned scatter plot, representing the mean value of the vertical axis variable at each percentile of willingness to pay. This figure is referenced in footnote 60.

Figure A.6. Household Health State Distributions by Willingness to Pay



Notes: The figure shows the health state distributions faced by 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. This figure is referenced in Section V.B.

Figure A.7. Efficient Coverage Level by Willingness to Pay



Notes: The figure shows the percentage of family households at each percentile of willingness to pay for whom each contract is optimal. Households are ordered on the horizontal axis according to their willingness to pay. Overall, full insurance is efficient for 6 percent of households, Gold for 75 percent of households, Silver for 19 percent of households, and Bronze for less than one percent of households. Coverage lower than Bronze is not efficient for any household. This figure is referenced in Sections V.B and VI.A.

Table A.11. Outcomes Under Alternative Sets of Potential Contracts

Allocation at First Best (*FB*) and under the Optimal Menu (*Opt*)

	Metal-tier Contracts					No Deductible				
	Full	Gold	Silv.	Brnz.	Ctstr.	Full	25%	50%	75%	Ctstr.
FB:	0.06	0.75	0.19	<0.01	–	0.31	0.65	0.03	<0.01	–
Opt:	–	1.00	–	–	–	–	1.00	–	–	–

	No Coinsurance Region					Extended Coins. Region				
	Full	\$2.5k	\$5.0k	\$7.5k	Ctstr.	Full	12.5%	25%	37.5%	50%
FB:	–	0.82	0.17	0.01	–	0.66	0.31	0.01	0.01	–
Opt:	–	1.00	–	–	–	0.82	0.16	0.02	–	–

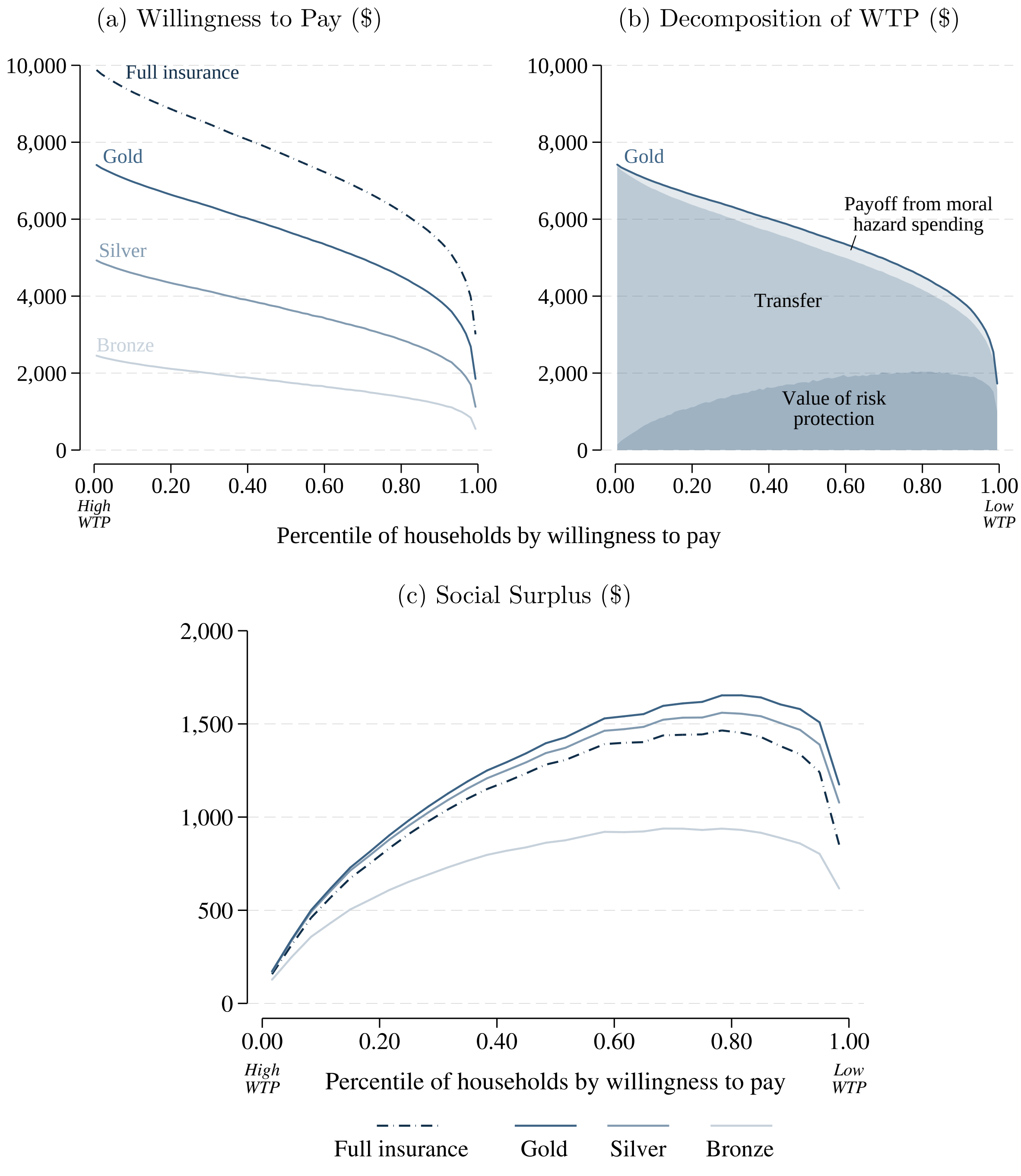
Notes: The table shows the percent of households allocated to each contract at the first best allocation (*FB*) and at the optimal feasible allocation (*Opt*), among alternative sets of potential contracts. *Metal-tier Contracts* are the primary set of contracts considered in the main text (and depicted in Fig. A.4a); *No Deductible* are a set of contracts that vary only in their coinsurance rate (see Fig. A.4c); *No Coinsurance Region* are a set of contracts between that vary only in their deductible (see Fig. A.4d); and *Extended Coins. Region* are a set of contracts that have no deductible and vary only in their coinsurance rate, and which have a stop-loss point of $20,000, twice as high as the other contracts (see Fig. A.4e). This table is referenced in Section V.C.

Table A.12. Parameter Estimates from Full Sample (Including Kaiser)

Variable	Parameter	Std. Err.	Variable	Parameter	Std. Err.
Employee Premium ($000s)	$-1.000^{\dagger}$		Kaiser * (Age−40)	−0.067	0.006
OOP spending, $-\alpha^{OOP}$	−1.429	0.026	Kaiser * **1**[Children]	−1.832	0.141
HRA/HSA contributions, α^{HA}	0.286	0.023	Kaiser * Region 1	−4.790	0.135
Vision/dental contributions, α^{VD}	1.285	0.024	Kaiser * Region 2	−7.930	0.323
Plan inertia intercept, γ^{plan}	5.119	0.065	Providence * (Age−40)	−0.047	0.007
Plan inertia * **1**[Children], γ^{plan}	−0.154	0.040	Providence * **1**[Children]	−0.629	0.151
Kaiser insurer inertia	9.750	0.262	Providence * Region 1	−1.655	0.132
Moda/Prov. insurer inertia,γ^{ins}	0.392	0.232	Providence * Region 2	−2.259	0.186
Insurer inertia * Risk score, γ^{ins}	0.553	0.073	Providence * Region 3	−1.551	0.213
Moda-specific inertia, 2013	2.162	0.199	κ * Risk Q_1	0.157	0.000
Narrow net. plan, $\nu^{NarrowNet}$	−2.639	0.166	κ * Risk Q_2	0.204	0.000
Kaiser utiliz. multiplier, ϕ_K	0.853	0.008	κ * Risk Q_3	0.188	0.000
Providence utiliz. multiplier, ϕ_P	1.118	0.001	κ * Risk Q_4	0.146	0.016
Risk aversion intercept, $\boldsymbol{\beta}^{\psi}$	−0.872	0.109	κ * Risk $Q_{n<4}$ * Risk score	0.005	0.000
Risk aversion * **1**[Children], $\boldsymbol{\beta}^{\psi}$	−0.096	0.071	κ * Risk Q_4 * Risk score	0.259	0.013
Moral hazard intercept, $\boldsymbol{\beta}^{\omega}$	1.160	0.002	μ * **1**[Female 18–35]	0.097	0.015
Moral hazard * **1**[Children], $\boldsymbol{\beta}^{\omega}$	0.425	0.000	μ * **1**[Age < 18]	0.018	0.015
Std. dev. of private health info., σ_μ	0.184	0.004	μ * Risk Q_1	−0.399	0.019
Std. dev. of log risk aversion, σ_ψ	0.621	0.064	μ * Risk Q_2	0.326	0.010
Std. dev. of moral hazard, σ_ω	0.097	0.001	μ * Risk Q_3	0.449	0.008
Corr(μ, ψ), $\rho_{\mu,\psi}$	0.373	0.004	μ * Risk Q_4	1.245	0.014
Corr(ψ, ω), $\rho_{\psi,\omega}$	−0.252	0.032	μ * Risk $Q_{n<4}$ * Risk score	1.127	0.018
Corr(μ, ω), $\rho_{\mu,\omega}$	0.135	0.007	μ * Risk Q_4 * Risk score	0.339	0.004
Scale of idiosyncratic shock, σ_ϵ	2.519	0.028	σ * Risk Q_1	1.431	0.008
			σ * Risk Q_2	1.240	0.004
			σ * Risk Q_3	1.191	0.003
			σ * Risk Q_4	1.031	0.004
Number of observations: 451,268					

Notes: The table presents parameter estimates using the full sample of households. The specification corresponds to column 3 of Tables 3 and A.10. The moral hazard parameter ω is estimated only within Moda and Providence plans. Standard errors are derived from the analytical Hessian of the likelihood function. The model is estimated on an unbalanced panel of 44,562 households, 14 plans, and 5 years. "Risk Q_n" is an indicator for an individual's risk quartile, where Q_4 is the sickest individuals. Higher risk scores correspond to worse predicted health. All parameters are measured in thousands of dollars. The insurer fixed effect of Moda is normalized to zero, and the utilization multiplier for Moda is normalized to one. This table is referenced in Section V.C. †By normalization.

Figure A.8. Results from Full Sample Parameter Estimates



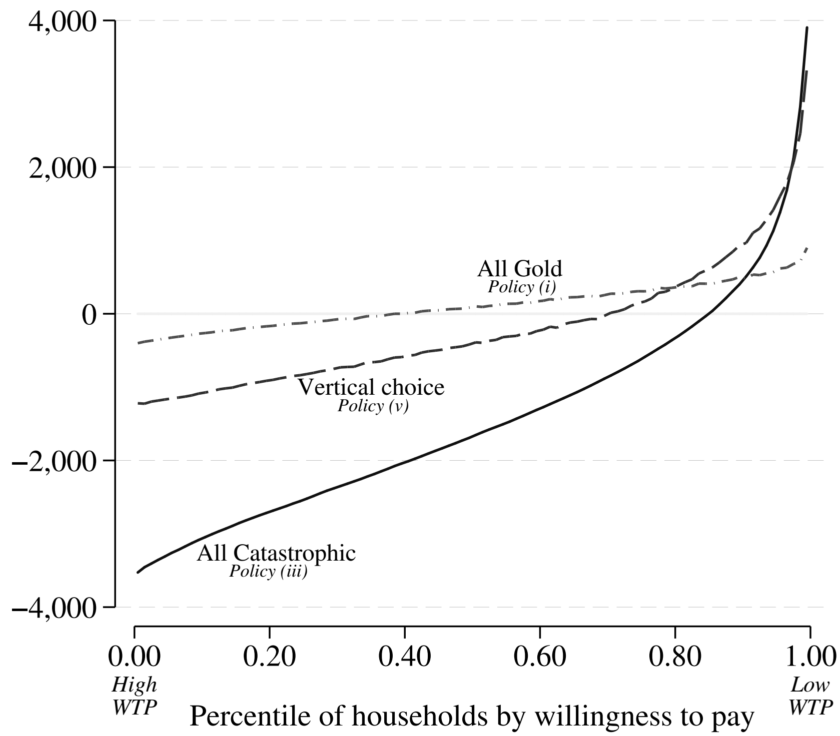
Notes: The figure shows the distribution across family households of (a) willingness to pay, (b) the decomposition of willingness to pay for the Gold contract, and (c) social surplus, using parameter estimates derived from the full sample of households (see Table A.12). The objects in all three panels are measured relative to the Catastrophic contract. Panel (a) consists of four connected binned scatter plots, with respect to 100 quantiles of households ordered by willingness to pay. Panel (b) consists of three connected binned scatter plots, with the area between each line shaded to indicate the component represented. Panel (c) consists of four connected binned scatter plots, with respect to 50 (to reduce noise) quantiles of households. This figure is referenced in Section V.C.

Table A.13. Outcomes Under Different Distributions of Consumer Types

		Outcomes at First Best (*FB*) and at the Optimal Menu (*Opt*), among:							
		Metal-tier contracts						Dense contracts	
Parameter Estimates		Full	Gold	Silv.	Brnz.	Ctstr.	*SS* ($)	Offer choice?	Δ *SS* ($)
Main estimates	*FB:*	0.06	0.75	0.19	<0.01	–	1,542		34
	Opt:	–	1.00	–	–	–	1,514	Yes	14
1. Double mean ω	*FB:*	–	0.29	0.64	0.07	–	1,091		42
	Opt:	–	–	1.00	–	–	1,069	Yes	4
2. Halve mean ω	*FB:*	0.39	0.61	<0.01	–	–	1,855		10
	Opt:	0.61	0.39	–	–	–	1,842	Yes	11
3. Double mean ψ	*FB:*	0.30	0.68	0.02	–	–	2,184		18
	Opt:	0.46	0.54	–	–	–	2,162	Yes	15
4. Halve mean ψ	*FB:*	–	0.35	0.63	0.02	<0.01	919		18
	Opt:	–	–	0.98	–	0.02	915	Yes	2
5. Increase var. ω	*FB:*	0.07	0.74	0.18	0.01	–	1,539		33
	Opt:	–	1.00	–	–	–	1,531	Yes	9
6. Increase var. ψ	*FB:*	0.13	0.64	0.21	0.02	<0.01	1,487		30
	Opt:	0.04	0.76	0.19	0.01	–	1,463	Yes	16
7. Fix F	*FB:*	0.06	0.83	0.11	–	–	1,410		17
	Opt:	–	1.00	–	–	–	1,407	Yes	6
8. Fix F and ω	*FB:*	0.16	0.67	0.17	–	–	1,457		14
	Opt:	0.14	0.68	0.18	–	–	1,456	Yes	12
9. Fix F and ψ	*FB:*	0.17	0.72	0.11	–	–	1,568		16
	Opt:	–	1.00	–	–	–	1,559	No	4

Notes: The table shows results under nine perturbations of our parameter estimates, as well as under our main estimates (column 3 of Tables 3 and A.10). Two sets of results are shown. First, the table shows the percent of households assigned to each of the five metal-tier contracts (Figure A.4a) under the first best allocation (*FB*) and under the optimal feasible allocation (*Opt.*), as well as the social surplus (*SS*) achieved by those allocations, relative to allocating all households to the Catastrophic contrat. Second, the table indicates whether or not the optimal menu features a choice when considering a dense set of contracts (Figure A.4b), as well as the associated social surplus gains (Δ *SS*) at the first best and under the optimal feasible allocation. The nine perturbation of parameter estimates are as follows: (1) double the moral hazard parameter ω for all households; (2) halve ω for all households; (3) double the risk aversion parameter ψ for all households; (4) halve ψ for all households; (5) double the amount of unobserved heterogeneity in moral hazard σ_ω; (6) double the amount of unobserved heterogeneity in log risk aversion σ_ψ; (7) fix household health type F in the population; (8) fix both health F and the moral hazard parameter ω in the population; and (9) fix both health F and risk aversion ψ in the population. This table is referenced in Section V.C.

Figure A.9. Distribution of Consumer Surplus ($), Relative to "All Full Insurance"



Notes: The figure shows the distribution of consumer surplus across households under three policies considered in Table 4. Households are arranged on the horizontal axis according to their willingness to pay. Consumer surplus equals marginal willingness to pay less marginal premium-plus-tax, *relative to the allocation of all households to full insurance*. That is, a policy of "All Full Insurance" would be represented by a horizontal line at zero. The premium-plus-tax that supports the single contract is $6,298 under "All Catastrophic," $10,619 under "All Gold," and $12,695 under "All full insurance." Premiums under "Vertical choice" are $7,059 for Full insurance, $4,594 for Gold, $2,173 for Silver, $375 for Bronze, $0 for Catastrophic, and a tax of $6,856. This figure is referenced in Section VI.B.