Nonparametric Estimates of Demand in the California Health Insurance Exchange *

Pietro Tebaldi† Alexander Torgovitsky‡ Hanbin Yang§

October 4, 2018

Abstract

We estimate the demand for health insurance in the California Affordable Care Act marketplace (Covered California) without using parametric assumptions about the unobserved components of utility. To do this, we develop a computational method for constructing sharp identified sets in a nonparametric discrete choice model. The model allows for endogeneity in prices (premiums) and for the use of instrumental variables to address this endogeneity. We use the method to estimate bounds on the effect of changes in premium subsidies on coverage choices, consumer surplus, and government spending. We find that a \$10 decrease in monthly premium subsidies would cause between a 3.3% and 8.4% decline in the proportion of low-income adults with coverage. The reduction in total annual consumer surplus would be between \$56 and \$70 million, while the savings in yearly subsidy outlays would be between \$441 and \$768 million. These nonparametric estimates reflect substantially greater price sensitivity than in comparable logit or probit models.

*We thank Nikhil Agarwal, Stéphane Bonhomme, Steve Berry, Øystein Daljord, Michael Dinerstein, Liran Einav, Phil Haile, Ali Hortaçsu, Simon Lee, Sanjog Misra, Magne Mogstad, Jack Porter, Bernard Salanié, and Thomas Wollman for helpful feedback.

†Department of Economics, University of Chicago. Research supported in part by the Becker Friedman Institute Health Economics Initiative.

‡Department of Economics, University of Chicago. Research supported in part by National Science Foundation grant SES-1426882.

§Harvard Business School.

1 Introduction

Under the Patient Protection and Affordable Care Act of 2010 ("ACA"), the United States federal government spends over $40 billion per year on subsidizing health insurance premiums for low-income households. The design of the ACA and the regulation of non-group health insurance remain objects of intense debate among policy makers. Addressing several key design issues, such as the structure of premium subsidies, requires estimating demand under counterfactual scenarios.

Recent research has filled this need using discrete choice models in the style of McFadden (1974). For example, Chan and Gruber (2010) and Ericson and Starc (2015) used conditional logit models to estimate demand in Massachusetts' Commonwealth Care program, Saltzman (2017) used a nested logit to estimate demand in the California and Washington ACA exchanges, and Tebaldi (2017) estimated demand in the California ACA exchange with a variety of logit, nested logit, and mixed (random coefficient) logit models.

These various flavors of logit models differ in the way they deal with the independence of irrelevant alternatives property (e.g. Goldberg, 1995; McFadden and Train, 2000), and in how they deal with potential endogeneity of prices (e.g. Berry, 1994; Hausman, 1996; Berry, Levinsohn, and Pakes, 1995). However, they are all fully parametric, with the logistic and normal distributions playing a central role in the parameterization. This raises the concerning possibility that these models generate demand predictions that are significantly driven by functional form.

In this paper, we use a nonparametric model to estimate the effects of changing premium subsidies on demand, consumer surplus, and government spending in the California ACA exchange (Covered California). The model is a distribution-free counterpart of a standard discrete choice model in which a consumer's indirect utility for an insurance option depends on the price (premium) and on their unobserved valuation for the option. In contrast to parametric models, we do not assume that these valuations follow a specific distribution such as normal (probit) or type I extreme value (logit). The main restriction of the model is that indirect utility is additively separable in premiums and latent valuations. The model allows for premiums to be endogenous (correlated with latent valuations), and allows a researcher to use instrumental variables to address this endogeneity.[1]

Point identification arguments in nonparametric discrete choice models with ex-

[1] While we develop our methodology with a focus on our health insurance application, we believe that it should also be of wider interest for demand analysis in other markets with product differentiation, as well as for other discrete choice settings more generally.

ogenous prices are often premised on the assumption of large variation in prices (e.g. Thompson, 1989; Matzkin, 1993). When prices are endogenous, these arguments shift the variation requirement to the instruments, sometimes with an additional completeness condition (Chiappori and Komunjer, 2009; Berry and Haile, 2010, 2014). In the Covered California data, we only observe limited variation in premiums, so these conditions will not be satisfied. This leads us to consider a partial identification framework.

The primary challenge with allowing for partial identification is finding a way to characterize and compute sharp bounds for target parameters of interest. We develop a characterization based on the observation that in a discrete choice model, many different realizations of latent valuations would lead to identical choice behavior under all relevant observed and counterfactual prices. Using this idea, we partition the space of unobserved valuations according to choice behavior by constructing a collection of sets that we call the Minimal Relevant Partition (MRP). We prove that sharp bounds for typical target parameters of interest can be characterized by considering only the way the distribution of valuations places mass over the MRP. We then use this result to develop estimators of these bounds, which we implement using linear programming.

We apply our empirical methodology with administrative data to estimate demand counterfactuals for the California ACA exchange. We focus on the choice of metal tier for low-income households who are not covered under employer-sponsored insurance or public programs. Our main counterfactual of interest is how changes in premium subsidies (or, equivalently, changes in the subsidized premium faced by the uninsured in the exchange) would affect the proportion of this population that chooses to purchase health insurance, as well as their chosen coverage tiers, and their realized consumer surplus. To identify these quantities, we use the additively separable structure of utility in the nonparametric model together with institutionally-induced variation in premiums across consumers of different ages and incomes. We exploit this variation by restricting the degree to which preferences (latent valuations) can differ across consumers of similar age and income who live in the same geographic market.

Since the nonparametric model is partially identified, this strategy yields bounds, rather than point estimates. However, the bounds are quite informative. Using our preferred specification, we estimate that a \$10 decrease in monthly premium subsidies would cause between a 3.3% and 8.8% decline in the proportion of low-income adults with coverage. The average consumer surplus reduction would be between \$1.91 and \$2.40 per person, per month, or between \$56 and \$70 million annually when aggregated. Savings on subsidy outlays would be \$441 to \$768 million. When we analyze heterogeneity by income, we find that poorer consumers incur the bulk of the surplus loss from decreasing subsidies. Overall, our estimates reinforce and amplify the finding that the demand for health insurance in this segment of the population is highly price elastic (e.g. Abraham, Drake, Sacks, and Simon, 2017; Finkelstein, Hendren, and Shepard, 2017).[2]

We show that comparable estimates using logit models tend to yield price responses close to our lower bounds, and so may substantially overstate the value that consumers place on health insurance. This possibility of understated price sensitivity is more severe when considering larger price changes that involve more distant extrapolations. It also remains when considering other parametric models, such as mixed logit, that allow for valuations to be correlated across options. Our findings provide an example in which the shape of the logistic (or similarly-shaped Gaussian) distribution can have an important impact on empirical conclusions.[3] The nonparametric model we use presents a remedy for this problem, and in this case provides empirical conclusions that differ significantly along a policy-relevant dimension.

The remainder of the paper is organized as follows. In Section 2, we begin with a discussion of the key institutional aspects of Covered California. In Section 3, we develop our nonparametric discrete choice methodology for estimating the demand for health insurance.[4] In Section 4, we discuss the data, our empirical implementation, and the main findings. In Section 5 we contrast these findings with estimates from parametric models. Section 6 contains some brief concluding remarks.

2 Covered California

Covered California is one of the largest state health insurance exchanges regulated by the ACA, accounting for more than 10% of national enrollment. The purpose of the exchange is to provide health insurance options for households not covered by an employer or a public program, such as Medicaid or Medicare.

The basic structure of Covered California is determined by federal regulation, and so is common to ACA marketplaces in all states. The regulation splits states into geographic rating regions comprised of groups of contiguous counties or zip codes. In California, there are 19 such rating regions. Insurers are allowed to vary premiums

[2] We do not model supply, so all of these estimates should be interpreted as holding insurers' decisions fixed while increasing the maximum affordable amount that determines the generosity of the federal subsidies. Tebaldi (2017) considers equilibrium price responses under different subsidy designs with a parametric demand model.

[3] Other examples include Ho and Pakes (2014) and Compiani (2018), who also found that logit models underestimate price elasticities relative to less parametric alternatives, albeit using different methods in different empirical settings.

[4] Appendix A contains a review of the related methodological literature on semi- and nonparametric discrete choice models.

across (but not within) rating regions, and consumers face the premiums set for their resident region. Each year in the spring, insurers announce their intention to enter a region in the subsequent calendar year and undergo a state certification process. Consumers are then able to purchase insurance for the subsequent year during an open enrollment period at the end of the year.

However, Covered California also differs from other ACA marketplaces in several important aspects. One difference is that an insurer who intends to participate in a rating region is required to offer a menu of four plans classified into metal tiers of increasing actuarial value: Bronze, Silver, Gold and Platinum.[5] Unlike other marketplaces, the insurer must provide the entire menu of four plans in any region where it enters.[6] Moreover, the actuarial features of the plans are standardized to have the characteristics shown in Table 1 (among others not shown). Insurers who enter a rating region must therefore offer each of the plans listed in Table 1 with the features shown there.

Insurers are also regulated in the way in which they can set premiums. Each insurer chooses a base premium for each metal tier in each rating region. This base premium is then transformed through federal regulation into premiums that vary by the consumer's age.[7] The insurer is not permitted to adjust premiums based on any other characteristic of the consumer.[8]

Households with income below 400% of the Federal Poverty Level (FPL) pay lower premiums than received by the insurer, with the difference being made up by premium subsidies. We focus our analysis on these households, since they constitute a large group of key policy interest.[9] The premium subsidies vary by household according to federal regulations. These ensure that the subsidized premium of Silver plans is lower than a maximum affordable amount, varying by income (measured in FPL) and household size.[10,11] To further incentivize insurance uptake, the ACA introduced a coverage mandate which determines an income tax penalty for remaining uninsured.

[5] There is a fifth coverage level called minimum (or catastrophic) coverage, but this is not available to subsidized buyers, so we omit it from our analysis.

[6] Under the ACA, insurers are mandated to offer one Silver and one Gold plan, while other plans offerings are optional.

[7] This transformation involves multiplying base premiums by an adjustment factor that starts at 1 for individuals at age 21 and increases smoothly to 3 at age 64. These factors are set by the Center for Medicaid & Medicare Services. Individuals 65 and older are covered by Medicare. See Orsini and Tebaldi (2017) for further discussion.

[8] Some states also allow for adjustments based on tobacco use, but California is not one of these states.

[9] In 2014, this group comprised over 94% of purchasing households in Covered California.

[10] See Tebaldi (2017) for details on how these subsidies are set.

[11] The reduction in subsidies we consider in our counterfactuals is equivalent to an increase in the maximum affordable amount, holding fixed insurers decisions.

Table 1: Standardized Financial Characteristics in Covered California

Panel (a): Characteristics by metal tier before cost-sharing reductions							
Tier	Annual deductible	Annual max out-of-pocket	Primary visit	E.R. visit	Specialist visit	Preferred drugs	Advertised AV*
Bronze	\$5,000	\$6,250	\$60	\$300	\$70	\$50	60%
Silver	\$2,250	\$6,250	\$45	\$250	\$65	\$50	70%
Gold	\$0	\$6,250	\$30	\$250	\$50	\$50	79%
Platinum	\$0	\$4,000	\$20	\$150	\$40	\$15	90%
Panel (b): Silver plan characteristics after cost-sharing reductions							
Income (%FPL)	Annual deductible	Annual max out-of-pocket	Primary visit	E.R. visit	Specialist visit	Preferred drugs	Advertised AV*
200-250% FPL	\$1,850	\$5,200	\$40	\$250	\$50	\$35	74%
150-200% FPL	\$550	\$2,250	\$15	\$75	\$20	\$15	88%
100-150% FPL	\$0	\$2,250	\$3	\$25	\$5	\$5	95%

Source: http://www.coveredca.com/PDFs/2015-Health-Benefits-Table.pdf

We treat this tax penalty as affecting the value of the outside option.

In addition to premium subsidies, the ACA also provides cost-sharing reduction (CSR) subsidies for households with income lower than 250% of the FPL. In Covered California, CSRs are implemented by changing the actuarial terms of the Silver plan for these households. As a result, the terms of a Silver plan vary across households according to income, with discrete changes at 150%, 200%, and 250% of the FPL; see Table 1. The CSRs make the Silver plan very attractive for low-income households relative to the more expensive Gold and Platinum plans.

3 Empirical Framework and Methodology

3.1 Nonparametric Model

We consider a model in which a population of consumers indexed by i each choose a single health insurance plan Y_i from a set $\mathcal{J} \equiv \{0, 1, \ldots, J\}$ of $J+1$ choices. Each plan j has a premium, P_{ij}, which is indexed by the consumer, i, since different consumers face different post-subsidy premiums depending on their sociodemographic characteristics. Choice $j = 0$ represents the outside option of not choosing any of the insurance plans, and has premium normalized to 0, so that $P_{i0} = 0$. When we take the model to the Covered California data in Section 4, we will have five choices ($J = 4$) with options $1, 2, 3$, and 4 representing Bronze, Silver, Gold, and Platinum plans, respectively.

Consumer i has a vector $V_i \equiv (V_{i0}, V_{i1}, \ldots, V_{iJ})$ of valuations for each plan, with the standard normalization that $V_{i0} = 0$.[12] The valuations are known to the consumer, but latent from the perspective of the researcher. We assume that consumer i's indirect utility from choosing plan j is given by $V_{ij} - P_{ij}$, so that their plan choice is given by

$$Y_i = \arg\max_{j \in \mathcal{J}} V_{ij} - P_{ij}. \tag{1}$$

We do not assume that the distribution of V_i follows a specific functional form such as i.i.d. type I extreme value (logit) or multivariate normal (probit). We also allow V_{ij} and V_{ik} to be dependent for $j \neq k$.

Models like (1) in which valuations and premiums are additively separable have been widely used in the recent literature on insurance demand, see e.g. Einav, Finkelstein, and Cullen (2010a), Einav, Finkelstein, and Levin (2010b), and Bundorf, Levin, and Mahoney (2012). In Appendix B, we derive (1) from an insurance choice model similar to the ones in Handel (2013) and Handel, Hendel, and Whinston (2015), in which consumers have quasilinear utility and constant absolute risk aversion preferences. In this model, differences in V_i across consumers arise from heterogeneity in their unobserved preferences, risk factors (and/or perception), and risk aversion.

The additive separability (quasilinearity) of premiums in (1) imposes restrictions on substitution patterns. In particular, if all premiums were to increase by the same amount, then a consumer who chose to purchase plan $j \geq 1$ before the premium increase will either continue to choose plan j after the premium increase, or will switch to the outside option ($j = 0$), but they will not switch to a different plan $k \geq 1$, $k \neq j$. This limits the role of income effects to the extensive margin of purchasing any insurance plan versus taking the outside option.

However, it is important to note that (1) is a model of a given consumer i. When we take (1) to the data, we combine observations on many consumers, so in practice we can allow for income effects by allowing for dependence between a consumer's income and their valuations. To formalize this, we treat a consumer's income and other observed characteristics as part of a vector, X_i, and then restrict the dependence between V_i and the various components of X_i. We discuss these restrictions in Section 3.4.1 and our specific implementation of them in Section 4.2.

One observable characteristic of consumer i that will be particularly important is their market, which in Covered California is their resident rating region. In particular, when we estimate demand we will do so conditional on a market, so that market-level

[12] Choosing $j = 0$ may imply the payment of a tax penalty, which affects the value of the outside option. Normalizing $V_{i0} = 0$ implies that we can interpret V_{ij} as the value of plan j for consumder i augmented by the expected tax penalty.

unobservables responsible for price endogeneity are held fixed in the counterfactual (e.g. see Berry and Haile, 2010, pg. 5). To emphasize this, we let M_i denote consumer i's market, and we denote M_i separately from X_i.

3.2 Comparison with a Common Parametric Model

A common parametric specification for discrete choice demand models is

$$Y_{im} = \arg\max_{j \in \mathcal{J}} \; X'_{ijm}\beta_{im} - \alpha_{im}P_{ijm} + \xi_{jm} + \epsilon_{ijm}, \tag{2}$$

where i, j, and m index consumers, products, and markets, P_{ijm} is price, X_{ijm} are observed characteristics, ξ_{jm} are unobserved product-market characteristics, β_{im} and α_{im} are individual-level random coefficients, and ϵ_{ijm} are idiosyncratic unobservables.[13] In the influential model of Berry et al. (1995), ϵ_{ijm} are assumed to be i.i.d. logit (type I extreme value), and $(\beta_{im}, \alpha_{im})$ are assumed to be normally distributed. Our motivation in considering (1) is to preserve the utility maximization structure in (2), while avoiding these types of parametric assumptions.[14]

The three indices in (2) reflect different possible levels of data aggregation. If only market-level data is available, as in Berry et al. (1995) or Nevo (2001), then (2) is aggregated to the (j, m) level, and the data is viewed as drawn from a population of markets and/or products (Berry, Linton, and Pakes, 2004b; Armstrong, 2016). Our model presumes richer individual-level choice data as in Berry, Levinsohn, and Pakes (2004a) or Berry and Haile (2010), but the number of markets we study is small and fixed. To emphasize this, we index the nonparametric model (1) only over i and j, and we record the identity of consumer i's market using the random variable M_i.

After replacing m subscripts with i subscripts, (1) can be seen to nest (2) by dividing through by α_i and taking $V_{ij} \equiv \alpha_i^{-1}(X'_{ij}\beta_i + \xi_{ij} + \epsilon_{ij})$.[15] This observation highlights some important considerations for our analysis. First, we do not want to assume that V_i and P_i are independent, since V_i depends on $\xi_i \equiv (\xi_{i1}, \ldots, \xi_{iJ})$, which captures unobservable characteristics of consumer i's market such as supply shocks (Berry, 1994). We address this by conditioning on the market, M_i, after which ξ_i is nonstochastic. Second, since V_{ij} depends on X_{ij}, we want to be careful about assuming

[13] For example, see equation (6) of Nevo (2011), or equation (1) of Berry and Haile (2015). We include i indices on X_{ijm} and P_{ijm} to maintain consistency with our notation.

[14] Fox, Kim, Ryan, and Bajari (2012) provide conditions under which the distribution of $(\beta_{im}, \alpha_{im})$ is nonparametrically point identified, and Fox, Kim, Ryan, and Bajari (2011) develop an estimator based on discretizing this distribution. Their results maintain the logit assumption on ϵ_{ijm}, and require additional structure to allow for price endogeneity.

[15] This requires the mild assumption that $\alpha_i > 0$ with probability 1.

that V_i and X_i are independent. This assumption is also threatened by the possibility that X_i is systematically related to the unobserved components of heterogeneity, α_i, β_i, and $(\epsilon_{i1}, \ldots, \epsilon_{iJ})$. Third, we want to allow for V_{ij} and V_{ik} to be arbitrarily dependent for $j \neq k$, in order to avoid imposing the unattractive substitution patterns associated with the logit model (Hausman and Wise, 1978; Goldberg, 1995; Berry et al., 1995; McFadden and Train, 2000).

3.3 Target Parameters

The primitive object of (1) is the distribution of valuations, V_i, conditional on premiums, P_i, market, M_i, and other covariates, X_i. We will assume throughout the paper that this distribution is continuous so that ties between choices in (1) occur with zero probability. In addition to ensuring no ties, this also means we can associate the conditional distribution of valuations with a conditional density function $f(\cdot|p, m, x)$ for each realization $P_i = p$, $M_i = m$, and $X_i = x$.[16] In Covered California, post-subsidy premiums are a deterministic function of market and consumer demographics. We denote this function by $P_i \equiv \pi(M_i, X_i)$, and we drop the redundant conditioning on P_i. In Appendix C, we discuss how to modify our methodology for settings in which prices vary at different levels.

The density f is a key object in the following sense. Common counterfactual quantities of interest can be written as integrals (or sums of integrals) of f. For example, a natural counterfactual is the proportion of consumers choosing plan j at a counterfactual premium vector, $p^\star$. This proportion can be written in terms of f as

$$\int \underbrace{\mathbb{1}[v_j - p_j^\star \geq v_k - p_k^\star \text{ for all } k]}_{\text{choose } j \text{ if premiums were } p^\star} f(v|m, x)\, dv, \tag{3}$$

where we are conditioning on market m and other consumer characteristics x. Another natural counterfactual quantity is the impact on average consumer surplus caused by changing premiums from p to $p^\star$. This can be written as

$$\underbrace{\int \left\{\max_{j \in \mathcal{J}} v_j - p_j^\star\right\} f(v|m, x)\, dv}_{\text{consumer surplus under } p^\star} - \underbrace{\int \left\{\max_{j \in \mathcal{J}} v_j - p_j\right\} f(v|m, x)\, dv}_{\text{consumer surplus under } p}, \tag{4}$$

where again market m is being held fixed in the counterfactual.

[16] More formally, this requires the assumption that the distribution of V_i, conditional on $(P_i, M_i, X_i) = (p, m, x)$ is absolutely continuously distributed with respect to Lebesgue measure on $\mathbb{R}^J$ for every (p, m, x) in the support of (P_i, M_i, X_i).

Conceptually, we view both (3) and (4) as scalar-valued functions of f. The functions vary in their form, and will further vary when we consider different counterfactual premiums $p^\star$, choice probabilities for plans other than j in (3), and different values of (or averages over) the covariates, x. In Section 4, we also estimate a third class of quantities that measure changes in subsidy outlays.

To handle this generality, we consider all such quantities to be examples of *target parameters*, $\theta : \mathcal{F} \to \mathbb{R}^{d_\theta}$, where $\mathcal{F}$ is the collection of all conditional density functions on $\mathbb{R}^J$. A target parameter is just a function of the conditional density of valuations, f. In the examples just given it is a scalar function, so that $d_\theta = 1$. However, we might also want to consider cases with $d_\theta > 1$, for example to understand the joint identified set for two related target parameters of interest, such as consumer surplus and government expenditure. Our goal is to infer the values of $\theta(f)$ that are consistent with both the observed data and our prior assumptions.

3.4 Assumptions

We augment (1) with two types of prior assumptions. The first assumption is that one or more components of X_i are suitable instruments. The second assumption exploits the vertical structure of the metal tiers in the ACA.

3.4.1 Instrumental Variables

To describe the first type of assumption, let W_i and Z_i be two subvectors (or more general functions) of the market and covariates, M_i and X_i. The Z_i subvector are instruments that satisfy an exogeneity assumption discussed ahead. This exogeneity assumption will be conditional on W_i, which are viewed as controls. Note that either or both of these subvectors could be chosen to be empty.

Stating the instrumental variable assumption requires the density of valuations conditional on W_i and Z_i. We can construct this object by averaging over f as follows:

$$f_{V|WZ}(v|w, z) \equiv \mathbb{E}\left[f(v|M_i, X_i) \middle| W_i = w, Z_i = z \right]. \tag{5}$$

Our assumption that Z_i is an instrument, conditional on W_i, can then be stated as:

$$f_{V|WZ}(v|w, z) = f_{V|WZ}(v|w, z') \quad \text{for all } z, z', w, \text{ and } v. \tag{6}$$

In words, (6) says that the distribution of valuations is invariant to shifts in Z_i, conditional on W_i. That is, Z_i is exogenous. In our application, W_i includes M_i and coarse age and income bins, and Z_i is residual variation in age and income within these bins.

Thus, (6) requires latent valuations to be invariant to small changes in age and income. In Section 4.4, we discuss one way to weaken this assumption.

In order for (6) to be a useful assumption, shifts in the instrument Z_i (still conditioning on W_i) should have an effect on premiums. This follows the usual intuition: If Z_i is exogenous, then changes in observed choice shares as Z_i varies reflect changes in premiums, rather than changes in valuations. The more that premiums vary with Z_i, the more information we will have to pin down different parts of the density of valuations, f, and hence the target parameter, θ. In our application, this premium variation comes from the age-rating and income subsidies legislated by the ACA.

It is common to justify point identification of nonparametric discrete choice models by assuming that the instrument has a large amount of variation.[17] However, in our data we can plainly see that this is not the case. For this reason, we consider the partial identification framework discussed in the next section. This framework does not require the instrument to have any particular amount of variation. However, greater variation is still rewarded in the form of more informative bounds.

3.4.2 Vertical Structure

The second assumption we use exploits the vertical structure of the ACA, i.e. the fact that the Platinum plan is actuarially more generous than the Bronze plan. For example, the Bronze plan has a higher deductible and higher out-of-pocket maximum than the Platinum plan (see Table 1). Our assumption is that, for equal premiums, a consumer would always prefer a plan that is more generous to one that is less generous. With $j = 4$ as the Platinum plan, and $j = 1$ is the Bronze plan, this means we assume that f places zero mass on regions where $v_1 > v_4$ or, equivalently, concentrates all of its mass on regions with $v_4 \geq v_1$.[18]

Implementing this assumption in the context of the ACA is complicated by the existence of CSR subsidies. As discussed in Section 2, CSRs are used in Covered California to change the terms of the Silver plan depending on a consumer's income. Lower-income consumers face more generous Silver plans, and this generosity gets gradually phased out at higher incomes. The effect of this is that, depending on a consumer's income, they might prefer a Silver plan to a Gold or even a Platinum plan

[17] These types of "large support" assumptions, and the closely related concept of identification-at-infinity, have had a prominent role in the literature on nonparametric identification more generally. Early examples of their use include Manski (1985), Thompson (1989), Heckman and Honoré (1990), and Lewbel (2000). More recent applications of this argument to discrete choice include Heckman and Navarro (2007) and Fox and Gandhi (2016).

[18] Note that we will not assume that consumers prefer any of the plans (inside options) to the outside option. That is, we always allow for $v_j \leq v_0 = 0$ for $j = 1, 2, 3, 4$.

(at equal premiums).

With this flexibility in mind, we formalize the verticality assumption as follows. For each realization of W_i defined as in the previous section, we choose a set $\overline{\mathcal{V}}(w)$ and then assume that f is such that

$$\int_{\overline{\mathcal{V}}(w)} f_{V|WZ}(v|w,z)\,dv = 1 \quad \text{for all } w, z. \tag{7}$$

This assumption captures the idea that the distribution of valuations is concentrated on a given region, e.g. by taking $\overline{\mathcal{V}}(w) = \{v : v_4 \geq v_1\}$ in the example above. Allowing $\overline{\mathcal{V}}(w)$ to change with covariates w will be used in our application to allow the vertical ordering to change with income, so as to account for CSRs. The case of no verticality assumption is nested by taking $\overline{\mathcal{V}}(w) = \mathbb{R}^J$.

3.5 The Identified Set

We now develop our method for determining the set of possible values that the target parameter $\theta(f)$ could take over valuation densities f that both satisfy the assumptions in the preceding section, and are consistent with the observed data. To do this, we assume that researcher has at their disposal a collection of conditional choice shares, denoted as

$$s(j|m,x) \equiv \mathbb{P}[Y_i = j | M_i = m, X_i = x]. \tag{8}$$

In our application, we estimate these shares from a combination of administrative data on enrollment and survey data used to construct the market size. Here, our analysis of identification is premised on the thought experiment of perfect knowledge of these choice shares.[19]

Each density of valuations implies a set of choice shares analogous to (8). In particular, a consumer would choose option j when faced with a premium p if and only if they have valuations in the set

$$\mathcal{V}_j(p) \equiv \left\{(v_1, \ldots, v_J) \in \mathbb{R}^J : v_j - p_j \geq v_k - p_k \ \text{ for all } k\right\}. \tag{9}$$

The choice shares for plan j implied by the density f are just determined by the mass that f places on this set when prices are $P_i = \pi(M_i, X_i)$. We denote these implied

[19] More formally, it is premised on perfect knowledge of the joint distribution of (Y_i, M_i, X_i).

choice shares by

$$s_f(j|m,x) \equiv \int_{\mathcal{V}_j(\pi(m,x))} f(v|m,x)\,dv. \tag{10}$$

We say that a valuation density f is observationally equivalent if its predicted choice shares match the observed choice shares, i.e. if

$$s_f(j|m,x) = s(j|m,x) \quad \text{for all } j,\ m \text{ and } x. \tag{11}$$

The identified set of valuation densities is the set of all f that are both observationally equivalent and satisfy the assumptions laid out in the previous section. We call this set $\mathcal{F}^\star$:

$$\mathcal{F}^\star \equiv \{f \in \mathcal{F} :\ f \text{ satisfies (6), (7), and (11)}\}. \tag{12}$$

However, our real interest centers on the target parameter, θ, examples of which include counterfactual demand (3) and changes in consumer surplus (4). The identified set for θ is the image of the identified set for $\mathcal{F}^\star$ under θ. That is,

$$\Theta^\star \equiv \{\theta(f) : f \in \mathcal{F}^\star\}.$$

The set $\Theta^\star$ consists of all values of the target parameter that are consistent with both the data and the instrumental variable and verticality assumptions (6) and (7). It is our central object of interest.

The difficulty lies in characterizing $\Theta^\star$. In the following, we develop an argument that enables us to compute $\Theta^\star$ exactly. The idea is to partition $\mathbb{R}^J$ into the smallest collection of sets within which choice behavior would remain constant under all premiums that were observed in the data, as well as all premiums that are required to compute the target parameter. We call this set the *minimal relevant partition* (MRP) of valuations. We then reduce the problem of characterizing $\Theta^\star$ from one of searching over densities f to one of searching over mass functions defined on the sets that constitute the MRP. For cases in which the target parameter is scalar-valued ($d_\theta = 1$), this latter problem can often be solved with two linear programs.

3.6 The Minimal Relevant Partition of Valuations

We illustrate the definition and construction of the MRP using a simple example with $J = 2$, so that a consumer's valuations (and the premiums of the plans in their choice set) can be represented as points in the plane. A general (and formal) definition of the

MRP is given in Section 3.8.

Suppose that the data consists of a single observed premium vector, $p^a \equiv (p_1^a, p_2^a)$, and that we are concerned with behavior under a counterfactual premium vector, $p^\star$, which we do not observe in the data. The idea behind the MRP is illustrated in Figure 1. Panel (a) shows that considering behavior under premium p^a divides $\mathbb{R}^2$ into three sets depending on whether a consumer would choose options 0, 1, or 2 when faced with p^a. Panel (b) shows the analogous situation under premium $p^\star$. Intersecting these two three-set collections creates the collection of six sets shown in panel (c). This collection of six sets is the MRP for this example.

The MRP is "minimally relevant" in the sense that any two consumers who have valuations in the same set would exhibit the same choice behavior under both premiums p^a and $p^\star$. Conversely, any two consumers with valuations in different sets would exhibit different choice behavior under at least one of these premiums. For example, consumers with valuations in the set marked $\mathcal{V}_2$ in Figure 1c make the same choices as those with valuations in $\mathcal{V}_4$ under p^a, but make different choices under $p^\star$, where the first group chooses the outside option, and the second group chooses plan 1. Similarly, consumers with valuations in $\mathcal{V}_2$ and $\mathcal{V}_6$ both choose the outside option at $p^\star$, but at p^a the first group chooses plan 2 and the second group chooses plan 1.

In Figure 1d, we show how the MRP would change if we were to observe a second premium, p^b. The MRP now consists of ten sets, but the idea is the same: Consumers with valuations within a given set have the same choice behavior under all premiums p^a, p^b, and $p^\star$, while consumers with valuations in different sets would make different choices for at least one of these premiums.

The way the MRP is constructed ensures that predicted choice shares for any valuation density can be computed by summing the mass that the density places on sets included in the MRP. For example, suppose that we fix $M_i = m$, and that there are two values of X_i such that $p^a = \pi(m, x^a)$, and $p^b = \pi(m, x^b)$. In Figure 1c, we can see that the share of consumers who would choose good 1 if premiums were p^a can be written as

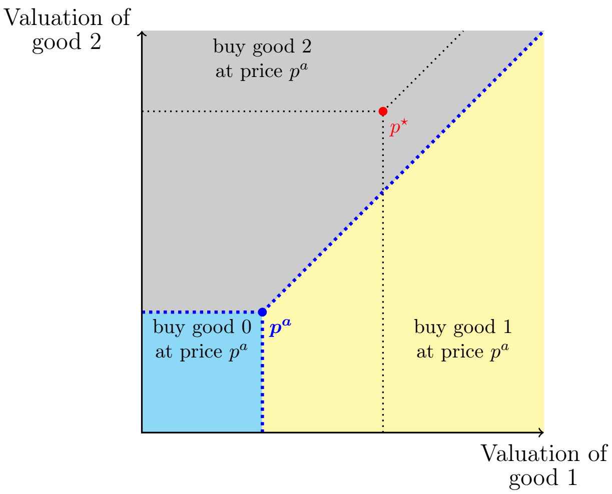
$$s_f(1|m, x^a) = \int_{\mathcal{V}_5} f(v|m, x^a)\, dv + \int_{\mathcal{V}_6} f(v|m, x^a)\, dv,$$

while the share of consumers who would choose good 2 is given by

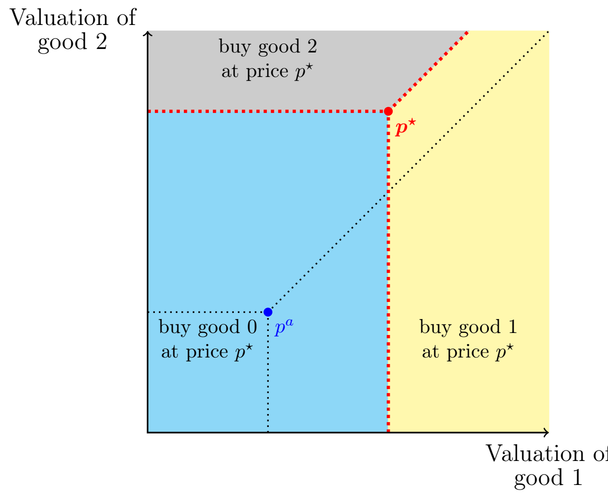
$$s_f(2|m, x^a) = \int_{\mathcal{V}_2 \cup \mathcal{V}_3 \cup \mathcal{V}_4} f(v|m, x^a)\, dv.$$

This allows us to simplify the determination of whether a given f is observationally

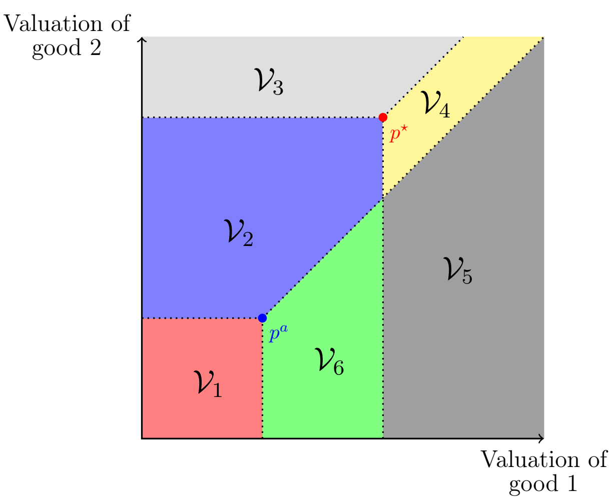
Figure 1: *Partitioning the Space of Valuations*



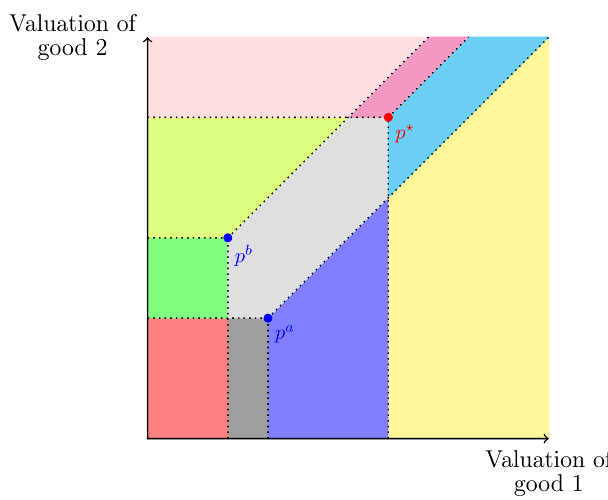
(a) *Choices if prices were p^a.*



(b) *Choices if prices were $p^\star$.*



(c) *The minimal relevant partition (MRP) constructed from $p^a, p^\star$.*



(d) *The minimal relevant partition (MRP) constructed from p^a, p^b, and $p^\star$.*

equivalent by considering only the total mass that f places in sets in the MRP, without having to be concerned with how this mass is distributed within these sets.

Since we constructed the MRP including $p^\star$, the same is also true when considering target parameters θ that measure choice behavior at $p^\star$. For example, suppose that our target parameter is the choice share of plan 2 if premiums were changed from p^a to $p^\star$. This is a particular case of (3), and can be written in terms of the MRP as

$$\theta(f) = \int_{\mathcal{V}_3} f(v|m, x^a)\, dv. \tag{13}$$

As another example, we could write the associated *change* in this choice share as

$$\theta(f) = \int_{\mathcal{V}_3} f(v|m, x^a)\, dv - \int_{\mathcal{V}_2 \cup \mathcal{V}_3 \cup \mathcal{V}_4} f(v|m, x^a)\, dv = -\int_{\mathcal{V}_2 \cup \mathcal{V}_4} f(v|m, x^a)\, dv.$$

In both of these quantities, we have fixed the density conditional on the market, m, and observed covariates, x^a. This corresponds to the usual counterfactual of changing prices while holding fixed unobservable factors that are correlated with price.

3.7 Computing Bounds on the Target Parameter

Now suppose that we observe the following choice shares:

$$s(0|m, x^a) = .20, \quad s(1|m, x^a) = .14, \quad \text{and} \quad s(2|m, x^a) = .66,$$

For simplicity, we will start by assuming that X_i is exogenous, i.e. we limit our attention to f for which $f(v|m, x^a) = f(v|m, x^\star) = f(v|m)$.[20] In this case, the observational equivalence condition (11) can be written as

$$\begin{aligned}
\int_{\mathcal{V}_1} f(v|m)\, dv &= s(0|m, x^a) = .20, \\
\text{and} \quad \int_{\mathcal{V}_5} f(v|m)\, dv + \int_{\mathcal{V}_6} f(v|m)\, dv &= s(1|m, x^a) = .14, \\
\text{and} \quad \int_{\mathcal{V}_2} f(v|m)\, dv + \int_{\mathcal{V}_3} f(v|m)\, dv + \int_{\mathcal{V}_4} f(v|m)\, dv &= s(2|m, x^a) = .66.
\end{aligned} \tag{14}$$

As shown in (13), if the target parameter is the choice share of plan 2 at $p^\star$, this can be written as

$$\theta(f) = \int_{\mathcal{V}_3} f(v|m)\, dv. \tag{15}$$

[20] In terms of (6), this would be like taking $W_i = M_i$ and $Z_i = X_i$.

The key observation is that even though all of these quantities depend on a density f, they can be computed with knowledge of just six non-negative numbers:

$$\left\{ \phi_l \equiv \int_{\mathcal{V}_l} f(v|m)\, dv \right\}_{l=1}^{6}.$$

This suggests that we can focus only on the total mass placed on the sets in the MRP without losing any information. To find the largest value that $\theta(f)$ can take while still respecting (14), we rephrase all quantities in terms of $\{\phi_l\}_{l=1}^6$ and then maximize (15) subject to (14):

$$\begin{aligned} t^\star \equiv \max_{\phi \in \mathbb{R}^6} \ & \phi_3 \\ \text{subject to:} \quad & \phi_1 = .20 \\ & \phi_5 + \phi_6 = .14 \\ & \phi_2 + \phi_3 + \phi_4 = .66 \\ & \phi_l \geq 0 \quad \text{for } l = 1, \ldots, 6. \end{aligned} \tag{16}$$

This is a linear program. In this simple example, one can see by inspection that the solution of the program is to take $\phi_3 = .66$, so that $t^\star = .66$. To find the smallest value of $\theta(f)$ we solve the analogous minimization problem, the optimal value of which we call $t_\star$. In this example, $t_\star = 0$.

In the next section, we formally prove that $\Theta^\star = [t_\star, t^\star]$. This result shows that the procedure of reducing f to a collection of six numbers $\{\phi_l\}_{l=1}^6$ is a sharp characterization of $\Theta^\star$ in the sense that it entails no loss of information. The intuition behind the sharpness is as follows. First, for any value $t \in \Theta^\star$, there must exist (by definition) an $f \in \mathcal{F}^\star$ such that $\theta(f) = t$. This f generates a collection of numbers $\{\phi_l = \int_{\mathcal{V}_l} f(v|m)\, dv\}_{l=1}^6$, which must satisfy the constraints in (16), since every $f \in \mathcal{F}^\star$ satisfies (14). Conversely, given any value of $t \in [t_\star, t^\star]$, there exists a set of numbers $\{\phi_l\}_{l=1}^6$ satisfying the constraints in (16), and such that $\phi_3 = t$.[21] From this set of numbers $\{\phi_l\}_{l=1}^6$, we can construct a density f that satisfies (14) by distributing mass in the amount of ϕ_l arbitrarily within each $\mathcal{V}_l$. Evidently, this density will also satisfy $\theta(f) = \phi_3 = t$. Thus, $\Theta^\star = [t_\star, t^\star]$.

Now suppose that we have a second observed premium, p^b, so that the MRP is as shown in Figure 1d. In this case, the MRP contains 10 sets, so the linear program analogous to (16) will have 10 variables of optimization. In addition to the observational

[21] This follows because the constraint set in (16) is closed and connected and the objective function is continuous.

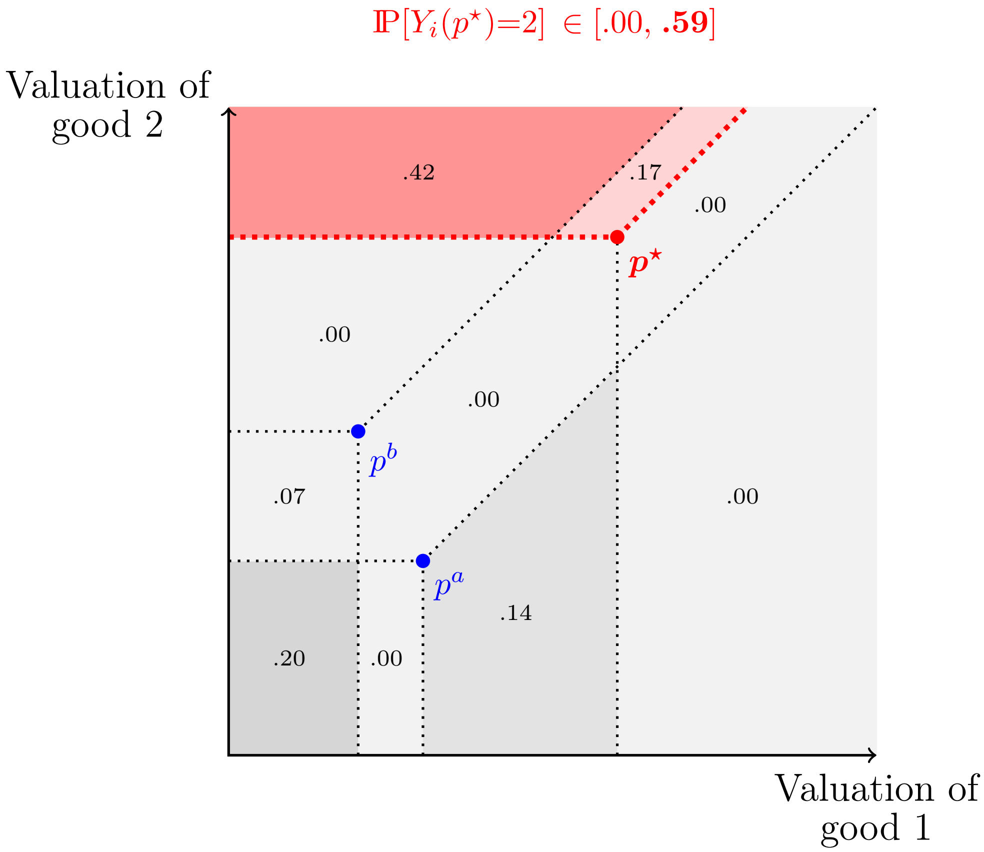


Figure 2: *The numbers in each set show a solution to the linear program when the target parameter is the proportion of consumers who choose plan 2 at* $p^\star$ *and the objective is to find the upper bound (maximize) this proportion. Matching the share of consumers who choose the outside option at the new observed premium,* p^b*, means there is now* .07 *less mass to devote to this objective.*

equivalence constraints for x^a in (16), these variables will also need to satisfy three more observational equivalence constraints corresponding to the observed shares for x^b, which we will suppose here are given by

$$s(0|m, x^b) = .27, \quad s(1|m, x^b) = .31, \quad \text{and} \quad s(2|m, x^b) = .42.$$

Reasoning through the solution to the resulting program is more complicated. Since the observed shares for p^a still need to be matched, it is still the case that a total mass of .66 must be placed over consumers who would choose plan 2 under p^a. Some of these consumers might choose the outside option under p^b. In fact, as shown in Figure 2, this must be the case for a proportion of at least $s(0|m, x^b) - s(0|m, x^a) = .07$ of consumers. Given this new requirement, the maximum amount of mass remaining to distribute over consumers who would choose plan 2 under $p^\star$ has decreased from .66 to

$.66 - .07 = .59$. This is the new upper bound, $t^\star$. The fact that it is smaller than the previous upper bound reflects the additional information contained in p^b. The lower bound, $t_\star$, is still zero, because it is still possible to match the observed choice shares for p^a and p^b by concentrating all mass to the south of $p^\star$.

When we take this procedure to the data, the linear programs will have thousands of variables and constraints, which makes this sort of case-by-case reasoning impossible. Instead, we will use state of the art solvers to obtain $t^\star$ and $t_\star$.[22] In practice, we also do not assume that $f(v|m, x)$ is invariant in x. This makes a graphical interpretation unweildy, since a separate diagram like Figure 2 would be needed for each value of x. The mass placed over sets within each diagram is linked together by imposing constraints on these masses that are analogous to the instrumental variable assumption (6). Part of the formal analysis in the next section involves showing that such a procedure retains sharpness.

3.8 Formalization

In this section, we formalize the discussion in the previous three sections in the following ways. First, we provide a precise definition of the MRP. Second, we generalize the transformation from densities f to mass functions over the sets in the MRP, which, as in the previous section, we refer to as ϕ. Third, we show how to compute bounds for any target parameter under the instrumental variable and verticality assumptions. Fourth, we provide the general statement and proof of the result that these bounds are sharp. Lastly, we consider the conditions under which these bounds can be computed by solving linear programs. Throughout the analysis, we model (M_i, X_i) as discretely distributed with finite support, although this is not essential to the methodology.

Beginning with the MRP, we let $\mathcal{P}$ denote a finite set of premiums that is chosen by the researcher and always contains at least the marginal support of premiums, $\text{supp}(P_i)$. The premiums in $\mathcal{P}$ are used to construct the MRP, so a given MRP depends on $\mathcal{P}$. For example, in Figure 1c we had $\mathcal{P} = \{p^a, p^\star\}$, while in Figure 1d, $\mathcal{P} = \{p^a, p^b, p^\star\}$.

The choice of which additional points to include in $\mathcal{P}$ is determined by the parameter of interest, θ. In Figure 1, our focus was on demand at a new premium, $p^\star$, so $\mathcal{P}$ had to include $p^\star$. This restriction will be formalized below as the statement that $\theta(f)$ can be evaluated for any f by only considering the total mass that f places on sets in the MRP. Additional points can always be added to $\mathcal{P}$ to help satisfy this restriction.

We use the set $\mathcal{P}$ to formally define the MRP as follows.

[22] In particular, we use Gurobi (Gurobi Optimization, 2015) and double-check a subset of the results using CPLEX (IBM, 2010). We formulate and presolve the problems using AMPL (Fourer, Gay, and Kernighan, 2002).

Definition 1. *Let* $Y(v, p) \equiv \arg\max_{j \in \mathcal{J}} v_j - p_j$ *for any* $(v_1, \ldots, v_J), (p_1, \ldots, p_J) \in \mathbb{R}^J$*, where* $v \equiv (v_0, v_1, \ldots, v_J)$ *and* $p \equiv (p_0, p_1, \ldots, p_J)$ *with* $v_0 = p_0 = 0$*. The* minimal relevant partition of valuations *(MRP) is a collection* $\mathbb{V}$ *of sets* $\mathcal{V} \subseteq \mathbb{R}^J$ *for which the following property holds for almost every* $v, v' \in \mathbb{R}^J$ *(with respect to Lebesgue measure):*

$$v, v' \in \mathcal{V} \text{ for some } \mathcal{V} \in \mathbb{V} \quad \Leftrightarrow \quad Y(v, p) = Y(v', p) \text{ for all } p \in \mathcal{P}. \tag{17}$$

Definition 1 creates a collection of sets that is "minimally relevant" in the sense that any two consumers who have valuations in a set in the collection would exhibit the same choice behavior for every premium vector in $\mathcal{P}$. Conversely, any two consumers with valuations in different sets would exhibit different choice behavior for at least one premium in $\mathcal{P}$. Constructing the MRP is intuitive, but somewhat involved both notationally and algorithmically. Since the details of constructing the MRP are not necessary for understanding the methodology, we relegate our discussion of this to Appendix D.[23]

The utility of the MRP as a concept is that it allows us to express the choice probabilities associated with any density of valuations, f, in terms of the mass that f places on sets in $\mathbb{V}$. In particular, for every $p \in \mathcal{P}$ and $j \in \mathcal{J}$, let $\mathbb{V}_j(p) \subseteq \mathbb{V}$ denote the sets in the MRP for which a consumer with valuations in these sets would choose j when facing premiums p.[24] Then the probability that a consumer chooses j under premiums p is the probability that V_i lies in the union of $\mathcal{V} \in \mathbb{V}_j(p)$. Since sets in $\mathbb{V}$ are disjoint, the observational equivalence condition (10) can be written as the sum of the masses that a given f places on sets in $\mathbb{V}_j(p)$, that is

$$s_f(j|m, x) = \sum_{\mathcal{V} \in \mathbb{V}_j(\pi(m,x))} \int_{\mathcal{V}} f(v|m, x)\, dv. \tag{18}$$

Having defined the MRP, we now define mass functions over the MRP. To do this, let $\phi(\cdot|\cdot, \cdot)$ denote a function with domain $\mathbb{V} \times \operatorname{supp}(M_i, X_i)$. Such a function ϕ can be viewed as an element of $\mathbb{R}^{d_\phi}$, where d_ϕ is the product of the cardinalities of these sets.

[23] We should, however, note two small misnomers in our terminology that become evident in the construction, or perhaps by inspecting Figure 1. First, the MRP may not be a strict partition, because adjacent sets in $\mathbb{V}$ could overlap on their boundary. Since we are limiting attention to continuously distributed valuations, this distinction does not have any practical or empirical relevance, and does not violate Definition 1. Second, and for the same reason, although we have described the MRP as "the" MRP, it may not be unique, since one could consider a boundary region to be in either of the sets to which it is a boundary without violating (17) on a set of positive measure. Again, this is not important for our analysis given our focus on continuously distributed valuations.

[24] Using the notation of Definition 1, $\mathbb{V}_j(p) \equiv \{\mathcal{V} \in \mathbb{V} : Y(v, p) = j \text{ for almost every } v \in \mathcal{V}\}$.

Let $\mathbb{R}_+^{d_\phi}$ denote the subset of $\mathbb{R}^{d_\phi}$ whose elements are all non-negative and define

$$\Phi \equiv \left\{ \phi \in \mathbb{R}_+^{d_\phi} : \sum_{\mathcal{V} \in \mathbb{V}} \phi(\mathcal{V}|m, x) = 1 \quad \text{for all } (m, x) \in \operatorname{supp}(M_i, X_i) \right\}. \tag{19}$$

The set Φ contains all functions that could represent a conditional probability mass function with domain given by the finite collection of sets, $\mathbb{V}$.

Each conditional valuation density f generates a mass function $\phi_f \in \Phi$ defined by

$$\phi_f(\mathcal{V}|m, x) \equiv \int_{\mathcal{V}} f(v|m, x)\, dv. \tag{20}$$

We assume that the value of the target parameter for any f is fully determined by ϕ_f. Formally, the assumption is that there exists a known function $\overline{\theta}$ with domain Φ such that $\theta(f) = \overline{\theta}(\phi_f)$ for every $f \in \mathcal{F}$. Since Φ depends on the MRP, and the MRP depends on $\mathcal{P}$, satisfying this requirement is a matter of choosing $\mathcal{P}$ to be sufficiently rich to evaluate the target parameter of interest, θ.

We have now phrased both the target parameter and observational equivalence condition in terms of ϕ. The last step is to translate the instrumental variable and verticality assumptions into statements about ϕ. For the instrumental variable assumption, we first define for any $\phi \in \Phi$ a function $\phi_{\mathbb{V}|WZ}$ in analogy to (5) as

$$\phi_{\mathbb{V}|WZ}(\mathcal{V}|w, z) \equiv \mathbb{E}\left[\phi(\mathcal{V}|M_i, X_i) \middle| W_i = w, Z_i = z\right], \tag{21}$$

where W_i and Z_i are as in the statement of that condition. Then, a condition appropriately analogous to (6) is

$$\phi_{\mathbb{V}|WZ}(\mathcal{V}|w, z) = \phi_{\mathbb{V}|WZ}(\mathcal{V}|w, z') \quad \text{for all } z, z', w, \text{ and } \mathcal{V}. \tag{22}$$

Similarly, for the verticality assumption, we define in analogy to (7),

$$\sum_{\mathcal{V} \in \overline{\mathbb{V}}(w)} \phi_{\mathbb{V}|WZ}(\mathcal{V}|w, z) = 1 \quad \text{for all } w, z, \tag{23}$$

where $\overline{\mathbb{V}}(w)$ is the subset of $\mathbb{V}$ that intersects $\overline{\mathcal{V}}(w)$, i.e. $\overline{\mathbb{V}}(w) \equiv \{\mathcal{V} \in \mathbb{V} : \lambda(\mathcal{V} \cap \overline{\mathcal{V}}(w)) > 0\}$, with λ denoting Lebesgue measure on $\mathbb{R}^J$.

The next proposition shows that $\Theta^\star$ can be characterized exactly by solving systems of equations in ϕ.

Proposition 1. *Let $t \in \mathbb{R}^{d_\theta}$. Then $t \in \Theta^\star$ if and only if there exists a $\phi \in \Phi$ such that*

$$\overline{\theta}(\phi) = t, \tag{24}$$

$$\sum_{\mathcal{V} \in \mathbb{V}_j(\pi(m,x))} \phi(\mathcal{V}|m,x) = s(j|m,x) \quad \textit{for all } j \in \mathcal{J} \textit{ and } (m,x), \tag{25}$$

$$\phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z) = \phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z') \quad \textit{for all } z, z', w, \textit{ and } \mathcal{V}, \tag{26}$$

$$\textit{and} \quad \sum_{\mathcal{V} \in \overline{\mathbb{V}}(w)} \phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z) = 1 \quad \textit{for all } w, z. \tag{27}$$

Observe that each of (25)–(27) are linear in ϕ.[25] If $\overline{\theta}$ is also linear in ϕ, then Proposition 1 shows that $\Theta^\star$ can be exactly characterized by solving linear systems of equations. This linearity is satisfied for common target parameters, such as demand and consumer surplus.[26] One byproduct of this linearity is that $\Theta^\star$ will be connected, and so when $d_\theta = 1$ it can also be characterized by solving two linear programs. We record this point in the following proposition.

Proposition 2. *If $\overline{\theta}$ is continuous on Φ, then $\Theta^\star$ is a compact, connected set. In particular, if $d_\theta = 1$, then $\Theta^\star = [t_\star, t^\star]$, where*

$$t_\star \equiv \min_{\phi \in \Phi} \ \overline{\theta}(\phi) \qquad \textit{subject to } (25)\text{–}(27), \tag{28}$$

and with $t^\star$ defined as the solution to the analogous maximization problem.

In practice, one often finds that the feasible set in (28) is empty, so that $\Theta^\star$ is also empty.[27] This is an indication of either sampling error in the observed shares, $s(j|m,x)$, or model misspecification, or both. Instead of reporting empty identified sets, we modify Proposition 2 to construct a set estimator of $\Theta^\star$. The estimator uses a procedure analogous to the method of moments for point identified models. First, we minimize a criterion function that measures the extent to which the observational equivalence equality in (25) is violated. Second, we find the set of values t that $\overline{\theta}(\phi)$ can take while coming close to the optimal value of the criterion.[28] By choosing an absolute deviations criterion, this procedure amounts to solving three linear programs in cases where $\overline{\theta}$ is linear. We provide more detail on the estimation procedure in Appendix G.

[25] This requires noting from (21) that $\phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z)$ is itself a linear function of ϕ.
[26] The former is clear from e.g. (15), but the latter is not obvious; see Appendix F.
[27] We follow the usual convention here of letting $t_\star = +\infty$ and $t^\star = -\infty$ if the feasible set is empty, in which case $\Theta^\star = \emptyset$.
[28] This second step would not be needed if the model were point identified.

4 Demand in Covered California

4.1 Data

Our primary data are administrative records on the universe of households who purchased a plan through Covered California in 2014. The data contain unique person and household identifiers for each individual in each household, as well as their age, income measured in percentage of FPL, gender, zipcode of residence, and choice of plan. We focus on the subpopulation of subsidy-eligible households (100-400% FPL) in which the uninsured members consist of either one or two adults aged 27 and older.[29] In addition, we drop the relatively small number of purchasing households with income under 140% of FPL, since these households are typically eligible for public health insurance (Medi-Cal, the California Medicaid program). These restrictions reduce our analysis sample to 630,924 of the 877,365 households who purchased coverage. Of these, 436,224 are singles, and 194,700 are couples.

We characterize each household by their resident rating region, M_i, and a vector X_i of observables consisting of their age, income, and distinguishing between singles and couples. Household age takes 38 unit values between 27 and 64, defined as the age of a single household member, or the (rounded) average age of a couple. We discretize household income into 52 bins of 5% of the FPL. (These bins are $[140, 145), [145, 150), \ldots, [395, 400]$.) When crossed with the 19 rating regions in Covered California, this yields 59,176 unique rating region $\times$ household type $\times$ age $\times$ income bins of the observable characteristics, (M_i, X_i). Since the number of households per bin varies greatly by region, we will report parameters that average over (M_i, X_i), and therefore put greater weight on larger geographic markets.

We reconstruct the post-subsidy per-person premiums faced by each household (P_{ij}) using their demographic information together with knowledge of insurers' base prices. As described in Section 2, individual pre-subsidy premiums for a given metal tier, rating region, and insurer only vary by age, while the post-subsidy premiums also vary by income and household structure. As a consequence, $P_i \equiv \pi(M_i, X_i)$ is a deterministic function of (M_i, X_i).

Our analysis is focused on a household's choice of coverage level (metal tier). The implicit assumption here is that a household's choice of coverage level is separable from their choice of insurer. We view this as a reasonable assumption for Covered California because the regulations ensure that the metal tiers offered—as well as the characteristics of the tiers—do not vary by insurer. We define premiums for each tier

[29] That is, the household either is childless, or the children are insured through a public program such as Medi-Cal. The maximum age is 64, after which individuals become eligible for Medicare.

in each market by taking the median post-subsidy premium across insurers.[30]

As in most demand analysis, we do not directly observe individuals who chose the outside option, i.e. to not purchase a plan through Covered California. This means that we first need to transform data on quantities chosen for the inside choices into choice shares by estimating the size of the market. To do this, we use the 2013 American Community Survey public use file (via IPUMS, Ruggles et al., 2015) to estimate uninsurance rates conditional on (M_i, X_i). Our estimation procedure for this part closely follows those used by Finkelstein et al. (2017) and Tebaldi (2017). For more detail, see Appendix H.

Table 2 provides some summary statistics. We consider demand for insurance for each bin of observables (M_i, X_i), and this is the level at which we summarize the data. Each bin contains on average 32 enrollees. The average participation rate in Covered California is 32%, and varies widely across markets and demographics, with a standard deviation of 34%. Older and poorer buyers are significantly more likely to purchase coverage. The impact of the CSRs is evident in panel (b) of the table: Buyers with income below 200% face premiums of less than \$100 per month to purchase a Silver plan with actuarial value of 88% or more (recall Table 1). Likely as a consequence, a Silver plan is chosen by roughly 23% of such consumers—more than twice as many as the 10% of consumers with income over 250% FPL who face a more expensive and less generous Silver plan.

4.2 Identifying Assumptions

In this section, we describe our specific implementations of assumptions (6) and (7).

An insurer's primary decision in Covered California is the base price for each rating region and coverage level. This decision likely depends on differences in demand and cost specific to each rating region, for example due to the underlying socioeconomic or health characteristics of the residents in a region, or due to differences in hospitals of medical providers. These factors are unobserved in our data, so we will not assume that variation in premiums across regions is exogenous. That is, we define a market M_i to be a rating region, and we will not impose any restriction on how preferences (the density of valuations f) varies across markets.

Instead, we will assume—in a limited way—that preferences are invariant to changes in age and income. Since premiums vary with age due to the age-rating, and with income due to the premium subsidies, this will provide variation in premiums that we

[30] Our results are robust to the use of other measures such as the mean, minimum, and second-cheapest premium across insurers.

Table 2: *Summary Statistics*

Panel (a) : Data by region and household type

	Observations	Mean	St. Dev.	P-10	Median	P-90
Number of buyers	59,176	32.25	56.06	1	10	90
Age	59,176	44.83	11.34	29	45	60
Income (FPL%)	59,176	227.92	63.58	155	215	320
Household size	59,176	1.24	0.43	1	1	2
Takeup rate	59,176	0.33	0.34	0.03	0.18	1.00
Average premium paid	59,176	148.78	85.17	58	128	274

Panel (b) : Heterogeneity in premiums and market shares

	Bronze		Silver		Gold		Platinum	
	Premium	Share	Premium	Share	Premium	Share	Premium	Share
By age:								
27-34	88	0.051	143	0.098	198	0.011	241	0.012
35-49	84	0.054	148	0.143	215	0.012	267	0.013
50-64	76	0.059	181	0.187	296	0.012	386	0.010
By income (FPL%):								
140-150	4	0.008	54	0.259	132	0.004	193	0.005
150-200	21	0.034	87	0.229	166	0.006	228	0.007
200-250	70	0.063	151	0.119	231	0.011	293	0.010
250-400	161	0.078	255	0.076	344	0.020	413	0.019
By household size:								
Singles	93	0.051	174	0.119	257	0.011	323	0.012
Couples	45	0.068	115	0.244	196	0.016	259	0.012

Note: Each observation in panel (a) is a unique combination of rating region × household type × age × income bins of the observable characteristics, (M_i, X_i). In panel (b), premium is calculated as the average premium paid across buyers of a given age/income group, while market shares are calculated as proportion of potential buyers.

can use to help identify demand counterfactuals. The way in which premiums evolve with age and income is prescribed by the ACA regulations, so the behavior of insurers is not likely to be an important threat to this strategy. Rather, our main concern is that valuations also change with age or income due to changes in latent risk factors or preferences. For this reason, we will use only local variation in age and income.

We formulate this approach using the notation of Section 3 by letting W_i denote a coarser aggregate of a group of X_i realizations. We group X_i into age bins given by {27–30, 31–35, 36–40, ..., 56–60, 61–64} and income bins given in percentage of FPL by {140–150, 150–200, 200–250, 250–300, 300–350, 350–400}. A value of W_i is then taken to be the market indicator M_i crossed between all possibilities of these coarser age-income bins. Conditioning on a value of W_i, we observe multiple premiums corresponding to variation in age and income within the W_i bin. Our assumption is that the distribution of latent valuations does not change as X_i varies within this coarser bin.

For example, one value of $W_i = w$ corresponds to the North Coast rating region, singles, ages between 36–40 and incomes between 150–200% of the FPL. Within this bin, we have 50 values of X_i, comprised of the 5 ages 36, 37, ..., 40 crossed with 10 income bins between 150 and 200 (in 5% steps). For each of these 50 values we observe a different premium vector. Since the variation we want to use is now in X_i, conditioning on a value of W_i, the notation we developed in Section 3 corresponds to taking $Z_i = X_i$. The assumption we use is now precisely (6) in that discussion, repeated here for emphasis:

$$\underbrace{f_{V|WZ}(v|w,z) = f_{V|WZ}(v|w,z') \quad \text{for all } z, z', w, \text{ and } v.}_{\text{within a coarse bin } (W_i = w), \text{ valuations are invariant to age and income } (z \neq z')} \tag{29}$$

The income aspect of assumption (29) gives empirical content to the separability between income and valuations in (1). The implied behavioral restriction is that a change in income that leaves a household inside a given W_i bin would not affect their choice of metal tier, although it could lead a household to change to or from the outside option. That is, the assumption is that there are no income effects with respect to the choice of metal tier within a coarse income bin. This assumption becomes weaker as the width of the W_i bins gets smaller.

The other aspect of (29) is the invariance to age. We are more concerned about this assumption, since health risks certainly increase in age, and likely at an increasing rate. We begin with (29) primarily for ease of interpretability. In Section 4.4, we relax assumption (6)/(29) to a strictly weaker "imperfect instrument" assumption that allows for some variation with age. Our estimated bounds there are wider, but more credible.

The other assumption we utilize is the verticality assumption (7), adjusted to account for CSRs as shown in Table 1. To account for the CSRs, we have chosen the coarse bins (W_i) so as not to cross the CSRs thresholds of 150, 200 and 250% of the FPL. We impose a verticality assumptions only when there are dominance relationships between all the characteristics of two products (as it is always the case for Bronze and Silver, for example), but avoid to assume any vertical ordering otherwise. In no case do we assume that any of the plans are preferred to the outside option.[31]

[31] For consumers with income above 250% FPL, we assume that for equal prices everyone would prefer

4.3 Results

Our focus is on measuring the effect of an equal change in post-subsidy premiums for all consumers on demand, consumer surplus, and government subsidy expenditure. We do not model supply, so all of our results should be interpreted as holding supply fixed. Integrating our nonparametric methodology with a model of supply-side behavior is a promising avenue for future research.

We consider counterfactual premium vectors of the form $\pi(M_i, X_i) + \delta$, for various choices of δ. That is, the counterfactuals we consider can be represented as the impact of shifting every households' price from the observed price, $P_i \equiv \pi(M_i, X_i)$ to a counterfactual price, $P_i^\star \equiv \pi(M_i, X_i) + \delta$. For each value of W_i, we construct the MRP using the set formed from all P_i and $P_i^\star$.

Figure 3 illustrates Bronze and Silver observed and counterfactual prices for buyers with income lower than 250% of the FPL. We use this case for illustration since it can be plotted on the plane, and because most buyers (over 93%) in this income range choose Bronze and Silver, presumably due to the CSR subsidies. In Figure 3b, the counterfactual is an increase in the price of the Bronze plan by \$10 for all consumers, while Figure 3c illustrates the analogous change in the Silver plan. In Figure 3d, both the Bronze and the Silver plan are increased by \$10, which from the consumer's perspective would be equivalent to a \$10 reduction in premium subsidies (or, more correctly, an increase in the maximum affordable amount holding premiums fixed) if Bronze and Silver were the only two choices.[32]

The first set of target parameters we consider is the change in choice shares for each good. For market m, consumer characteristics x, and good j, this can be written as

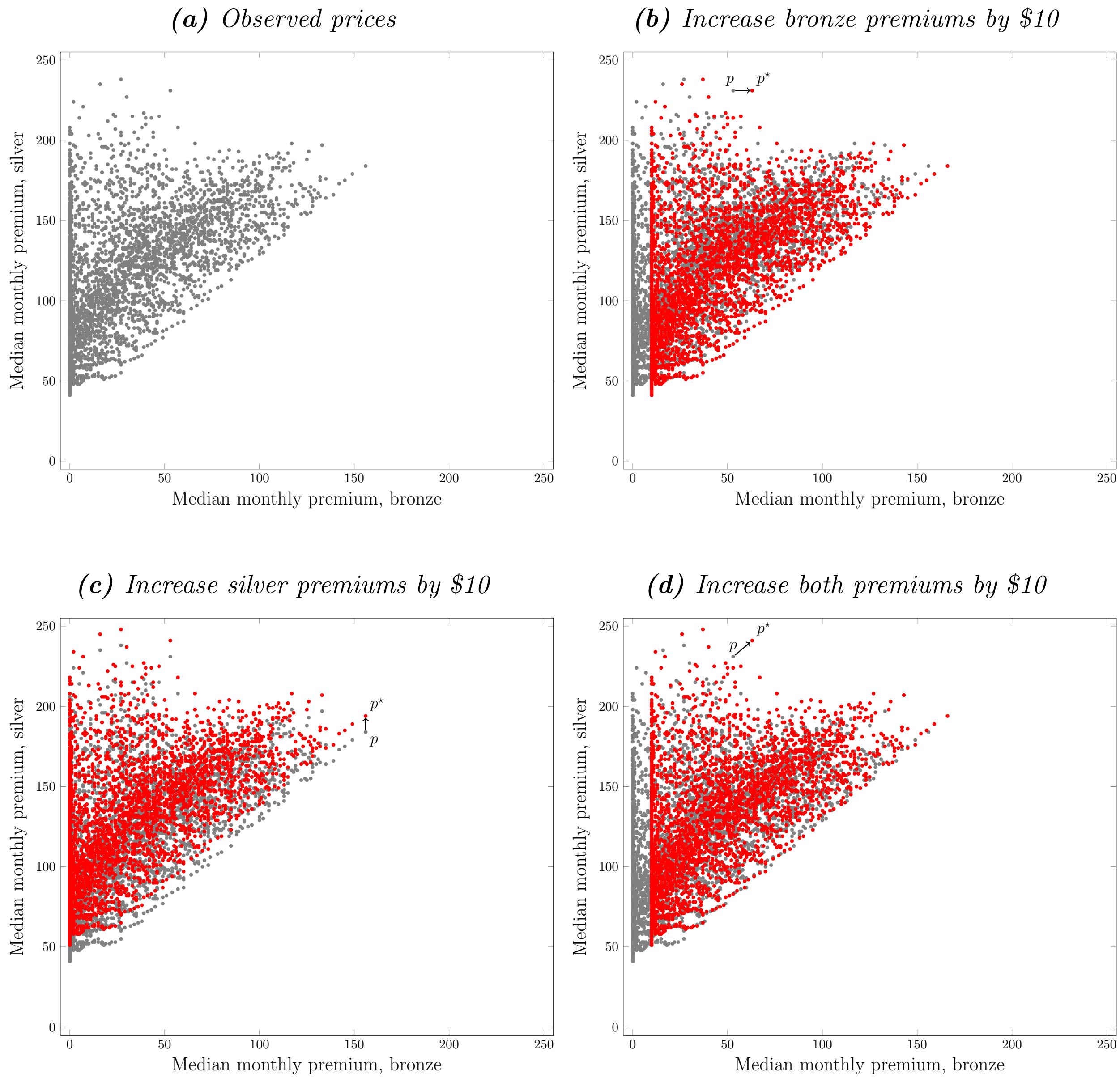
$$\Delta\text{Share}_j(f|m, x) \equiv \int_{\mathcal{V}_j(\pi(m,x)+\delta)} f(v|m, x)\, dv - \int_{\mathcal{V}_j(\pi(m,x))} f(v|m, x)\, dv, \tag{30}$$

where $\mathcal{V}_j(p)$ was defined in (9). Note that in (30), we omit the dependence on the price change, δ, since this will be clear from the way we present our results. In order

Platinum over Gold, Gold over Silver, and Silver over Bronze. Below 150, we assume that Silver is preferred to Platinum, which is itself preferred over Gold, and Gold over Bronze. In the 150–200% FPL range, we assume that Silver is preferred to Bronze, and that Platinum is preferred to Gold, and Gold to Bronze. According to Table 1, we assume no ordering between Silver and Gold, nor Silver and Platinum. Lastly, for households with income between 200 and 250% FPL, we assume that Platinum is preferred to Gold, which is preferred to Bronze, and that Platinum is preferred to Silver, which is preferred to Bronze. We do not assume any ordering between Silver and Gold.

[32] In this example with $J = 2$, $j = 1$ denoting Bronze, and $j = 2$ denoting Silver, Figures 3b, 3c, and 3d would correspond to taking $\delta = (10, 0)$, $\delta = (0, 10)$ and $\delta = (10, 10)$, respectively, where the price of the outside option is always fixed at 0.

Figure 3: *Observed and Counterfactual Prices*



Note: *The figure shows observed and counterfactual prices of Bronze and Silver plans for households with income between 140-250% of the FPL. Panel (a) plots the prices observed in the data in grey, where each observation is a unique region-age-income combination. Panel (b) overlays in red the counterfactual prices representing an increase in $10 per person, per month for Bronze premiums. Panel (c) is like Panel (b), but the price increases are for Silver premiums. Panel (d) is like Panels (b) and (c) with price increases of $10 for both Silver and Bronze premiums.*

Table 3: Substitution Patterns

$10/month premium increase for	Bronze LB	Bronze UB	Silver LB	Silver UB	Gold LB	Gold UB	Platinum LB	Platinum UB	Any plan LB	Any plan UB
	Change in probability of choosing									
Panel (a): 140 - 400% FPL										
Bronze	-0.048	-0.012	+0.004	+0.041	+0.000	+0.024	+0.000	+0.020	-0.017	-0.004
Silver	+0.002	+0.105	-0.169	-0.029	+0.000	+0.123	+0.000	+0.098	-0.059	-0.008
Gold	+0.000	+0.006	+0.000	+0.010	-0.013	-0.002	+0.000	+0.011	-0.004	-0.000
Platinum	+0.000	+0.005	+0.000	+0.008	+0.000	+0.012	-0.012	-0.002	-0.004	-0.000
All plans	-0.020	-0.007	-0.062	-0.022	-0.005	-0.001	-0.005	-0.001	-0.084	-0.033
Panel (b): 140 - 250% FPL										
Bronze	-0.041	-0.010	+0.004	+0.035	+0.000	+0.020	+0.000	+0.016	-0.014	-0.002
Silver	+0.002	+0.132	-0.221	-0.036	+0.000	+0.159	+0.000	+0.126	-0.079	-0.012
Gold	+0.000	+0.004	+0.000	+0.008	-0.010	-0.002	+0.000	+0.009	-0.003	-0.000
Platinum	+0.000	+0.003	+0.000	+0.007	+0.000	+0.010	-0.010	-0.002	-0.003	-0.000
All plans	-0.016	-0.005	-0.082	-0.030	-0.004	-0.001	-0.004	-0.001	-0.099	-0.039
Panel (c): 250 - 400% FPL										
Bronze	-0.061	-0.016	+0.005	+0.052	+0.000	+0.033	+0.000	+0.026	-0.024	-0.006
Silver	+0.001	+0.055	-0.073	-0.015	+0.000	+0.056	+0.000	+0.044	-0.021	-0.002
Gold	+0.000	+0.009	+0.000	+0.014	-0.017	-0.003	+0.000	+0.016	-0.006	-0.000
Platinum	+0.000	+0.007	+0.000	+0.010	+0.000	+0.015	-0.016	-0.003	-0.006	-0.000
All plans	-0.027	-0.010	-0.023	-0.009	-0.008	-0.001	-0.007	-0.001	-0.056	-0.024

to aggregate (30) into a single measure, we average it over markets and covariates:

$$\Delta\text{Share}_j(f) \equiv \sum_{m,x} \Delta\text{Share}_j(f|x,m)\,\mathbb{P}[M_i = m, X_i = x]. \tag{31}$$

We will average other parameters in an analogous way. In the notation of Section 3, ΔShare_j is an example of a target parameter, θ.

Table 3 reports estimated bounds for ΔShare_j across the four metal tier choices together with bounds on overall participation, i.e. $1 - \Delta\text{Share}_0$. The rows of Table 3 reflect different types of premium increases, δ. The nominal premium increase is taken to be \$10 per person, per month, which represents a moderate to large price increase for many consumers (see Table 2). Our estimated bounds are quite informative. For example, for the full sample in panel (a), we estimate that a simultaneous $10 increase in all premiums reduces the proportion of households that purchase coverage by between 3.3 and 8.4%. Panel (b) shows that these estimates are larger in magnitude for low-income households, at between 3.9 and 9.9%, and panel (c) shows that they are smaller in magnitude for higher-income households, who we estimate would reduce participation in Covered California by between 2.4 and 5.6%. Comparing panels (b) and (c) more generally, we find a pattern of higher price sensitivity for low-income households.

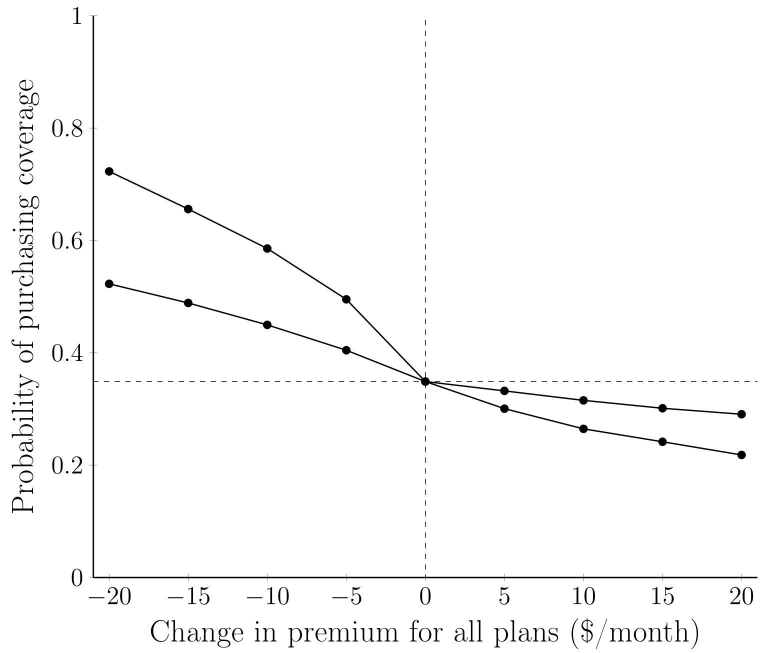
The other columns of Table 3 measure substitution patterns within and between coverage tiers. For example, panel (a) shows that an increase in Bronze premiums by $10 per person, per month would lead to a decrease of between 1.2 and 4.8% in the share of consumers choosing the Bronze plan, and an increase in the share choosing Silver of between 0.4 and 4.1%. The increase in the share choosing Gold or Platinum is significantly smaller, reflecting the closer substitutability of the Bronze and Silver plans. The extensive margin change of participation for a Bronze premium increase is between 0.4 and 1.7%, which is naturally both smaller and tighter than the change when all premiums are increased together. In contrast, increasing Platinum premiums by the same amount would lead to a much smaller decline in the proportion of buyers not purchasing coverage. Overall, Table 3 indicates substitution patterns inconsistent with the independence of irrelevant alternatives property of the logit model.

One consequence of adopting a partial identification framework is that the amount of information that the data and assumptions yield about a specific counterfactual quantity is reflected in the width of the bounds. The bounds for more ambitious (more distant) counterfactuals will be wider than for more modest counterfactuals that are closer to what was observed in the data. This situation is evident in Figure 4, which plots the average extensive margin (enrollment) response as a function of a given increase or decrease in all premiums. Our bounds are relatively tight for small changes in premiums, and then widen as the premiums get farther from what was observed in the data. We consider this an attractive feature of our approach, since it reflects the increasing difficulty of drawing inference about objects that involve larger departures from the observed data, and so captures an important dimension of model uncertainty. In contrast, a fully parametric model point identifies any counterfactual quantity regardless of how distant the extrapolation involved.[33]

The second set of parameters we consider measure the effects of changing premium

[33] Note that confidence intervals on point estimates from a parametric model will tend to widen as one extrapolates further. However, for the parametric models we consider in Section 5, the width of these confidence intervals is basically zero even for distant extrapolations.

Figure 4: Extensive Margin Demand Changes for Different Counterfactuals



subsidies on consumer surplus and government spending. From the household's perspective, a decrease in premium subsidies—which in terms of policy can be thought of as an increase in the maximum affordable amount—is the same as an increase in premiums faced.[34] Such a subsidy change generates an average change in consumer surplus for a household in market m with characteristics x of

$$\Delta\mathrm{CS}(f|m,x) \equiv \int \left[\max_{j\in\mathcal{J}} \{v_j - \pi_j(m,x) - \delta_j\} - \max_{j\in\mathcal{J}} \{v_j - \pi_j(m,x)\} \right] f(v|m,x)dv,$$

which we aggregate by averaging over markets and demographics into

$$\Delta\mathrm{CS}(f) \equiv \sum_{m,x} \Delta\mathrm{CS}(f|m,x)\, \mathbb{P}[M_i = m, X_i = x].$$

We will contrast the change in consumer surplus to the change in government spending

[34] Our analysis here requires maintaining a partial equilibrium framework in which there are no other supply side responses in base prices due to an adjustment in subsidy schemes. As noted above, integrating our approach with a model of insurance supply is an interesting avenue for future research.

on premium subsidies. This is given by

$$\Delta\mathrm{GS}(f|m,x) \equiv \sum_{j>0} (\mathrm{sub}_j(m,x) - \delta_j) \times \left[\int_{\mathcal{V}_j(\pi(m,x)+\delta)} f(v|m,x)\, dv \right] - \sum_{j>0} \mathrm{sub}_j(m,x) \times \left[\int_{\mathcal{V}_j(\pi(m,x))} f(v|m,x)\, dv \right],$$

where $\mathrm{sub}_j(m,x)$ denotes the baseline premium subsidy for purchasing plan j. We denote aggregated government spending as

$$\Delta\mathrm{GS}(f) \equiv \sum_{m,x} \mathrm{GS}(f|m,x)\, \mathbb{P}[M_i = m, X_i = x].$$

Both ΔCS and ΔGS are other examples of target parameters θ.[35]

Figure 5 depicts our bounds on ΔCS for a \$10 decrease in subsidies as the shaded areas between the two demand curves. The lower bound on the change in consumer surplus is the area to the left of the flatter demand curve, while the upper bound also includes the entire area to the right of the steeper demand curve. Intuitively, the lower bound is attained at the upper bound (smallest magnitude) of price elasticity for the extensive margin, while the upper bound of the change in consumer surplus is attained at the lower bound (largest magnitude) of this price elasticity. Note that while the bounds on ΔCS shown here are sharp and unique, the demand curves we have plotted are not, since there are many ways to draw a demand curve up to a \$10 premium increase that can yield the same area to the left, while still respecting the data and assumptions.

Table 4 tabulates the estimated bounds on ΔCS for the same \$10 decrease in premium subsidies. The first column shows estimated bounds using the entire sample, while the second and third columns split the estimates on income. In the fourth column of Table 4, we report bounds on the corresponding reduction in government spending that results from the lower subsidies. We estimate these by fixing average consumer surplus at its lower or upper bound, then solving for the bounds on government spending that could be realized for this consumer surplus change.[36]

Our bounds imply that a \$10 decrease in monthly subsidies would lead to a reduction in average monthly consumer surplus of between \$1.91 and \$2.40 per person. The

[35] In Appendix F, we show how to construct sharp bounds on ΔCS by deriving corresponding $\overline{\theta}$ functions that are linear in ϕ.

[36] We do this because a given consumer surplus change could be attained in a variety of different ways, each of which might be associated with different changes in government spending.

Figure 5: *Change in Consumer Surplus Resulting from a Change in Premiums*

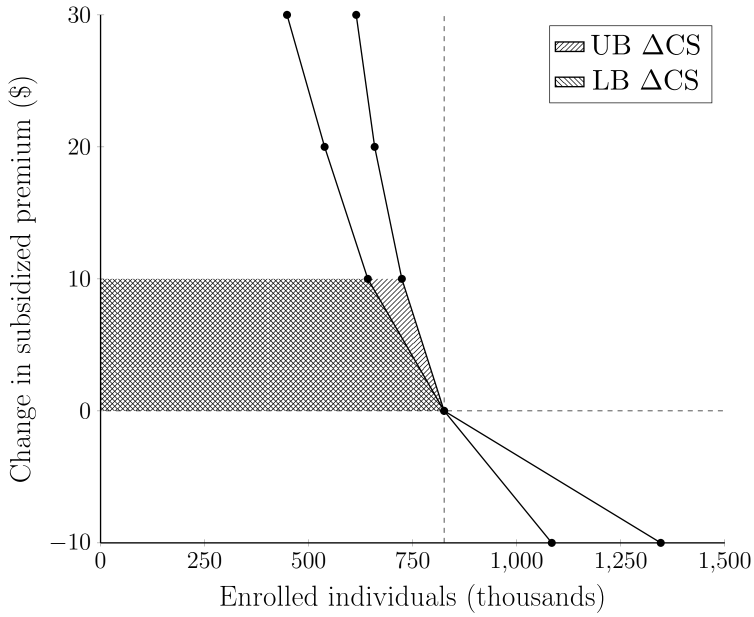


Table 4: *The Impacts of Reducing Monthly Subsidies by $10*

	140 - 400% FPL Change in consumer surplus		140 - 250% FPL Change in consumer surplus		250 - 400% FPL Change in consumer surplus		140 - 400% FPL Associated change in subsidy outlets	
	LB	UB	LB	UB	LB	UB	LB	UB
Average ($/person-month)	-2.397	-1.905	-2.881	-2.317	-1.500	-1.142	-26.125	-14.995
Aggregate ($ million/year)	-70.485	-56.012	-56.036	-45.059	-16.434	-12.506	-768.142	-440.884

impacts for the lower-income sample ($2.32–$2.88) are estimated to be approximately twice as large as the impacts for the higher-income sample ($1.14–$1.5). This reflects the fact that individuals with income lower than 250% of the FPL have a higher up-take of insurance and are covered under more generous plans due to the CSRs. Our estimates of changes in consumer surplus are dwarfed by the corresponding change in government expenditure on premium subsidies, which we estimate to be between $15.00 and $26.13 per consumer, per month. The large magnitude of the expenditure savings is due to the large number of marginal buyers who exit the market due to the post-subsidy premium increase. When these buyers exit, they relinquish their entire premium subsidy, which in most cases is significantly greater than $10.

The bottom row of Table 4 shows the aggregate yearly impact of a $10 reduction in subsidies in Covered California. The total consumer surplus impact would be between \$56 and \$70 million, with the majority of the losses concentrated among households with income below 250% of the FPL. At the same time, government subsidy outlays would decline by between \$441 and \$768 million per year. Overall, our findings suggest that consumers value health insurance significantly less than it would cost in premium subsidies to induce them to purchase a plan. This finding is consistent with a growing number of empirical analyses, see e.g. Finkelstein et al. (2017). In interpreting this finding, we caution that our estimates do not account for the existence of potentially large externalities such as the cost of uncompensated care, debt delinquency, or bankruptcy (Finkelstein et al., 2012; Mahoney, 2015; Garthwaite, Gross, and Notowidigdo, 2018).

4.4 Allowing Valuations to Change Within Coarse Age Bins

The primary assumption that drives our results is (29). As we noted, the part of this assumption that imposes independence between valuations and age within coarse age bins is probably questionable, since valuations likely change with risk factors, and risk factors change with age.[37] In this section, we consider a strictly weaker version of (29) that allows for deviations away from perfect invariance. This can be viewed as a sensitivity analysis, and is similar in spirit to proposals by Conley, Hansen, and Rossi (2010), Nevo and Rosen (2012), and Manski and Pepper (2017).

The way in which we do this is to relax (29) into two inequalities controlled by a slackness parameter. The relaxed assumption is that

$$(1-\kappa(z,z'))f_{V|WZ}(v|w,z') \leq f_{V|WZ}(v|w,z) \leq (1+\kappa(z,z'))f_{V|WZ}(v|w,z') \quad \text{for all } z, z', w, \text{ and } v, \tag{32}$$

where $\kappa(z,z') \geq 0$ is the slackness parameter. We specify κ in the following way:

$$\kappa(z,z') = \begin{cases} \kappa, & \text{if } z \text{ and } z' \text{ differ only in age, and only by a single bin} \\ 0, & \text{if } z \text{ and } z' \text{ differ only in income} \\ +\infty, & \text{otherwise.} \end{cases}$$

In words, the assumption is that within any coarse bin (i.e., conditional on $W_i = w$), the pointwise difference in conditional valuation densities corresponding to any two adjacent two-year age bins (with identical income) can be no greater than κ%. The constant κ must be chosen, but we will report results for various choices. Taking $\kappa = 0$

[37] Indeed, the importance of age heterogeneity in health insurance demand is the emphasis of existing work, see e.g. Ericson and Starc (2015), Geruso (2017), and Tebaldi (2017).

Table 5: *Allowing for Valuations to Vary Within Coarse Age Bins*

Allowed variation in preference with age	Change in probability of not enrolling if all per-person premiums increase by \$10/month LB	UB	Change in consumer surplus (\$/person-month) if per-person subsidies decrease by \$10/month LB	UB	Change in government spending (\$/person-month) if per-person subsidies decrease by \$10/month LB	UB
$\kappa = 0$	+0.033	+0.084	-2.397	-1.905	-26.125	-14.995
$\kappa = 0.2$	+0.034	+0.086	-2.450	-1.941	-26.504	-14.909
$\kappa = 0.4$	+0.034	+0.088	-2.481	-1.966	-26.991	-15.032
$\kappa = 0.8$	+0.035	+0.092	-2.541	-1.998	-28.339	-14.825
$\kappa = 1$	+0.033	+0.098	-2.574	-1.951	-29.374	-14.169
$\kappa = 3$	+0.032	+0.104	-2.641	-1.952	-31.027	-13.496
$\kappa = +\infty$	+0.027	+0.114	-2.714	-1.846	-34.215	-11.632

reduces (32) back to our previous assumption of (29). Alternatively, taking $\kappa = +\infty$ completely relaxes the age restriction, so that the only variation we are using is with respect to income.

Table 5 reports bounds on some of our main target parameters under (29) for different values of κ.[38] The row with $\kappa = 0$ are the same as the estimates reported in the previous section, $\kappa = +\infty$ corresponds to estimates that only use variation in income. We include a variety of choices of κ so that the reader can make their own judgment, where for example $\kappa = 0.4$ can be interpreted as allowing for a change of up to 40% in valuations between two adjacent age groups.[39]

Overall, our findings are robust to allowing increasing dependence between valuations and age, including the case in which this dependence is unrestricted ($\kappa = +\infty$). In this case, we estimate that reducing monthly premium subsidies by \$10 leads to a drop in coverage between 2.7% and 11.4%, to a corresponding reduction in con-

[38] Note that it is straightforward to modify the sharp characterization in Proposition 1 to allow for an assumption like (29) instead of (6). The difference in implementation just amounts to replacing (26) with an inequality analogous to (29).

[39] Note that the bounds generally widen with κ, since larger values correspond to weaker assumptions. However, this is not always the case, due to the fact that we are estimating these bounds using the procedure in Appendix G. Essentially, that procedure works by restricting attention to densities that come close to fitting the observed choice shares the best. This fit mechanically improves as κ increases, because more densities are considered. As a result, densities that seemed to fit well for smaller values of κ might no longer be deemed to fit well when κ increases, since the best fit has improved. This creates a countervailing effect to changing κ, which can lead to non-monotonicity in the estimated bounds even though monotonicity must hold for the population bounds.

sumer surplus between \$1.85 and \$2.71 per person, per month, and to a reduction in government spending between \$11.63 and \$34.22.

5 Estimates from Parametric Models

The motivation of this paper has been to provide estimates of key policy parameters using a model that does not use parametric distributional assumptions. In this section, we compare our nonparametric bounds to estimates from some fully parametric logit and probit models which do use such assumptions. These models all follow a specification similar to (2):

$$Y_{im} = \arg\max_{j \in \mathcal{J}} \mathbb{1}[j \geq 1] \left(\gamma_{im} + \beta_{im}\text{AV}_{ijm} - \alpha_{im}P_{ijm} + \xi_{jm}\right) + \epsilon_{ijm}, \tag{33}$$

where γ_{im} is an individual-specific (and market-specific) intercept, and AV_{ijm} is the actuarial value of tier j for individual i in market m (see Table 1; in Covered California this characteristic does not vary with m). Additionally, ξ_{jm} represents the shock to preferences due to unobservable characteristics specific to tier j in market m. These may represent, for example, differences in the set of participating insurers, differences in the set of medical providers, or differences in health risk or risk attitudes of individuals residing in market m. The presence of the indicator sets the contribution of these terms to 0 for the outside option ($j = 0$). The logit class of models restrict ϵ_{ijm} to follow a type I extreme value distribution, independently across j, while probit models restrict ϵ_{ijm} to follow a standard normal distribution.

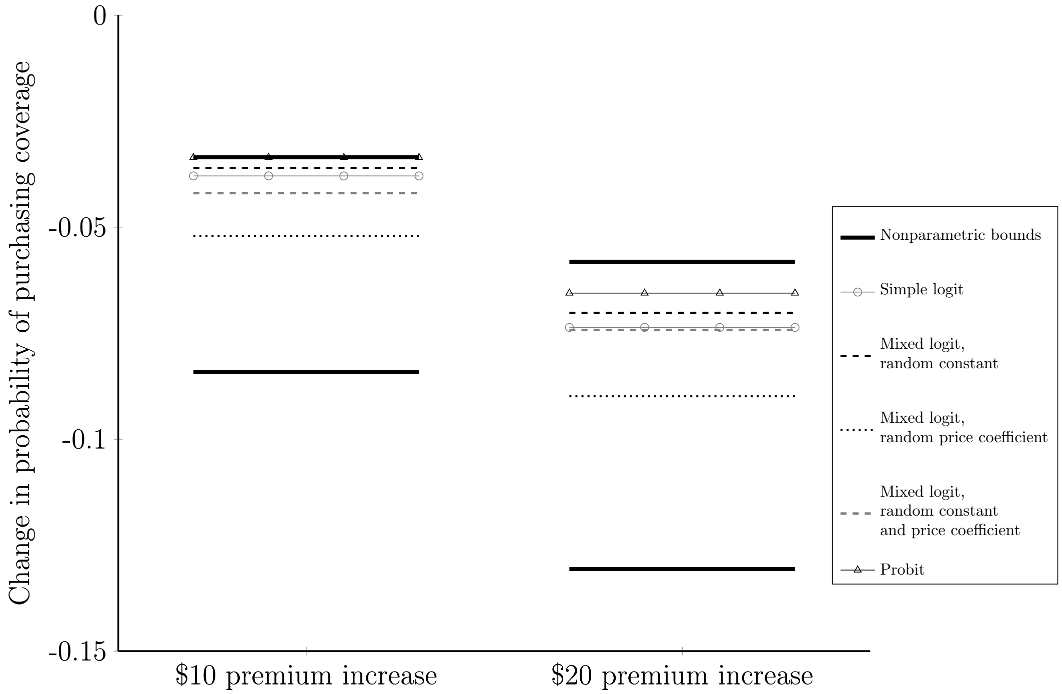
The first model we estimate is a logit in which the price parameter, α_{im}, is constant within a market, similarly to ξ_{jm}, but both γ_{im} and β_{im} vary with observables in a rich way.[40] The second model is a probit with the same specification.[41] We then consider three mixed logit models. In all of these models, γ_{im} and β_{im} vary with observables as in the baseline model, and the premium coefficient α_{im} still varies with the market. The three models differ in whether γ_{im}, α_{im}, or both have an additional unobservable component that is normally distributed with unknown variance. In the latter case, we also assume that the two unobservable components are uncorrelated.

Figure 6 illustrates how our nonparametric bounds on the extensive margin re-

[40] The specification allows β_{im} to vary freely by market with a different value in each of the following four age bins: {27–34, 35–44, 45–54, 54–64}. It allows γ_{im} to also vary freely by market, and within each region restricts $\gamma_{im} = \gamma_{im}^{\text{Inc}} + \gamma_{im}^{\text{Age}}$, where γ_{im}^{Inc} varies in three FPL income bins {140–200, 200–250, 250–400}, and γ_{im}^{Age} varies in the same four age bins as β_{im}.

[41] We have had difficulty estimating a similar probit with correlated ϵ_{ijm} because the likelihood is very flat, suggesting a potential failure of point identification.

Figure 6: *Comparison Between Nonparametric Bounds and Parametric Estimates*



sponses compare to the estimates one obtains from these five parametric models. The estimates shown are for the counterfactuals of a \$10 and \$20 increase in all premiums (or decrease in subsidies). All of the point estimates are within the nonparametric bounds, but clustered near the upper bound, where price sensitivity is smallest. The one exception, with point estimates close to the midpoint of our nonparametric bounds, is the mixed logit in which the premium coefficient α_{im} varies across individuals within a market. The implication is that different distributional assumptions on ϵ_{ijm} other than logit and probit could yield estimates near the lower bound, while still preserving the same degree of fit to the observed choice shares. As we showed in Table 4, these estimates would have substantially different policy implications in terms of consumer surplus and government spending. Thus, the assumption of a type I extreme value (or similarly-shaped normal) distribution appears here to have a significant impact on the empirical conclusions that would be drawn.

6 Conclusion

We estimated the demand for health insurance in California's ACA marketplace using a new nonparametric methodology. While we designed our methodology with health insurance in mind, it should be applicable to other discrete choice problems as well. The central idea of the method is to divide realizations of a consumer's valuations into sets for which behavior remains constant. We showed how to define the collection of such sets, which we referred to as the minimal relevant partition (MRP) of valuations. Using the MRP, we developed a computationally reliable linear programming procedure for consistently estimating sharp identified sets for target parameters of interest.

Our estimates of demand using this methodology point to the possibility of substantially greater price sensitivity than would be recognized using comparable parametric models. This is consistent with the commonly-heard folklore that logits are "flat" models. We showed that this finding has potentially important policy implications, since it implies that the impact of decreasing subsidies on consumer surplus could be much smaller—and the impact of government expenditure much larger—than would be recognized using standard parametric methods. More broadly, our results provide a clear example in which functional form assumptions are far from innocuous, and actually play a leading role in driving empirical conclusions.

A Methodology Literature Review

In this section, we discuss the relationship of our methodology to the existing literature. We focus our attention first on semi- and non-parametric approaches to unordered discrete choice analysis. This literature can be traced back to Manski (1975). The focus of Manski's work, as well as most of the subsequent literature, has been on relaxing parameterizations on the distribution of unobservables, while the observable component of utility is usually assumed to be linear-in-parameters.[42] The motivation of our approach is also to avoid the need to parameterize distributions of latent variables, however we have chosen to keep the entire analysis nonparametric.[43]

Our approach has three key properties that, when taken together, make it distinct in the literature on semi- and nonparametric discrete choice. First, much of the literature has focused on identification of the observable components of indirect utility, while treating the distribution of unobservables as an infinite-dimensional nuisance parameter. For example, in (2), this would correspond to identifying α_i and β_i when these random coefficients are restricted to be constant. Examples of work with this focus include Manski (1975), Matzkin (1993), Lewbel (2000), Fox (2007), Pakes (2010), Ho and Pakes (2014), Pakes, Porter, Ho, and Ishii (2006, 2015), Pakes and Porter (2016), and Shi, Shum, and Song (2016). Identification of the relative importance of observable factors for explaining choices is insufficient for our purposes, because the policy counterfactuals we are interested in, such as choice probabilities and consumer surplus, also depend on the distribution of unobservables. Treating this distribution as a nuisance parameter would not allow us to make sharp statements about quantities relevant to these counterfactuals.

Second, we allow for prices (premiums in our context) to be endogenous in the sense of being correlated with the unobservable determinants of utility. This differentiates our paper from work that focuses on identification of counterfactuals, but which assumes exogenous explanatory variables. Examples of such work includes Thompson (1989), Manski (2007, 2014), Briesch, Chintagunta, and Matzkin (2010), Chiong, Hsieh, and Shum (2017), and Allen and Rehbeck (2017). The importance of allowing for endogenous explanatory variables in discrete choice demand analysis was emphasized by

[42] Matzkin (1991) considered the opposite case in which the distribution of the unobservable component is parameterized, but the observable component is treated nonparametrically. See also Briesch, Chintagunta, and Matzkin (2002).

[43] Extending our methodology to a semiparametric model is an interesting avenue for future work, but not well-suited to our application since there is no variation in choice (plan) characteristics in Covered California. Conceptually though, one could use our strategy with a semiparametric model by fixing the parametric component and then repeatedly applying our characterization argument, similar to the strategy in Torgovitsky (2018).

Berry (1994) and Hausman, Leonard, and Zona (1994), and motivated the influential work of Berry et al. (1995, 2004a). In our application, it is essential that we can make statements about demand counterfactuals while still recognizing that premiums could be dependent with unobservable valuations.

This leads us to the third way that our approach differs from existing literature, which is that we do not place strong demands on the available exogenous variation in the data. In particular, we do not require the existence of a certain number of instruments, or that such instruments satisfy strong support or rank conditions. For example, Lewbel (2000) and Fox and Gandhi (2016) require exogenous "special regressors" with large support, which are not available in our data. Alternatively, Chiappori and Komunjer (2009) and Berry and Haile (2014) provide identification results that require a sufficient number of continuous instruments that satisfy certain "completeness" conditions, which can be viewed as high-level analogs to traditional rank conditions.[44] Besides the difficulty of finding a sufficient number of continuous instruments, one might also be concerned with the interpretability and/or testability of the completeness condition (Canay, Santos, and Shaikh, 2013). Not maintaining these types of support and completeness conditions leads naturally to a partial identification framework.

Other authors have also considered taking a partial identification approach to unordered discrete choice models. Pakes (2010), Ho and Pakes (2014), Pakes et al. (2006, 2015), Pakes and Porter (2016) developed moment inequality approaches that can be used to bound coefficients on observables in specifications like (2) without parametric assumptions on the unobservables. As noted, this is insufficient for our purposes, since we are concerned with demand counterfactuals. Manski (2007), Chiong et al. (2017) and Allen and Rehbeck (2017) bound counterfactuals, but assume that all explanatory variables are exogenous. In parametric contexts, Nevo and Rosen (2012) have considered partial identification arising from allowing instruments to be partially endogenous, and Gandhi, Lu, and Shi (2017) treated the problem of non-purchases in scanner data as one of partial identification.

On a more specific technical level, our work is related to a literature on computational approaches to characterizing identified sets in the presence of partial identification. In particular, the linear programming structure we exploit has been noted by many other authors, see e.g. Balke and Pearl (1994, 1997) and Hansen, Heaton, and Luttmer (1995) for early examples. Previous work that has implemented linear programming to characterize sharp identified sets includes Honoré and Tamer (2006), Honoré and Lleras-Muney (2006), Manski (2007, 2014), Lafférs (2013), Freyberger and

[44] See also Compiani (2018), who has shown how to construct and implement estimators based on the results of Berry and Haile (2014).

Horowitz (2015), Demuynck (2015), Kline and Tartari (2016), Torgovitsky (2016, 2018), Kamat (2017), and Mogstad, Santos, and Torgovitsky (2018). Of this work, ours is closest to Manski (2007), who also considered discrete choice problems. Methodologically, our work differs from Manski's because we maintain and exploit more structure on preferences (via (1)), and in addition we do not assume that explanatory variables (or choice sets in Manski's framework) are exogenous.

B A Model of Insurance Choice

In this section, we provide a model of choice under uncertainty for a risk averse consumer which leads to (1). The model is quite similar to those discussed in Handel (2013, pp. 2660–2662) and Handel et al. (2015, pp. 1280–281). Throughout, we suppress observable factors other than price (components of X_i) that could affect a consumer's decision. All quantities can be viewed as conditional on these observed factors, which is consistent with the nonparametric implementation we use in the main text.

Suppose that each consumer i chooses a plan j to maximize their expected utility taken over uncertain medical expenditures, so that

$$Y_i = \arg\max_{j \in \mathcal{J}} \int U_{ij}(\text{ex})\, dF_{ij}(\text{ex}), \tag{34}$$

where $U_{ij}(\text{ex})$ is consumer i's ex-post utility from choosing plan j given realized expenditures of ex, and F_{ij} is the distribution of these expenditures, which varies both by consumer i (due to risk factors) and by plan j (due to coverage levels). Assume that U_{ij} takes the constant absolute risk aversion (CARA) form

$$U_{ij}(\text{ex}) = -\frac{1}{A_i} e^{-A_i C_{ij}(\text{ex})}, \tag{35}$$

where A_i is consumer i's risk aversion, and $C_{ij}(\text{ex})$ is their ex-post consumption when choosing plan j and realizing expenditures ex. We assume that ex-post consumption takes the additively separable form

$$C_{ij}(\text{ex}) = \text{Inc}_i - P_{ij} - \text{ex} + \widetilde{V}_{ij}, \tag{36}$$

where Inc_i is consumer i's income, P_{ij} is the price they paid for plan j, and $\widetilde{V}_{ij}$ is an idiosyncratic preference parameter.

Substituting (36) into (35) and then into (34), we obtain

$$Y_i = \arg\max_{j \in \mathcal{J}} -\frac{1}{A_i}\left[e^{A_i(P_{ij}-\mathrm{Inc}_i-\widetilde{V}_{ij})}\int e^{A_i \mathrm{ex}}\, dF_{ij}(\mathrm{ex})\right]$$

Transforming the objective using $u \mapsto -\log(-u)$, which is strictly increasing for $u < 0$, we obtain an equivalent problem

$$\begin{aligned} Y_i &= \arg\max_{j \in \mathcal{J}} -\log\left(\frac{1}{A_i}\left[e^{A_i(P_{ij}-\mathrm{Inc}_i-\widetilde{V}_{ij})}\int e^{A_i \mathrm{ex}}\, dF_{ij}(\mathrm{ex})\right]\right) \\ &= \arg\max_{j \in \mathcal{J}} -\log\left(\frac{1}{A_i}\right) + A_i\left(\mathrm{Inc}_i - P_{ij} + \widetilde{V}_{ij}\right) + \log\left(\int e^{A_i \mathrm{ex}}\, dF_{ij}(\mathrm{ex})\right). \end{aligned}$$

Eliminating additive terms that don't depend on plan choice yields

$$Y_i = \arg\max_{j \in \mathcal{J}} -A_i P_{ij} + A_i \widetilde{V}_{ij} + \log\left(\int e^{A_i \mathrm{ex}}\, dF_{ij}(\mathrm{ex})\right).$$

Suppose that $A_i > 0$, so that all consumers are risk averse.[45] Then we can express the consumer's choice as

$$Y_i = \arg\max_{j \in \mathcal{J}} \left[\widetilde{V}_{ij} + \frac{1}{A_i}\log\left(\int e^{A_i \mathrm{ex}}\, dF_{ij}(\mathrm{ex})\right)\right] - P_{ij},$$

which takes the form of (1) with

$$V_{ij} \equiv \left[\widetilde{V}_{ij} + \frac{1}{A_i}\log\left(\int e^{A_i \mathrm{ex}}\, dF_{ij}(\mathrm{ex})\right)\right].$$

Examining the components of V_{ij} reveals the factors that contribute to heterogeneity in valuations in this model. Heterogeneity across i can come from variation in risk aversion (A_i), from differences in risk factors or beliefs (F_{ij}), and from idiosyncratic differences in the valuation of health insurance ($\widetilde{V}_{ij}$). Differences in valuations across j arise from the interaction between risk factors and the distribution of corresponding expenditures (F_{ij}), as well as from idiosyncratic differences in valuations across plans ($\widetilde{V}_{ij}$). The main restrictions in this model are the assumption of CARA preferences in (35) and the quasilinearity of ex-post consumption in (36). However, as noted in the main text, it is important to realize that for these restrictions to have empirical content, they must be combined with an assumption about the dependence between income (here called Inc_i) and the preference parameters, A_i and $\widetilde{V}_{ij}$.

[45] Showing that (1) would arise from risk neutral consumers is immediate.

C Modifications for More or Less Price Variation

In Covered California, post-subsidy premiums are a deterministic function of the market (rating region) and consumer demographics. Our discussion in the main text was tailored to this case. In this section, we discuss how to modify our approach to settings in which prices vary either more or less.

The more straightforward (and probably less interesting) case is when P_i still varies conditional on (M_i, X_i). This could occur if prices vary at the individual level due to demographic or geographic variables that the researcher does not observe. In this case, our methodology can be applied with little more than notational changes. The conditioning on P_i would need to be carried along, since it is no longer redundant after conditioning on (M_i, X_i). Demand and consumer surplus parameters like (3) and (4) would be defined as before, but there would be an additional integration step to construct the density of V_i given (M_i, X_i) from that of V_i given (P_i, M_i, X_i). A similar comment applies to the assumptions in Section 3.4. The observational equivalence condition (11) would be modified so that it is defined for all (p, m, x).

The less straightforward (and more interesting) case is when one observes only a single price for each market, as in Berry et al. (1995); Berry and Haile (2014). Notationally, this means $P_i = \pi(M_i)$ depends on M_i only, and not X_i. As a technical matter, our methodology applies exactly as before to this case. However, since there is only a single price per market, and since we are not assuming anything about how demand varies across markets, the resulting bounds will be uninformative. Here, we suggest two additional assumptions that could potentially be used to compensate for limited price variation.

The first assumption is that there is another observable variable that varies within markets that can be made comparable to prices.[46] This is implicit in standard discrete choice models like (2). Consider modifying (1) to

$$Y_i = \arg\max_{j \in \mathcal{J}} V_{ij} + X_i' \beta_j - P_{ij}, \tag{37}$$

where $\beta \equiv (\beta_1, \ldots, \beta_J)$ are unknown parameter vectors. For each fixed β, this model is like (1) but with prices given by $\tilde{P}_{ij}(\beta_j) \equiv P_{ij} - X_i' \beta_j$. While P_{ij} does not vary within markets, $\tilde{P}_{ij}(\beta_j)$ will if a component of X_i does. In order to make use of this variation, that component of X_i needs to be independent of V_i, which can be incorporated by modifying the instrumental variable assumptions in Section 3.4.1.

[46] Berry and Haile (2010) show how such information can be used to improve on the nonparametric point identification arguments in Berry and Haile (2014).

The second assumption is that the unobservables that vary across markets can be made comparable to prices. In (2), these unobservables are called ξ_{ijm}. In our notation, we can incorporate these by replacing (1) with

$$Y_i = \arg\max_{j \in \mathcal{J}} V_{ij} + \xi_j(M_i) - P_{ij}, \tag{38}$$

where ξ_j is an unknown function of the consumer's market. For each fixed ξ, this model is like (1) but with valuations given by $\tilde{V}_{ij}(\xi) \equiv V_{ij} + \xi_j(M_i)$. After incorporating unobserved market effects in this way, one may be willing to assume that V_{ij} is independent of $P_i = \pi(M_i)$, as is common in implementations of (2). This can be incorporated by modifying the instrumental variables assumptions in Section 3.4.1.

Implementing either (37) or (38) requires looping over possible parameter values β or ξ. However, for each candidate β and ξ, one can characterize and compute the identified set exactly as before. This suggests that such a procedure will still be sharp. Developing a feasible computational strategy appears more challenging, but not impossible. Since neither (37) or (38) are needed for our application, we leave fuller investigations of these extensions to future work.

D Construction of the Minimal Relevant Partition

We first observe that any price (premium) vector $p \in \mathbb{R}^J$ divides $\mathbb{R}^J$ into the sets $\{\mathcal{V}_j(p)\}_{j=0}^J$, as shown in Figures 1a and 1b. Intuitively, we view such a division as a partition, although formally this is not correct, since these sets can overlap on the hyperplanes like $v_j - p_j = v_k - p_k$ where ties occurs. These regions of overlap have Lebesgue measure zero in $\mathbb{R}^J$, so this caveat is unimportant given our focus on continuously distributed valuations. To avoid confusion, we refer to a collection of sets that would be a partition if not for regions of Lebesgue measure zero as an almost sure (a.s.) partition.

Definition 2. *Let $\{\mathcal{A}_t\}_{t=1}^T$ be a collection of Lebesgue measurable subsets of $\mathbb{R}^J$. Then $\{\mathcal{A}_t\}_{t=1}^T$ is an almost sure (a.s.) partition of $\mathbb{R}^J$ if*

a) $\bigcup_{t=1}^T \mathcal{A}_t = \mathbb{R}^J$; and

b) $\lambda(\mathcal{A}_t \cap \mathcal{A}_{t'}) = 0$ for any $t \neq t'$, where λ denotes Lebesgue measure on $\mathbb{R}^J$.

Next, we enumerate the price vectors in $\mathcal{P}$ as $\mathcal{P} = \{p_1, \ldots, p_L\}$ for some integer L. Let $\mathcal{Y} \equiv \mathcal{J}^L$ denote the collection of all L–tuples from the set of choices $\mathcal{J} \equiv \{0, 1, \ldots, J\}$. Then, since $\{\mathcal{V}_j(p_l)\}_{j=0}^J$ is an a.s. partition of $\mathbb{R}^J$ for every p_l, it follows that

$$\left\{\widetilde{\mathcal{V}}_y : y \in \mathcal{Y}\right\} \quad \text{where} \quad \widetilde{\mathcal{V}}_y \equiv \bigcap_{l=1}^{L} \mathcal{V}_{y_l}(p_l) \tag{39}$$

also constitutes an a.s. partition of $\mathbb{R}^J$.[47] Intuitively, each vector $y \equiv (y_1, \ldots, y_L)$ is a profile of L choices under the price vectors $(p_1, \ldots, p_L)$ that comprise $\mathcal{P}$. Each set $\widetilde{\mathcal{V}}_y$ in the a.s. partition (39) corresponds to the subset of valuations in $\mathbb{R}^J$ for which a consumer would make choices y when faced with prices $\mathcal{P}$.

The collection $\mathbb{V} \equiv \{\widetilde{\mathcal{V}}_y : y \in \mathcal{Y}\}$ is the MRP, since it satisfies Definition 1 by construction. To see this, note that if $v, v' \in \widetilde{\mathcal{V}}_y$ for some y, then by (39), $v, v' \in \mathcal{V}_{y_l}(p_l)$ for all $l = 1, \ldots, L$, at least up to collections of v, v' that have Lebesgue measure zero. Recalling (9), this implies (using the notation of Definition 1) that $Y(v, p) = Y(v', p)$ for all $p \in \mathcal{P}$. Conversely, if $Y(v, p) = Y(v', p)$ for all $p \in \mathcal{P}$, then taking

$$y \equiv (Y(v, p_1), \ldots, Y(v, p_L)) = (Y(v', p_1), \ldots, Y(v', p_L)), \tag{40}$$

yields an L–tuple $y \in \mathcal{Y}$ such that $v, v' \in \mathcal{V}_{y_l}(p_l)$, again barring ambiguities that occur with Lebesgue measure zero.

From a practical perspective, this is an inadequate representation of the MRP, because if choices are determined by the quasilinear model (1), then many of the sets $\widetilde{\mathcal{V}}_y$ must have Lebesgue measure zero. This makes indexing the partition by $y \in \mathcal{Y}$ excessive; for computation we would prefer an indexing scheme that only includes sets that are not already known to have measure zero. For this purpose, we use an algorithm that starts with the set of prices $\mathcal{P}$ and returns the collection of choice sequences $\overline{\mathcal{Y}}$ that are not required to have Lebesgue measure zero under (1). We use this set $\overline{\mathcal{Y}}$ in our computational implementations. Note that since $\widetilde{\mathcal{V}}_y$ has Lebesgue measure zero for any $y \in \mathcal{Y} \setminus \overline{\mathcal{Y}}$, the collection $\mathbb{V} \equiv \{\widetilde{\mathcal{V}}_y : y \in \overline{\mathcal{Y}}\}$ still constitutes an a.s. partition of $\mathbb{R}^J$ and still satisfies the key property (17) of the MRP in Definition 1.

The algorithm works as follows.[48] We begin by partitioning $\mathcal{P}$ into T sets (or blocks) of prices $\{\mathcal{P}_t\}_{t=1}^T$ that each contain (give or take) ψ prices. For each t, we then construct the set of all choice sequences $\overline{\mathcal{Y}}_t \subseteq \mathcal{J}^{|\mathcal{P}_t|}$ that are compatible with the

[47] Note that these sets are Lebesgue measurable, since $\mathcal{V}_j(p)$ is a finite intersection of half-spaces and $\widetilde{\mathcal{V}}_y$ is a finite intersection of sets like $\mathcal{V}_j(p)$.

[48] We expect that this algorithm leaves room for significant computational improvements, but we leave more sophisticated developments for future work. In practice, we also use some additional heuristics based on sorting the price vectors. These have useful but second-order speed improvements that are specific to our application, so for brevity we do not describe them here.

quasilinear choice model in the sense that $y^t \in \overline{\mathcal{Y}}_t$ if and only if the set

$$\left\{v \in \mathbb{R}^J : v_{y^t_l} - p_{y^t_l} \geq v_j - p_j \text{ for all } j \in \mathcal{J} \text{ and } p \in \mathcal{P}_t\right\} \tag{41}$$

is empty. In practice, we do this by sequentially checking the feasibility of a linear program with (41) as the constraint set. The sense in which we do this sequentially is that instead of checking (41) for all $y^t \in \mathcal{J}^{|\mathcal{P}_t|}$—which could be a large set even for moderate ψ—we first check whether it is nonempty when the constraint is imposed for only 2 prices in $\mathcal{P}_t$, then 3 prices, etc. Finding that (41) is empty when restricting attention to one of these shorter choice sequences implies that it must also be infeasible for all other sequences that share the short component. This observation helps speed up the algorithm substantially.

One we have found $\overline{\mathcal{Y}}_t$ for all t, we combine blocks of prices into pairs, then repeat the process with these larger, paired blocks. For example, if we let $\mathcal{P}_{12} \equiv \mathcal{P}_1 \cup \mathcal{P}_2$—i.e. we pair the first two blocks of prices—then we know that the set of $y^{12} \in \mathcal{J}^{|\mathcal{P}_1|+|\mathcal{P}_2|}$ that satisfy (41) must be a subset of $\{(y_1, y_2) : y_1 \in \overline{\mathcal{Y}}_1, y_2 \in \overline{\mathcal{Y}}_2\}$. We sequentially check the non-emptyness of (41) for all y^{12} in this set, eventually obtaining a set $\overline{\mathcal{Y}}_{12}$. Once we have done this for all pairs of price blocks, we then combine pairs of pairs of blocks (e.g. $\mathcal{P}_{12} \cup \mathcal{P}_{34}$) and repeat the process. Continuing in this way, we eventually end up with the original set of price vectors, $\mathcal{P}$, as well as the set of all surviving choice sequences, $\overline{\mathcal{Y}} \subseteq \mathcal{Y}$.

The key input to this algorithm is the number of prices in the initial price blocks, which we have denoted by ψ. The optimal value of ψ should be something larger than 2, but smaller than L. With small ψ, the sequential checking of (41) yields less payoff, since each detection of infeasibility eliminates fewer partial choice sequences. On the other hand, large ψ makes the strategy of combining pairs of smaller blocks of prices into larger blocks less fruitful. For our application, we use $\psi = 8$–10, which seems to be fairly efficient, although it is likely specific to our setting.

E Proofs for Propositions 1 and 2

E.1 Proposition 1

If $t \in \Theta^\star$, then by definition there exists an $f \in \mathcal{F}^\star$ such that $\theta(f) = t$. Let ϕ_f be defined as in (20), which we reproduce here for convenience:

$$\phi_f(\mathcal{V}|m, x) \equiv \int_{\mathcal{V}} f(v|m, x)\, dv. \tag{20}$$

Note that $\phi_f \in \Phi$, because the MRP $\mathbb{V}$ is (almost surely) a partition of $\mathbb{R}^J$, and f is a conditional probability density function on $\mathbb{R}^J$. Due to the assumed properties of $\overline{\theta}$, we also know that $\overline{\theta}(\phi_f) = \theta(f) = t$. To see that ϕ_f satisfies (25), observe that

$$\sum_{\mathcal{V} \in \mathbb{V}_j(\pi(m,x))} \phi_f(\mathcal{V}|m,x) \equiv \sum_{\mathcal{V} \in \mathbb{V}_j(\pi(m,x))} \int_{\mathcal{V}} f(v|m,x)\,dv = s_f(j|m,x) = s(j|m,x),$$

where the first equality follows by definition (20), the second follows from (18), and the third follows from the definition of $\mathcal{F}^\star$. Similarly, ϕ_f satisfies (26) because

$$\begin{aligned}
&\mathbb{E}[\phi_f(\mathcal{V}|M_i,X_i)|W_i = w, Z_i = z] \\
&\quad = \mathbb{E}\left[\int_{\mathcal{V}} f(v|M_i,X_i)\,dv \middle| W_i = w, Z_i = z\right] \\
&\quad = \int_{\mathcal{V}} \mathbb{E}\left[f(v|M_i,X_i)|W_i = w, Z_i = z\right] dv \\
&\quad = \int_{\mathcal{V}} \mathbb{E}\left[f(v|M_i,X_i)|W_i = w, Z_i = z'\right] dv = \mathbb{E}[\phi_f(\mathcal{V}|M_i,X_i)|W_i = w, Z_i = z']
\end{aligned}$$

where the second equality follows by Tonelli's Theorem (e.g. Shorack, 2000, pg. 82), the third uses (6), which holds (by assumption) for all $f \in \mathcal{F}^\star$, and the final equality reverses the steps of the first two equalities. That ϕ_f also satisfies (27) follows using a similar argument and the hypothesis that $f \in \mathcal{F}^\star$ satisfies (7), i.e.

$$\begin{aligned}
\sum_{\mathcal{V} \in \overline{\mathbb{V}}(w)} (\phi_f)_{\mathbb{V}|WZ}(\mathcal{V}|w,z) &= \sum_{\mathcal{V} \in \overline{\mathbb{V}}(w)} \int_{\mathcal{V}} \mathbb{E}\left[f(v|M_i,X_i)|W_i = w, Z_i = z\right] dv \\
&= \int_{\cup\{\mathcal{V}:\mathcal{V} \in \overline{\mathbb{V}}(w)\}} f_{V|WZ}(v|w,z)\,dv \\
&\geq \int_{\overline{\mathcal{V}}(w)} f_{V|WZ}(v|w,z)\,dv = 1. \qquad (42)
\end{aligned}$$

The inequality in (42) follows because the definition of $\overline{\mathbb{V}}(w)$, together with the fact that $\mathbb{V}$ is an a.s. partition of $\mathbb{R}^J$, implies that $\overline{\mathcal{V}}(w)$ is contained in the union of sets in $\overline{\mathbb{V}}(w)$. This inequality implies that ϕ_f satisfies (27), because

$$\begin{aligned}
\sum_{\mathcal{V} \in \overline{\mathbb{V}}(w)} (\phi_f)_{\mathbb{V}|WZ}(\mathcal{V}|w,z) &\leq \sum_{\mathcal{V} \in \mathbb{V}} (\phi_f)_{\mathbb{V}|WZ}(\mathcal{V}|w,z) \\
&= \mathbb{E}\left[\sum_{\mathcal{V} \in \mathbb{V}} \phi_f(\mathcal{V}|M_i,X_i) \middle| W_i = w, Z_i = z\right] = 1,
\end{aligned}$$

as a result of ϕ_f being an element of Φ. We have now established that if $t \in \Theta^\star$, then there exists a $\phi \in \Phi$ satisfying (25)–(27) for which $\overline{\theta}(\phi) = t$.

Conversely, suppose that such a $\phi \in \Phi$ exists for some t. Recall that W_i was assumed to be a subvector (or more generally, a function) of (M_i, X_i), and denote this function by ω, so that $W_i = \omega(M_i, X_i)$. Then define

$$f_\phi(v|m,x) \equiv \sum_{\mathcal{V} \in \overline{\mathbb{V}}(\omega(m,x))} \frac{\mathbb{1}\left[v \in \mathcal{V} \cap \overline{\mathcal{V}}(\omega(m,x))\right]}{\lambda\left(\mathcal{V} \cap \overline{\mathcal{V}}(\omega(m,x))\right)} \phi(\mathcal{V}|m,x),$$

noting that the summands are well-defined by the definition of $\overline{\mathbb{V}}(w)$. We will show that $t \in \Theta^\star$ by establishing that $f_\phi \in \mathcal{F}^\star$ and $\theta(f_\phi) = t$.

First observe that for any $\mathcal{V} \in \mathbb{V}$,

$$\begin{aligned}\int_{\mathcal{V}} f_\phi(v|m,x)\, dv &\equiv \sum_{\mathcal{V}' \in \overline{\mathbb{V}}(\omega(m,x))} \int_{\mathcal{V}} \frac{\mathbb{1}\left[v \in \mathcal{V}' \cap \overline{\mathcal{V}}(\omega(m,x))\right]}{\lambda\left(\mathcal{V}' \cap \overline{\mathcal{V}}(\omega(m,x))\right)} \phi(\mathcal{V}'|m,x)\, dv \\ &= \mathbb{1}[\mathcal{V} \in \overline{\mathbb{V}}(\omega(m,x))]\phi(\mathcal{V}|m,x), \end{aligned} \tag{43}$$

since the sets in $\mathbb{V}$ and hence $\overline{\mathbb{V}}(\omega(m,x))$ are disjoint (almost surely). Using (43), we have that

$$\int_{\mathbb{R}^J} f_\phi(v|m,x)\, dv = \sum_{\mathcal{V} \in \mathbb{V}} \int_{\mathcal{V}} f_\phi(v|m,x)\, dv = \sum_{\mathcal{V} \in \overline{\mathbb{V}}(\omega(m,x))} \phi(\mathcal{V}|m,x) = 1, \tag{44}$$

where the first equality uses the fact that $\mathbb{V}$ is a partition. The final equality is implied by the hypothesis that ϕ satisfies (27), since

$$1 = \sum_{\mathcal{V} \in \overline{\mathbb{V}}(w)} \phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z) = \mathbb{E}\left[\sum_{\mathcal{V} \in \overline{\mathbb{V}}(\omega(M_i,X_i))} \phi(\mathcal{V}|M_i,X_i) \Big| W_i = w, Z_i = z \right],$$

and every $\phi \in \Phi$ satisfies

$$\sum_{\mathcal{V} \in \overline{\mathbb{V}}(\omega(m,x))} \phi(\mathcal{V}|m,x) \leq \sum_{\mathcal{V} \in \mathbb{V}} \phi(\mathcal{V}|m,x) = 1.$$

Thus, from (44), and since f_ϕ inherits non-negativity from $\phi \in \Phi$, we conclude that f_ϕ is a conditional density, i.e. $f_\phi \in \mathcal{F}$.

To see that f_ϕ satisfies (6), notice that

$$
\begin{aligned}
(f_\phi)_{V|WZ}(v|w,z) &\equiv \mathbb{E}\left[f_\phi(v|M_i,X_i)|W_i=w,Z_i=z\right] \\
&\equiv \mathbb{E}\left[\sum_{\mathcal{V}\in\overline{\mathbb{V}}(w)} \frac{\mathbb{1}\left[v\in\mathcal{V}\cap\overline{\mathcal{V}}(w)\right]}{\lambda\left(\mathcal{V}\cap\overline{\mathcal{V}}(w)\right)}\phi(\mathcal{V}|M_i,X_i)\Big|W_i=w,Z_i=z\right] \\
&= \sum_{\mathcal{V}\in\overline{\mathbb{V}}(w)} \frac{\mathbb{1}\left[v\in\mathcal{V}\cap\overline{\mathcal{V}}(w)\right]}{\lambda\left(\mathcal{V}\cap\overline{\mathcal{V}}(w)\right)}\phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z) \\
&= \sum_{\mathcal{V}\in\overline{\mathbb{V}}(w)} \frac{\mathbb{1}\left[v\in\mathcal{V}\cap\overline{\mathcal{V}}(w)\right]}{\lambda\left(\mathcal{V}\cap\overline{\mathcal{V}}(w)\right)}\phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z') = (f_\phi)_{V|WZ}(v|w,z'),
\end{aligned}
$$

where the fourth equality uses (26), and the final equality reverses the steps of the first four. The satisfaction of the verticality condition, (7), follows in a similar way from (27) and Tonelli's Theorem, since

$$
\begin{aligned}
\int_{\overline{\mathcal{V}}(w)} (f_\phi)_{V|WZ}(v|w,z)\,dv &\equiv \int_{\overline{\mathcal{V}}(w)} \mathbb{E}\left[f_\phi(v|M_i,X_i)|W_i=w,Z_i=z\right]dv \\
&= \mathbb{E}\left[\sum_{\mathcal{V}\in\overline{\mathbb{V}}(w)}\phi(\mathcal{V}|M_i,X_i)\Big|W_i=w,Z_i=z\right] \\
&= \sum_{\mathcal{V}\in\overline{\mathbb{V}}(w)}\phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z) = 1.
\end{aligned}
$$

That f_ϕ satisfies the observational equivalence condition (11) follows from (18), (25), and (43), i.e.

$$
\begin{aligned}
s_{f_\phi}(j|m,x) &\equiv \sum_{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))}\int_{\mathcal{V}} f_\phi(v|m,x)\,dv \\
&= \sum_{\substack{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))\\ \cap\overline{\mathbb{V}}(\omega(m,x))}}\phi(\mathcal{V}|m,x) \\
&= \sum_{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))}\phi(\mathcal{V}|m,x) - \sum_{\substack{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))\\ \cap\overline{\mathcal{V}}(\omega(m,x))^c}}\phi(\mathcal{V}|m,x) = s(j|m,x),
\end{aligned}
$$

for all $j\in\mathcal{J}$ and $(p,x)\in\text{supp}(P_i,X_i)$. The last equality here follows using (44) because

$$0 \leq \sum_{\substack{\mathcal{V} \in \mathbb{V}_j(\pi(m,x)) \\ \cap \overline{\mathbb{V}}(\omega(m,x))^c}} \phi(\mathcal{V}|m,x) \leq \sum_{\mathcal{V} \in \overline{\mathbb{V}}(\omega(m,x))^c} \phi(\mathcal{V}|m,x) = 1 - \sum_{\mathcal{V} \in \overline{\mathbb{V}}(\omega(m,x))} \phi(\mathcal{V}|m,x) = 0.$$

Finally, note that in the notation of (20), (43) says

$$\phi_{f_\phi}(\mathcal{V}|m,x) = \mathbb{1}[\mathcal{V} \in \overline{\mathbb{V}}(\omega(m,x))]\phi(\mathcal{V}|m,x).$$

This equality implies that $\phi_{f_\phi}(\mathcal{V}|m,x) = \phi(\mathcal{V}|m,x)$ for all $\mathcal{V}$, since for $\mathcal{V} \notin \overline{\mathbb{V}}(\omega(m,x))$ we must have $\phi(\mathcal{V}|m,x) = 0$, as implied by (44). Thus, $\theta(f_\phi) = \overline{\theta}(\phi_{f_\phi}) = \overline{\theta}(\phi) = t$, and therefore $t \in \Theta^\star$. *Q.E.D.*

E.2 Proof of Proposition 2

Observe that Φ is a compact and connected subset of $\mathbb{R}^{d_\phi}$. Since (25)–(27) are linear equalities, the subset of Φ that satisfies them is also compact and connected. Thus, if $\overline{\theta}$ is continuous on this subset and $d_\theta = 1$, it follows that its image over it—which Proposition 1 established to be $\Theta^\star$— is compact and connected as well. If $d_\theta = 1$, then $\Theta^\star$ is a compact interval, so by definition its endpoints must be given by $t_\star$ and $t^\star$. *Q.E.D.*

F Implementing Bounds on Consumer Surplus

In this section, we show that setting the target parameter to be the change in consumer surplus (as defined in (4)) results in a reduced target parameter function ($\overline{\theta}$) that is linear in ϕ. For shorthand, we denote average consumer surplus at price $p^\star$, conditional on $(M_i, X_i) = (m, x)$ as

$$\text{CS}_{p^\star}(f|m,x) \equiv \int \left\{ \max_{j \in \mathcal{J}} v_j - p^\star_j \right\} f(v|m,x)\, dv.$$

Suppose that $\mathbb{V}$ is a minimal relevant partition constructed from a set of premiums $\mathcal{P}$ that contains two premiums of interest, p and $p^\star$. Then

$$\text{CS}_{p^\star}(f|m,x) = \sum_{\mathcal{V} \in \mathbb{V}} \int_{\mathcal{V}} \left\{ \max_{j \in \mathcal{J}} v_j - p^\star_j \right\} f(v|m,x)\, dv, \tag{45}$$

since the MRP is an (almost sure) partition of $\mathbb{R}^J$. By definition of the MRP, the optimal choice of plan is constant as a function of v within any MRP set $\mathcal{V}$. That is, using the notation in Definition 1, $\arg\max_{j\in\mathcal{J}} v_j - p_j \equiv Y(v,p) = Y(v',p) \equiv Y(\mathcal{V},p)$ for all $v, v' \in \mathcal{V}$ and any $p \in \mathcal{P}$. Consequently, we have from (45) that

$$\mathrm{CS}_{p^\star}(f|m,x) = \sum_{\mathcal{V}\in\mathbb{V}} \int_{\mathcal{V}} v_{Y(\mathcal{V},p^\star)} f(v|m,x)\, dv - p^\star_{Y(\mathcal{V},p^\star)}$$

Replacing $p^\star$ by p, it follows that the change in consumer surplus resulting from a shift in prices from $p \to p^\star$ can be written as

$$\begin{aligned}\Delta\mathrm{CS}_{p\to p^\star}(f|m,x) &\equiv \mathrm{CS}_{p^\star}(f|m,x) - \mathrm{CS}_p(f|m,x)\\ &= \sum_{\mathcal{V}\in\mathbb{V}} \int_{\mathcal{V}} \left(v_{Y(\mathcal{V},p^\star)} - v_{Y(\mathcal{V},p)}\right) f(v|m,x)\, dv + p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)}.\end{aligned}$$

Now define the smallest and largest possible change in valuations within any partition set $\mathcal{V}$ as

$$\begin{aligned} &\underline{v}_{p\to p^\star}(\mathcal{V}) \equiv \min_{v\in\mathcal{V}} v_{Y(\mathcal{V},p^\star)} - v_{Y(\mathcal{V},p)},\\ \text{and}\quad &\overline{v}_{p\to p^\star}(\mathcal{V}) \equiv \max_{v\in\mathcal{V}} v_{Y(\mathcal{V},p^\star)} - v_{Y(\mathcal{V},p)}.\end{aligned}$$

Since we do not restrict the distribution of valuations within each MRP set, the sharp lower bound on a change in consumer surplus is attained when this distribution concentrates all of its mass on $\underline{v}_{p\to p^\star}(\mathcal{V})$ in every $\mathcal{V} \in \mathbb{V}$. That is,

$$\Delta\mathrm{CS}_{p\to p^\star}(f|m,x) \geq \sum_{\mathcal{V}\in\mathbb{V}} \underline{v}_{p\to p^\star}(\mathcal{V})\phi_f(\mathcal{V}|m,x) + p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)} \equiv \underline{\Delta\mathrm{CS}}_{p\to p^\star}(f|m,x).$$

Similarly, the sharp upper bound for any f is given by

$$\Delta\mathrm{CS}_{p\to p^\star}(f|m,x) \leq \sum_{\mathcal{V}\in\mathbb{V}} \overline{v}_{p\to p^\star}(\mathcal{V})\phi_f(\mathcal{V}|m,x) + p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)} \equiv \overline{\Delta\mathrm{CS}}_{p\to p^\star}(f|m,x).$$

Therefore, a sharp upper bound on the change in consumer surplus can be found by taking $\theta(f) \equiv \underline{\Delta\mathrm{CS}}_{p\to p^\star}(f|m,x)$, setting

$$\overline{\theta}(\phi) \equiv \sum_{\mathcal{V}\in\mathbb{V}} \underline{v}_{p\to p^\star}(\mathcal{V})\phi(\mathcal{V}|m,x) + p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)}.$$

and applying Propositions 1 or 2. The key requirement that $\theta(f) = \overline{\theta}(\phi_f)$ can be seen to be satisfied here by examining the expression for $\underline{\Delta\mathrm{CS}}_{p\to p^\star}(f|m,x)$ above. The sharp upper bound is found analogously.

G Estimation

Our analysis in Section 3 concerns the identification problem under which the joint distribution of (Y_i, M_i, X_i) is treated as known. In practice, features of this distribution, such as the choice shares $s(j|m,x)$, need to be estimated from a finite data set, so we want to model them as potentially contaminated with statistical error. In this section, we show how to modify Proposition 2 to account for such error in our primary case of interest with $\overline{\theta}$ linear. A formal justification for this procedure is developed in Mogstad et al. (2018).

The estimator proceeds in two steps. First, we minimize the discrepancy in the observational equivalence conditions (25) by solving

$$\widehat{Q}^\star \equiv \min_{\phi \in \Phi} \widehat{Q}(\phi) \qquad \text{subject to (26)–(27),}$$

$$\text{where} \quad \widehat{Q}(\phi) \equiv \sum_{j,m,x} \widehat{\mathbb{P}}[M_i = m, X_i = x] \left| \widehat{s}(j|m,x) - \sum_{\mathcal{V} \in \mathbb{V}_j(\pi(m,x))} \phi(\mathcal{V}|m,x) \right|, \tag{46}$$

with $\widehat{s}(j|m,x)$ the estimated share of choice j, conditional on $(M_i, X_i) = (m,x)$, and $\widehat{\mathbb{P}}[M_i = m, X_i = x]$ an estimate of the density of (M_i, X_i). The use of absolute deviations in the definition of $\widehat{Q}$ means that (46) can be reformulated as a linear program by replacing terms in absolute values by the sum of their positive and negative parts.[49] We weight these absolute deviations by the estimated density of (M_i, X_i) so that regions of smaller density do not have an outsized impact on the estimated bounds.

In the second step, we collect values of $\overline{\theta}(\phi)$ among ϕ that come close to minimizing (46). That is, we construct the set:

$$\widehat{\Theta}^\star \equiv \left\{ \overline{\theta}(\phi) : \phi \in \Phi, \text{ and } \widehat{Q}(\phi) \leq \widehat{Q}^\star + \eta_n, \text{ and } \phi \text{ satisfies (26)–(27)} \right\} \tag{47}$$

The qualifier "close" here reflects the tuning parameter η_n, which must converge to zero at an appropriate rate with the sample size, n. The purpose of this tuning parameter is to smooth out possible discontinuities caused by set convergence. In our empirical estimates, we set $\eta_n = .1$, and found very little sensitivity to values of η_n that were bigger or smaller by an order of magnitude. However, there are currently no theoretical results to guide the choice of this parameter.

In our main case of interest when $\overline{\theta}$ is linear and scalar-valued, we estimate $\widehat{\Theta}^\star$ by solving two linear programs that replace (25) with the condition in (47). That is, we

49 This is a common reformulation argument, see e.g. Bertsimas and Tsitsiklis (1997, pp. 19–20).

solve

$$\widehat{t}_\star \equiv \min_{\phi\in\Phi} \overline{\theta}(\phi) \quad \text{s.t} \quad \widehat{Q}(\phi) \leq \widehat{Q}^\star + \eta_n, \tag{48}$$

and an analogous maximization problem defining $\widehat{t}^\star$. The set estimator for $\Theta^\star$ is then $\widehat{\Theta}^\star \equiv [\widehat{t}_\star, \widehat{t}^\star]$. For this case, Mogstad et al. (2018) show that $\widehat{t}_\star$ and $\widehat{t}^\star$ are consistent for $t_\star$ and $t^\star$ under weak conditions on $\widehat{s}$. When $\overline{\theta}$ is linear, (48) can be reformulated as a linear program, again by appropriately rephrasing the absolute value terms in terms of their positive and negative parts. In this case, the overall procedure of the estimator is to solve three linear programs: One for (46), one for (48), and one for the analogous maximization problem.

H Estimation of Potential Buyers

In this section, we describe how we use the American Community Survey (ACS) to estimate the number of potential buyers in each market × age × income bin, or each value of $(M_i, X_i) = (m, x)$. As is often the case in empirical demand analysis, our administrative data only contains observations of individuals who buy health insurance in Covered California, but not those who were eligible yet chose the outside option. That is, we do not have data on the quantity who chose choice 0.[50] Instead, we will construct conditional choice probability (market shares) by estimating the number of potential buyers, constructing shares of the inside choices ($j \geq 1$) by dividing quantity by potential buyers, and then taking the difference between the sum of the inside shares and 1 to be the share of the outside choice.

The key step here is estimating the number of potential buyers (market size), $\mu(m, x)$, for each $(M_i, X_i) = (m, x)$. We do this using the California 2013 3-year subsample of the American Community Survey (ACS) public use file, downloaded from IPUMS (Ruggles et al., 2015).[51] We define an individual as a potential buyer, denoted by the indicator $I_i = 1$, if they report being either uninsured or privately insured. Individuals with $I_i = 0$ include those who are covered by employer-sponsored plans, Medi-Cal (Medicaid), Medicare, or other types of public insurance. Then our estimator of $\mu(m, x)$ is

$$\widehat{\mu}(m, x) = \sum_{i=1}^{N} \text{weight}_i I_i \mathbb{1}[M_i = m, X_i = x], \tag{49}$$

[50] For further discussion, see for example (Berry, 1994, pg. 247).

[51] The 3 year sample includes information from 2011 to 2013. We use the entire 3 year sample to increase our sample size.

where weight_i are the individual sampling weights provided in the ACS, and N is the total sample size. This sample reflects the selection rules discussed in Section 4.1. To impose the restriction to households with 1 or 2 adults, we combine age with the IPUMS definition of a health insurance unit (HIU), and keep only individiuals in HIUs of size 1 or 2.

An adjustment to this procedure is needed to account for the fact that the PUMA (public use micro area) geographic identifier in the ACS can be split across multiple counties, and so in some cases also multiple ACA rating regions. For a PUMA that is split in such a way, we allocate HIUs to each rating region it overlaps using the population of the zipcodes in the PUMA as weights. This is the same adjustment factor used in the PUMA to county crosswalk.[52] Since the definition of a PUMA changed after 2011, we also use this adjustment scheme to convert the 2011 PUMA definitions to 2012–2013 definitions.

A final adjustment is needed for situations in which our estimate $\widehat{\mu}(m, x)$ is smaller than the number of enrollees in the Covered California administrative data. Some fix for such a case is needed in order to keep all shares bounded between 0 and 1. The fix we use is to replace $\widehat{\mu}(m, x)$ by the total enrollment observed in the administrative data, so that the estimated share of the outside option is 0. In practice, we find that this only happens for smaller x bins in sparsely-populated rating regions, and we expect the cause is statistical error in $\widehat{\mu}(m, x)$. While our solution is not ideal, it seems to be the best that we can do given the available data. Since our estimates are weighted by bin size, the adjustment we use turns out to affect our results little when compared to some other (also ad hoc) adjustments we have tried.

[52] For example, suppose that an HIU is in a PUMA that spans counties A and B, and that this HIU has a total sampling weight of 10, so that it represents 10 observationally equivalent households. If the adjustment factor is 0.3 in county A and 0.7 in county B, we assume there are 3 identical HIUs in county A and 7 in county B.

References

Abraham, J., C. Drake, D. W. Sacks, and K. Simon (2017): "Demand for health insurance marketplace plans was highly elastic in 2014–2015," *Economics Letters*, 159, 69–73. 3

Allen, R. and J. Rehbeck (2017): "Identification with Additively Separable Heterogeneity," *Working paper*. 38, 39

Armstrong, T. B. (2016): "Large Market Asymptotics for Differentiated Product Demand Estimators With Economic Models of Supply," *Econometrica*, 84, 1961–1980. 7

Balke, A. and J. Pearl (1994): "Counterfactual Probabilities: Computational Methods, Bounds, and Applications," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI-94)*, ed. by R. Lopez de Mantras and D. Poole, 46–54. 39

——— (1997): "Bounds on Treatment Effects From Studies With Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176. 39

Berry, S. and P. Haile (2015): "Identification in Differentiated Products Markets," *Annu. Rev. Econ.*, –. 7

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890. 1, 7, 8, 39, 42

——— (2004a): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112, 68–105. 7, 39

Berry, S., O. B. Linton, and A. Pakes (2004b): "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *Review of Economic Studies*, 71, 613–654. 7

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25, 242–262. 1, 7, 39, 52

Berry, S. T. and P. A. Haile (2010): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," *Cowles Foundation Discussion Paper 1718*. 2, 7, 42

——— (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797. 2, 39, 42

Bertsimas, D. and J. N. Tsitsiklis (1997): *Introduction to linear optimization*, vol. 6, Athena Scientific Belmont, MA. 51

Briesch, R. A., P. K. Chintagunta, and R. L. Matzkin (2002): "Semiparametric Estimation of Brand Choice Behavior," *Journal of the American Statistical Association*, 97, 973–982. 38

——— (2010): "Nonparametric Discrete Choice Models With Unobserved Heterogeneity," *Journal of Business & Economic Statistics*, 28, 291–307. 38

Bundorf, M. K., J. Levin, and N. Mahoney (2012): "Pricing and Welfare in Health Plan Choice," *American Economic Review*, 102, 3214–3248. 6

Canay, I. A., A. Santos, and A. M. Shaikh (2013): "On the Testability of Identification in Some Nonparametric Models With Endogeneity," *Econometrica*, 81, 2535–2559. 39

Chan, D. and J. Gruber (2010): "How Sensitive are Low Income Families to Health Plan Prices?" *American Economic Review*, 100, 292–296. 1

Chiappori, P.-A. and I. Komunjer (2009): "On the Nonparametric Identification of Multiple Choice Models," *Working paper*. 2, 39

Chiong, K. X., Y.-W. Hsieh, and M. Shum (2017): "Counterfactual Estimation in Semiparametric Multinomial Choice Models," *SSRN Electronic Journal*. 38, 39

Compiani, G. (2018): "Nonparametric Demand Estimation in Differentiated Products Markets," *Working paper*. 3, 39

Conley, T. G., C. B. Hansen, and P. E. Rossi (2010): "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260–272. 33

Demuynck, T. (2015): "Bounding average treatment effects: A linear programming approach," *Economics Letters*, 137, 75–77. 40

Einav, L., A. Finkelstein, and M. R. Cullen (2010a): "Estimating Welfare in Insurance Markets Using Variation in Prices," *Quarterly Journal of Economics*, 125, 877–921. 6

Einav, L., A. Finkelstein, and J. Levin (2010b): "Beyond Testing: Empirical Models of Insurance Markets," *Annual Review of Economics*, 2, 311–336. 6

Ericson, K. M. and A. Starc (2015): "Pricing Regulation and Imperfect Competition on the Massachusetts Health Insurance Exchange," *The Review of Economics and Statistics, forthcoming*, 97, 667–682. 1, 33

Finkelstein, A., N. Hendren, and M. Shepard (2017): "Subsidizing Health Insurance for Low-Income Adults: Evidence from Massachusetts," Working Paper 23668, National Bureau of Economic Research. 3, 23, 33

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, K. Baicker, and O. H. S. Group (2012): "The Oregon health insurance experiment: evidence from the first year," *The Quarterly journal of economics*, 127, 1057–1106. 33

Fourer, R., D. M. Gay, and B. W. Kernighan (2002): *AMPL: A Modeling Language for Mathematical Programming*, Cengage Learning. 18

Fox, J. T. (2007): "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *The RAND Journal of Economics*, 38, 1002–1019. 38

Fox, J. T. and A. Gandhi (2016): "Nonparametric identification and estimation of random coefficients in multinomial choice models," *The RAND Journal of Economics*, 47, 118–139. 10, 39

Fox, J. T., K. i. Kim, S. P. Ryan, and P. Bajari (2011): "A simple estimator for the distribution of random coefficients," *Quantitative Economics*, 2, 381–418. 7

——— (2012): "The random coefficients logit model is identified," *Journal of Econometrics*, 166, 204–212. 7

Freyberger, J. and J. L. Horowitz (2015): "Identification and shape restrictions in non-parametric instrumental variables estimation," *Journal of Econometrics*, 189, 41–53. 39

Gandhi, A., Z. Lu, and X. Shi (2017): "Estimating Demand for Differentiated Products with Zeroes in Market Share Data," *Working paper*. 39

Garthwaite, C., T. Gross, and M. J. Notowidigdo (2018): "Hospitals as Insurers of Last Resort," *American Economic Journal: Applied Economics*, 10, 1–39. 33

Geruso, M. (2017): "Demand Heterogeneity in Insurance Markets: Implications for Equity and Efficiency," *Quantitative Economics*, 8, 929–975. 33

Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63, 891–951. 1, 8

Gurobi Optimization, I. (2015): "Gurobi Optimizer Reference Manual," . 18

Handel, B., I. Hendel, and M. D. Whinston (2015): "Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk," *Econometrica*, 83, 1261–1313. 6, 40

Handel, B. R. (2013): "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103, 2643–82. 6, 40

Hansen, L. P., J. Heaton, and E. G. J. Luttmer (1995): "Econometric Evaluation of Asset Pricing Models," *The Review of Financial Studies*, 8, 237–274. 39

Hausman, Leonard, and Zona (1994): "Competitive Analysis with Differenciated Products," *Annales d'Économie et de Statistique*, 159. 39

Hausman, J. A. (1996): *Valuation of New Goods under Perfect and Imperfect Competition*, University of Chicago Press, 207–248. 1

Hausman, J. A. and D. A. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403–426. 8

Heckman, J. J. and B. E. Honoré (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1149. 10

Heckman, J. J. and S. Navarro (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136, 341–396. 10

Ho, K. and A. Pakes (2014): "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, 104, 3841–84. 3, 38, 39

Honoré, B. E. and A. Lleras-Muney (2006): "Bounds in Competing Risks Models and the War on Cancer," *Econometrica*, 74, 1675–1698. 39

Honoré, B. E. and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629. 39

IBM (2010): *IBM ILOG AMPL Version 12.2*, International Business Machines Corporation. 18

Kamat, V. (2017): "Identification with Latent Choice Sets: The Case of the Head Start Impact Study," *Working paper.* 40

Kline, P. and M. Tartari (2016): "Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach," *American Economic Review*, 106, 972–1014. 40

Lafférs, L. (2013): "A note on bounding average treatment effects," *Economics Letters*, 120, 424–428. 39

Lewbel, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177. 10, 38, 39

Mahoney, N. (2015): "Bankruptcy as Implicit Health Insurance," *American Economic Review*, 105, 710–46. 33

Manski, C. F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3, 205–228. 38

——— (1985): "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27, 313–333. 10

——— (2007): "Partial Indentification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393–1410. 38, 39, 40

——— (2014): "Identification of income-leisure preferences and evaluation of income tax policy," *Quantitative Economics*, 5, 145–174. 38, 39

Manski, C. F. and J. V. Pepper (2017): "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions," *The Review of Economics and Statistics*. 33

Matzkin, R. L. (1991): "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica*, 59, 1315–1327. 38

——— (1993): "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 58, 137–168. 2, 38

McFadden, D. (1974): "Conditional logit analysis of qualitative choice behavior," in *Frontiers of Econometrics*, ed. by P. Zarembka, Academic Press: New York, 105–142. 1

McFadden, D. and K. Train (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447–470. 1, 8

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters," *Econometrica (forthcoming)*. 40, 51, 52

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342. 7

——— (2011): "Empirical Models of Consumer Behavior," *Annual Review of Economics*, 3, 51–75. 7

Nevo, A. and A. M. Rosen (2012): "Identification With Imperfect Instruments," *Review of Economics and Statistics*, 94, 659–671. 33, 39

Orsini, J. and P. Tebaldi (2017): "Regulated Age-based Pricing in Subsidized Health Insurance: Evidence from the Affordable Care Act," *Available at SSRN 2464725.* 4

Pakes, A. (2010): "Alternative Models for Moment Inequalities," *Econometrica*, 78, 1783–1822. 38, 39

Pakes, A. and J. Porter (2016): "Moment Inequalities for Multinomial Choice with Fixed Effects," Tech. rep. 38, 39

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006): "Moment Inequalities and Their Application," *Working paper.* 38, 39

——— (2015): "Moment Inequalities and Their Application," *Econometrica*, 83, 315–334. 38, 39

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015): "Integrated Public Use Microdata Series: Version 6.0 [dataset]," *Minneapolis: University of Minnesota.* 23, 52

Saltzman, E. (2017): "Demand for Health Insurance: Evidence from the California and Washington ACA Marketplaces," . 1

Shi, X., M. Shum, and W. Song (2016): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *SSRN Electronic Journal.* 38

Shorack, G. R. (2000): *Probability for Statisticians*, Springer-Verlag. 46

Tebaldi, P. (2017): "Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA," *Working paper.* 1, 3, 4, 23, 33

Thompson, T. S. (1989): "Identification of Semiparametric Discrete Choice Models," *Discussion paper 249, Center for Economics Research, Department of Economics, University of Minnesota.* 2, 10, 38

Torgovitsky, A. (2016): "Nonparametric Inference on State Dependence with Applications to Employment Dynamics," *SSRN Working Paper 2564305.* 40

——— (2018): "Partial Identification by Extending Subdistributions," *Quantitative Economics*, Forthcoming. 38, 40