Nonparametric Estimates of Demand in the California Health Insurance Exchange *

Pietro Tebaldi[†] Alexander Torgovitsky[‡] Hanbin Yang[§]

April 8, 2019

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 effects of changing premium subsidies on coverage choices, consumer surplus, and government spending. We find that a $10 decrease in monthly premium subsidies would cause between a 1.6% and 7.0% decline in the proportion of low-income adults with coverage. The reduction in total annual consumer surplus would be between $63 and $78 million, while the savings in yearly subsidy outlays would be between $238 and $604 million. Comparable logit models yield price sensitivity estimates towards the lower end of the bounds.

*We thank Nikhil Agarwal, Stéphane Bonhomme, Steve Berry, Øystein Daljord, Michael Dinerstein, Liran Einav, Phil Haile, Kate Ho, Ali Hortaçsu, Simon Lee, Sanjog Misra, Magne Mogstad, Ariel Pakes, Jack Porter, Bernard Salanié, and Thomas Wollman for helpful feedback. We thank conference participants at The Buenos Aires Meetings in Economics at Universidad Torcuato Di Tella, the 2018 Cowles Foundation Summer Conference on Structural Microeconomics, The Workshop on Structural Industrial Organization at Pontificia Universidad Javeriana, Bogota, The Eleventh Annual Federal Trade Commission Microeconomics Conference, the 2019 Winter Econometric Society Meetings, and the 2019 NBER Industrial Organization Winter Meetings, as well as seminar participants at Georgetown University, Columbia University, University of Chicago Booth (Marketing), University of California, Irvine, University of California, San Diego, University of California, Los Angeles, The Einaudi Institute for Economics and Finance, Northwestern University, and Università Ca' Foscari Venezia.

[†]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 individuals (Congressional Budget Office, 2017). 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 at counterfactual prices.

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 (2019) 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 the potential endogeneity of prices (e.g. Berry, 1994; Hausman, 1996; Berry et al., 1995). However, they are all fully parametric, with the logistic and normal distributions playing a central role in the parameterization. This raises the possibility that demand estimates from these models 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 its 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]

Nonparametric point identification arguments for discrete choice models are often

[1] While we develop the methodology with a focus on health insurance, it may also be useful for analyzing demand in other markets, as well as for discrete choice analysis more generally. However, an important difference with many discrete choice analyses is that in our context we observe more than one price per market. See Sections 2, 3.3, and Appendix C for more detail.

premised on the assumption of a large amount of exogenous variation in prices or other observable characteristics (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 are unlikely to be satisfied. This leads us to consider a partial identification framework (see Ho and Rosen, 2017, for a recent review).

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 can be characterized by considering only the way the distribution of valuations places mass on sets in the MRP. We then use this result to develop estimators of these bounds, which we implement using linear programming.

We apply the empirical methodology with administrative data to estimate demand counterfactuals for Covered California. We focus on the choice of metal tier for low-income individuals who are not covered under employer-sponsored insurance or public programs. Our main counterfactual of interest is how changes in premium subsidies 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 market.

Since the nonparametric model is partially identified, this strategy yields bounds rather than point estimates. However, the estimated bounds are quite informative. Using our preferred specification, we estimate that a \$10 decrease in monthly premium subsidies would cause between a 1.6% and 7.0% decline in the proportion of low-income adults with coverage. The average consumer surplus reduction would be between \$1.99 and \$2.45 per person, per month, or between \$63 and \$78 million annually when aggregated. Total annual savings on subsidy outlays would be between \$238 and \$604 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 et al., 2017; Finkelstein et al., 2019).[2]

We show that comparable estimates using parametric logit and probit models tend to yield price responses close to the lower bounds, and so may substantially understate price sensitivity. This possibility becomes more acute when considering larger price changes that involve more distant extrapolations. It also remains when considering richer 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 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.

In Appendix A, we provide a detailed review of the related methodological literature on semi- and nonparametric discrete choice models. Here, we briefly mention the two papers most closely related to ours. Chesher et al. (2013) use random set theory to derive moment inequalities in general discrete choice models. They demonstrate their results by computing identified sets for some parametric models in numerical simulations. As we explain further in Appendix A, applying their approach to a nonparametric model is infeasible. Compiani (2019) develops a nonparametric estimator and applies it to study consumer demand for strawberries in California using aggregated scanner data. His approach is based on identification arguments developed by Berry and Haile (2014), which use assumptions different than ours.

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. 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] We do not model supply, so all of these estimates should be interpreted as holding insurers' decisions fixed. 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 (2019), who also found that logit models underestimate price elasticities relative to less parametric alternatives, albeit using different methods in different empirical settings.

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 individuals 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 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.[4] Unlike other marketplaces, the insurer must provide the entire menu of four plans in any region where it enters.[5] 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.[6] The insurer is not permitted to adjust premiums based on any other characteristic of the consumer.[7] Premiums are therefore a deterministic function of a consumer's age and resident rating region.

[4] There is a fifth coverage tier called minimum (or catastrophic) coverage. This tier is not available to the subsidized buyers we focus on (with a few, rare exceptions), so we omit it from the analysis.

[5] In other ACA marketplaces, insurers are required to offer one Silver and one Gold plan, while additional plans are optional.

[6] 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 Medicare and Medicaid Services. See Orsini and Tebaldi (2017) for further discussion. Individuals 65 and older are covered by Medicare.

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

Table 1: *Standardized Plan 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 .
(*): Actuarial value (AV) is advertised to consumers as a percentage of medical expenses covered by the plan.

Individuals with household 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 individuals, since they constitute a large group of key policy interest.[8] The premium subsidies vary across individuals according to federal regulations. These ensure that the subsidized premium of the second-cheapest Silver plan is lower than a maximum affordable amount that varies by household income.[9] Post-subsidy premiums are therefore a deterministic function of a consumer's age, resident rating region, and household income.

In addition to premium subsidies, the ACA also provides cost-sharing reductions (CSRs) for individuals with household income lower than 250% of the FPL. CSRs are implemented by changing the actuarial terms of the Silver plan for eligible individuals according to their income, with discrete changes at 150%, 200%, and 250% of the FPL; see Table 1. CSRs make Silver plans very attractive for low-income individuals relative to the more expensive Gold and Platinum plans.

To further incentivize insurance uptake, the ACA had a universal coverage mandate which determined an income tax penalty for remaining uninsured. We treat this tax penalty as affecting the value of the outside option of not purchasing any Covered

[8] In 2014, this group comprised nearly 90% of contracts in Covered California.

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

California plan. The universal mandate was weak in 2014, and generally unenforced between 2014–2017 (Miller, 2017). It was repealed under the Tax Cuts and Jobs Act of 2017.

3 Empirical Methodology

3.1 Nonparametric Discrete Choice 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 $V_{i0} = 0$.[10] 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 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 et al. (2010a), Einav et al. (2010b), and Bundorf et al. (2012). In Appendix B, we derive (1) from an insurance choice model similar to the ones in Handel (2013) and Handel et al. (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 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

[10] Choosing $j = 0$ may incur a tax penalty due to the universal coverage mandate. Normalizing $V_{i0} = 0$ means that V_{ij} also incorporates the value of not facing the tax penalty.

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 3.5.1 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 unobservables responsible for price endogeneity are held fixed in the counterfactual (see e.g. Berry and Haile, 2010, pg. 5). To emphasize this, we let M_i denote consumer i's market, and we treat M_i as separate 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.[11] 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 for considering (1) is to preserve the utility maximization structure in (2), while avoiding these types of parametric assumptions.[12]

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

[11] 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.

[12] Fox et al. (2012) provide conditions under which the distribution of $(\beta_{im}, \alpha_{im})$ is nonparametrically point identified, and Fox et al. (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.

aggregated to the (j, m) level, and the data is viewed as drawn from a population of markets and/or products (Berry et al., 2004b; Armstrong, 2016). Our analysis presumes richer individual-level choice data as in Berry et al. (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 subsuming m subscripts into i subscripts, (1) can be seen to nest (2) by dividing through by α_i and taking $V_{ij} \equiv \alpha_i^{-1}(X_{ij}'\beta_i + \xi_{ij} + \epsilon_{ij})$.[13] This relationship highlights two 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 unobserved product characteristics in consumer i's market (Berry, 1994). We address this by conditioning on the market, M_i, after which ξ_i is nonstochastic. Second, 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 Price Variation

In Covered California, post-subsidy premiums are a deterministic function of the market, M_i, and consumer demographics, X_i. We denote this function by $P_i \equiv \pi(M_i, X_i)$.

Throughout the paper, our estimates of demand condition on the market, so the price variation we use for identification comes from variation across consumer demographics within a market. This could be problematic if these characteristics are related to valuations, V_i. Our empirical strategy, which we describe in more detail later, will be to use demographic variation only within relatively homogenous groups of consumers, so that valuations can be reasonably assumed to be independent of prices within these groups.

Our setting is different than many discrete choice applications in which prices only vary at the market level, such as Berry et al. (1995) or Nevo (2001). In terms of our notation, these settings would have $\pi(M_i, X_i)$ constant in X_i. The methodology we develop in the main text is not immediately useful for this case. In Appendix C, we propose two ways in which one can extend our approach to handle more aggregated price variation. One proposal uses within-market variation in non-price product or consumer characteristics, as in Berry and Haile (2010), while the other uses an index restriction, as in Berry and Haile (2014).

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

3.4 Target Parameters

The primitive object in model (1) is the distribution of valuations, V_i, conditional on 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 $f(\cdot|m,x)$ for each realization $M_i = m$, and $X_i = x$.[14]

The density f is a key object in the following. Common counterfactual quantities of interest can be written as integrals or sums of integrals of f (see e.g. Section 4.2 of Berry and Haile, 2014, or Section 3.4.1 of Berry and Haile, 2015). For example, a natural counterfactual quantity is the proportion of consumers who would choose plan j at a new 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 the market, m, is being held fixed in the counterfactual.

Conceptually, we view both (3) and (4) as scalar-valued functionals (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 government spending on premium subsidies.

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, the target parameter is scalar-valued, so that $d_\theta = 1$.

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

However, we will also consider cases with $d_\theta > 1$, for example to understand the joint identified set for two related target parameters, 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 assumptions.

3.5 Assumptions

We augment (1) with two types of assumptions. The first assumption is that one or more components of X_i are suitable instruments. The second assumption is that the density of valuations has support contained within a known set.

3.5.1 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 consists of instruments that satisfy an exogeneity assumption discussed ahead. This exogeneity assumption will be conditional on W_i, which are viewed as control variables. Note that W_i 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.

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 therefore 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.[15] However, in our data this seems unlikely to be the case. For this reason, we consider the partial identification framework discussed ahead. 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.

3.5.2 Support

The second assumption we use is that the support of f is concentrated on a known set. For each realization of W_i, defined as in the previous section, we choose a set $\mathcal{V}^\bullet(w)$ and then assume that f is such that

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

By choosing $\mathcal{V}^\bullet(w) = \mathbb{R}^J$, one can make this assumption trivially satisfied.

We use (7) to exploit the vertical structure of the ACA. For example, a Platinum plan is actuarially more generous than a Bronze plan (see Table 1). We can use (7) to impose the assumption that consumers would always prefer Platinum ($j = 4$) to Bronze ($j = 1$) at equal premiums by taking $\mathcal{V}^\bullet(w) = \{v \in \mathbb{R}^J : v_4 \geq v_1\}$. Since $\mathcal{V}^\bullet(w)$ depends on w, we can allow the definition of this set to change with income, which allows us to account for CSRs. We list the support assumptions we use for the application in Section 4.2.

3.6 The Identified Set

We now define 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 the 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, the

[15] 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).

identification analysis is premised on the thought experiment of perfect knowledge of these choice shares.

Each density of valuations implies a set of choice shares. 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 determined by the mass that f places on $\mathcal{V}_j(p)$ when prices are $p = \pi(m, x)$. We denote these implied choice shares by

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

A density f is consistent with the observed choice shares if

$$s_j(m, x; f) = s_j(m, x) \quad \text{for all } j,\ m \text{ and } x. \tag{11}$$

The identified set of valuation densities is the set of all f that both match the observed choice shares 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 support assumptions (6) and (7). It is the 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 observed in the data, as well as all premiums that are required to compute the target parameter. We call this collection of sets 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.

3.7 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, formal definition of the MRP is given in Section 3.9.

Suppose that the data consists of a single observed premium vector, p^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.[16] 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.[17]

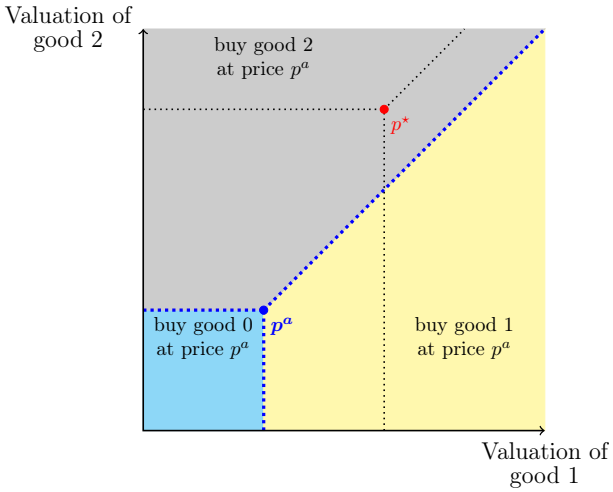
The MRP is minimal in the sense that any two consumers who have valuations in the same set would exhibit the same choice behavior under both premiums p^a and $p^\star$. Conversely, any two consumers with valuations in different sets would exhibit different choice behavior under at least one of these 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 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.

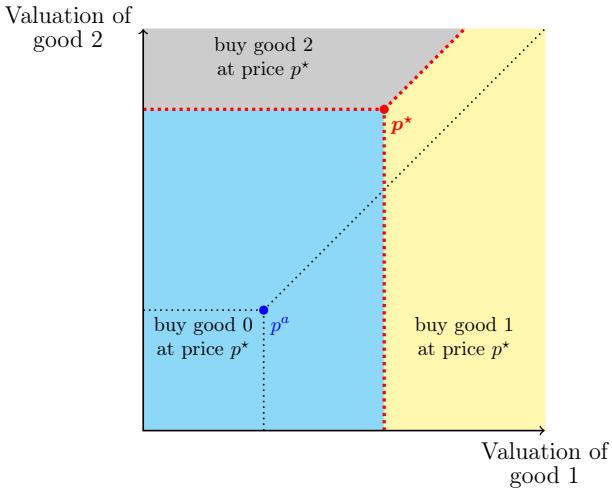
[16] Diagrams like panel (a) appear frequently in the literature on discrete choice, see e.g. Thompson (1989, Figure 1), Chesher et al. (2013, Figure 1), or Berry and Haile (2014, Figure 1).

[17] The MRP is related to the class of core-determining sets derived by Chesher et al. (2013). Comparing our Figure 1c to their Figures 2–3 shows that the MRP is a strict subset of the class of core-determining sets, since the latter also includes all connected unions of sets in the MRP.

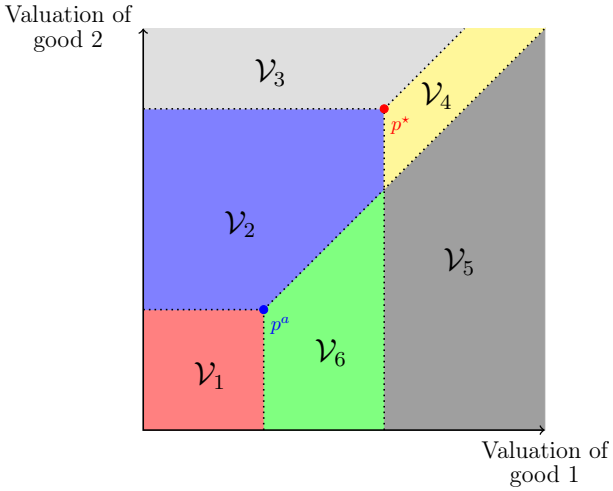
Figure 1: *Partitioning the Space of Valuations*



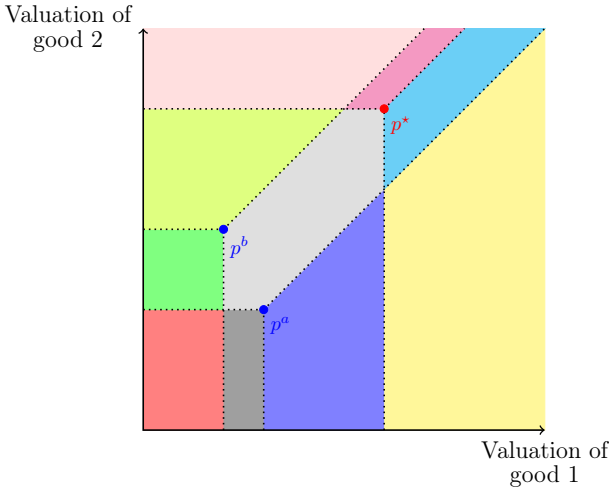
(a) *Choices if prices were p^a.*



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



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



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

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 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

$$s_1(m, x^a; f) = \int_{\mathcal{V}_5 \cup \mathcal{V}_6} f(v|m, x^a)\, dv = \int_{\mathcal{V}_5} f(v|m, x^a)\, dv + \int_{\mathcal{V}_6} f(v|m, x^a)\, dv,$$

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

$$s_2(m, x^a; f) = \int_{\mathcal{V}_2 \cup \mathcal{V}_3 \cup \mathcal{V}_4} f(v|m, x^a)\, dv.$$

This allows us to simplify the determination of whether a given f reproduces the observed choice shares by considering only the total mass that f places on sets in the MRP, without having to be concerned with how this mass is distributed within these sets.

Since we included $p^\star$ when constructing the MRP, the same is also true when considering target parameters θ that measure choice behavior at $p^\star$. For example, suppose that the 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 factors that might be correlated with price.

3.8 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 illustration, we assume that X_i is exogenous, i.e. we limit attention to f for which $f(v|m, x^a) = f(v|m, x^\star) = f(v|m)$. In terms of (6), this corresponds to $W_i = M_i$ and $Z_i = X_i$. In this case, (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

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

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):

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

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 $t^\star_{\text{ub}} = .66$. To find the smallest value of $\theta(f)$ we solve the analogous minimization problem, the optimal value of which we call $t^\star_{\text{lb}}$. In this example, $t^\star_{\text{lb}} = 0$.

In the next section, we formally prove that $\Theta^\star = [t^\star_{\text{lb}}, t^\star_{\text{ub}}]$. 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. A sketch of the proof 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 $t \in [t^\star_{\text{lb}}, t^\star_{\text{ub}}]$, there exists a set of numbers $\{\phi_l\}_{l=1}^6$ satisfying the constraints in (16), and such that $\phi_3 = t$.[18] 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, the sharp identified set for this target parameter is $\Theta^\star = [t^\star_{\text{lb}}, t^\star_{\text{ub}}]$. Intuitively, the reason there is no loss of information from reducing f to $\{\phi_l\}_{l=1}^6$ is that the MRP was constructed to represent all relevant differences in economic behavior.

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 matching the observed shares at x^a through (16), these variables will also need to match 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, $t^\star_{\text{ub}}$. The fact that it is smaller than the previous upper bound reflects the additional information contained in choice shares at p^b. The lower bound, $t^\star_{\text{lb}}$, is still zero, because it is still possible to match the observed choice shares for p^a and p^b by concentrating all mass southwest 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 $t^\star_{\text{ub}}$ and $t^\star_{\text{lb}}$.[19] In practice, we also do not assume that $f(v|m, x)$ is invariant in x. This makes a graphical interpretation

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

[19] In particular, we use Gurobi (Gurobi Optimization, 2015) and check a subset of the results using CPLEX (IBM, 2010). We formulate and presolve the problems using AMPL (Fourer et al., 2002).

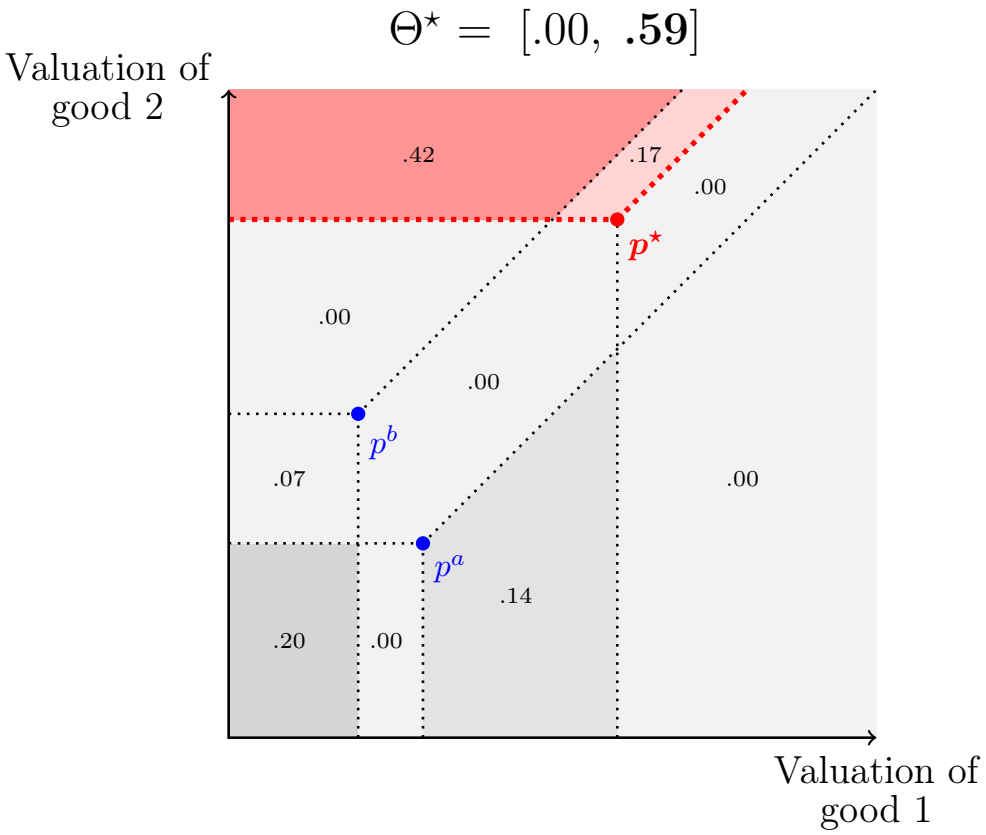


Figure 2: *The numbers in each set show a solution to the linear program when the target parameter is the proportion of consumers who choose plan 2 at $p^\star$ and the objective is to find the upper bound (maximize) this proportion. Matching the share of consumers who choose the outside option at the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

unwieldy, 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.

3.9 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 on the target parameter under the instrumental variable and support 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 discussion.

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 observed support of premiums. 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, $\mathcal{P} = \{p^a, p^b, p^\star\}$.[20]

The choice of which additional points to include in $\mathcal{P}$ is determined by the target parameter, θ. In Figure 1, the 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.

Definition MRP. *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 MRP creates a collection of sets $\mathbb{V}$ that is minimal in the sense that any two consumers who have valuations in a set in $\mathbb{V}$ would exhibit the same choice behavior for every premium vector in $\mathcal{P}$.[21] 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.[22]

[20] When implementing our methodology, we estimate demand separately for each market m, and thus also construct the set $\mathcal{P}$ separately for each market. We suppress this dependence in the notation because it does not affect our characterization of the identified set.

[21] Note that $\mathbb{V}$ depends on $\mathcal{P}$. We do not make this explicit in the notation because the following discussion only considers a single premium set, $\mathcal{P}$, and the single MRP it generates, $\mathbb{V}$.

[22] 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 MRP. Second, and for the same reason, although we have described the MRP as "the" MRP, it is not 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 the analysis given our focus on continuously distributed valuations.

The utility of the MRP 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, 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.[23] 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, this can be written as the sum of the masses that f places on sets in $\mathbb{V}_j(p)$, that is

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

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 $\mathbb{V} \times \text{supp}(M_i, X_i)$. Such a function ϕ can be viewed as an element of $\mathbb{R}^{d_\phi}$, where d_ϕ is the cardinality of its domain. Let $\mathbb{R}_+^{d_\phi}$ denote the subset of $\mathbb{R}^{d_\phi}$ whose elements are all non-negative and define

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

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

Each density f generates a mass function $\overline{\phi}(f) \in \Phi$ defined by

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

We assume that the value of the target parameter for any f is fully determined by $\overline{\phi}(f)$. Formally, the assumption is that there exists a known function $\overline{\theta}$ with domain Φ such that $\theta(f) = \overline{\theta}(\overline{\phi}(f))$ 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, θ.

To impose the instrumental variable assumption (6), we define for any $\phi \in \Phi$ the function

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

where W_i and Z_i are as in the statement of that condition. Similarly, to impose the

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

support assumption, we let $\mathbb{V}^\bullet(w)$ denote the subset of $\mathbb{V}$ that intersects $\mathcal{V}^\bullet(w)$, i.e.

$$\mathbb{V}^\bullet(w) \equiv \{\mathcal{V} \in \mathbb{V} : \lambda(\mathcal{V} \cap \mathcal{V}^\bullet(w)) > 0\}, \tag{22}$$

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. These equations replicate (6), (7), and (11) at a hypothesized parameter value, but in terms of the finite-dimensional mass function, ϕ, rather than the infinite-dimensional density, f. The interpretation of the result is that this dimension reduction entails no loss of information. A proof is in Appendix E.

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*

$$\overline{\theta}(\phi) = t, \tag{23}$$

$$\sum_{\mathcal{V} \in \mathbb{V}_j(\pi(m,x))} \phi(\mathcal{V}|m,x) = s_j(m,x) \quad \textit{for all } j \in \mathcal{J} \textit{ and } (m,x), \tag{24}$$

$$\phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z) = \phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z') \quad \textit{for all } z, z', w, \textit{ and } \mathcal{V}, \tag{25}$$

$$\textit{and} \quad \sum_{\mathcal{V} \in \mathbb{V}^\bullet(w)} \phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z) = 1 \quad \textit{for all } w, z. \tag{26}$$

Observe that each of (24)–(26) are linear in ϕ.[24] 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.[25] An implication of 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, also proved in Appendix E.

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

$$t^\star_{lb} \equiv \min_{\phi \in \Phi} \ \overline{\theta}(\phi) \qquad \textit{subject to } (24)\text{–}(26), \tag{27}$$

and with $t^\star_{ub}$ defined as the solution to the analogous maximization problem.

[24] This requires noting from (21) that $\phi_{\mathbb{V}|WZ}(\mathcal{V}|w,z)$ is itself a linear function of ϕ.

[25] For demand this is clear from e.g. (15). Consumer surplus (or changes in it) can be seen to be linear in f from (4). However, constructing $\overline{\theta}$ for consumer surplus is less obvious. We discuss how this is done in Appendix F, and we show there that the resulting $\overline{\theta}$ function is linear in ϕ.

3.10 Estimation

Our analysis thus far has concerned the identification problem in 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 a linear $\overline{\theta}$. A formal justification for this procedure is developed in Mogstad et al. (2018).

The estimator proceeds in two steps. First, we find the best fit to the observed choice shares by solving

$$\widehat{Q}^\star \equiv \min_{\phi \in \Phi} \widehat{Q}(\phi) \qquad \text{subject to (25) and (26)},$$

$$\text{where} \quad \widehat{Q}(\phi) \equiv \sum_{j,m,x} \widehat{\mathbb{P}}[M_i = m, X_i = x] \left| \widehat{s}_j(m,x) - \sum_{\mathcal{V} \in \mathbb{V}_j(\pi(m,x))} \phi(\mathcal{V}|m,x) \right|, \tag{28}$$

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 (28) can be reformulated as a linear program by replacing terms in absolute values by the sum of their positive and negative parts.[26] 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 $\overline{\theta}(\phi)$ for ϕ that come close to minimizing (28). That is, we construct the set:

$$\widehat{\Theta}^\star \equiv \left\{ \overline{\theta}(\phi) : \phi \in \Phi, \ \phi \text{ satisfies (25), (26), and } \widehat{Q}(\phi) \leq \widehat{Q}^\star + \eta, \right\}. \tag{29}$$

The qualifier "close" here reflects the tuning parameter η, which must converge to zero at an appropriate rate with the sample size. The purpose of this tuning parameter is to smooth out potential discontinuities caused by set convergence. In our empirical estimates, we set $\eta = .01$, and found very little sensitivity to values of η that were bigger or smaller by an order of magnitude. However, there are currently no theoretical results to guide the choice of this parameter.

We construct $\widehat{\Theta}^\star$ by solving two linear programs that replace (24) with the condition

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

in (29). That is, we solve

$$\widehat{t}^{\star}_{\text{lb}} \equiv \min_{\phi \in \Phi} \overline{\theta}(\phi) \quad \text{subject to (25), (26), and } \widehat{Q}(\phi) \leq \widehat{Q}^{\star} + \eta, \tag{30}$$

and an analogous maximization problem for $\widehat{t}^{\star}_{\text{ub}}$. The set estimator for $\Theta^{\star}$ is then $\widehat{\Theta}^{\star} \equiv [\widehat{t}^{\star}_{\text{lb}}, \widehat{t}^{\star}_{\text{ub}}]$. When $\overline{\theta}$ is linear, (30) can be reformulated as a linear program, again by appropriately redefining the absolute value terms. In this case, the overall procedure of the estimator is to solve three linear programs: One for (28), one for (30), and one for the analogous maximization problem.

4 Demand in Covered California

4.1 Data

Our primary data are administrative records on the universe of individuals who purchased a plan through Covered California in 2014. The data contain unique person and household identifiers for each individual, as well as their age, income measured in percentage of the FPL, gender, zipcode of residence, choice of plan, and premium paid. Since post-subsidy premiums are a deterministic function of demographics (see Section 2), this information also allows us to calculate premiums for plans a consumer did not choose. We focus on the subpopulation of subsidy-eligible adults aged 27–64 with household income between 140 and 400% of the FPL. This comprises 73% of enrollees in Covered California.[27]

We characterize each individual i by their resident rating region (market), M_i, and a vector X_i of observables consisting of their age and household income. We discretize age into 38 single-year bins running from 27 to 64, and household income into 52 FPL bins that are 5% wide.[28] When crossed with the 19 rating regions in Covered California, this yields 37,544 unique rating region $\times$ age $\times$ income bins of the observable characteristics, (M_i, X_i).

As in most demand analyses, we do not directly observe individuals who chose the outside option of not purchasing a plan through Covered California. This means that we need to transform data on quantities chosen for the inside options into choice shares by estimating the number of potential buyers. To do this, we use the 2011–2013 American Community Survey public use file (via IPUMS, Ruggles et al., 2015) to estimate the number of subsidy-eligible buyers not covered by employer-sponsored or

[27] Out of 1,291,214 covered individuals, 211,093 (16%) are dependents, younger than 26, while 137,714 (11%) are not beneficiaries of premium subsidies.

[28] The FPL bins are $[140, 145), [145, 150), \ldots, [395, 400]$.

public insurance for each (M_i, X_i) bin. Our estimation procedure for this part uses a flexible parametric model and is similar to procedures used by Finkelstein et al. (2019) and Tebaldi (2017). More detail is provided in Appendix G.

We combine the estimates of potential buyers with the administrative data to construct choice shares for each of the region $\times$ age $\times$ income bins. For 7,517 of these bins, we observe more enrollees in the administrative data than we estimate as potential buyers. We drop these bins and use the remaining 30,027 bins as the main estimation sample.[29] Since the number of individuals per bin varies greatly, we will report parameters that average over (M_i, X_i), and therefore put greater weight on larger bins.

Our analysis is focused on an individual's choice of coverage level (metal tier). Thus, $J = 4$, with $j = 1, 2, 3, 4$ denoting Bronze, Silver, Gold, and Platinum, respectively, and $j = 0$ denoting the outside option, as usual. The implicit assumption here is that the choice of coverage level is separable from the 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 financial characteristics of the tiers—do not vary by insurer. We define premiums P_{ij} for each tier j in each bin as the median post-subsidy premium across insurers.[30]

Table 2 provides some summary statistics. Each bin contains on average 85 potential buyers. The average participation rate in Covered California is 28%, and varies widely across rating regions and demographics, with a standard deviation across bins of 21%. 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 (see Table 1). Over 30% of such consumers purchase a Silver plan, whereas among consumers with income over 250% of the FPL, fewer than 9% purchase the more expensive and less generous non-CSR 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 costs 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 provider

[29] The bins that are dropped tend to have a small number of estimated potential buyers.

[30] We have also estimated a subset of the results using other measures of price, such as the mean, minimum, and second-cheapest premiums across insurers. The estimates turn out to be fairly insensitive to this choice.

Table 2: *Summary Statistics*

Panel (a): Data by region, age, income

	Obs. (# of bins)	Mean	St. Dev.	P-10	Median	P-90
Number of buyers$^{(\star)}$	30,027	85.27	90.86	14	55	194
Age	30,027	43.41	10.70	29	43	59
Income (FPL%)	30,027	243.98	72.05	155	230	355
Takeup rate	30,027	0.280	0.208	0.053	0.235	0.576
Average premium paid	30,027	175.51	89.06	69	163	298
Share choosing Bronze	30,027	0.065	0.073	0	0.045	0.147
Share choosing Silver	30,027	0.188	0.173	0.018	0.139	0.424
Share choosing Gold	30,027	0.015	0.021	0	0.009	0.038
Share choosing Platinum	30,027	0.012	0.018	0	0.007	0.030

Panel (b): Heterogeneity by age and income

	Bronze		Silver		Gold		Platinum	
	Premium	Share	Premium	Share	Premium	Share	Premium	Share
By age:								
27-34	120	0.050	175	0.122	229	0.010	271	0.009
35-49	118	0.058	182	0.175	248	0.013	300	0.011
50-64	105	0.086	210	0.259	321	0.022	409	0.016
By income (FPL%):								
140-150	5	0.011	59	0.338	133	0.005	191	0.006
150-200	29	0.046	95	0.318	170	0.008	229	0.009
200-250	87	0.084	164	0.193	241	0.018	302	0.015
250-400	197	0.074	278	0.084	357	0.019	419	0.014

Note: Each observation in panel (a) is a unique combination of rating region × age × income bins of the observable characteristics, (M_i, X_i). All statistics except the number of buyers are calculated across bins, weighted by number of buyers in each bin. Standard deviation refers to the standard deviation across bins of the within-bin median of the corresponding variable. In panel (b), premium is calculated as the average premium paid across buyers of a given age/income group, while market shares are calculated a proportion of potential buyers as estimated using the ACS.
(*): Number of buyers statistics are calculated across bins, not weighted by number of buyers.

networks. These factors are unobserved in the data, so we will not use variation in premiums across regions. That is, we define a market M_i to be a rating region, and we do not impose any restriction on how preferences (the density of valuations f) vary across markets.

Instead, we will assume—in a limited way—that preferences are locally invariant to 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 can use to help identify demand counterfactuals. The way in which premiums evolve with age

and income is prescribed by ACA regulations, so the behavior of insurers is not likely to be an important threat to this strategy. Rather, the 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 coarse aggregate of X_i bins. To do this, 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 the 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 due 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 individuals in the North Coast rating region who are aged between 36 and 40 with incomes between 150 and 200% of the FPL. Within this bin, we have 50 values of X_i, comprised of the 5 ages 36, 37, 38, 39, 40 crossed with the 10 income bins between 150 and 200 in steps of 5%. 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:

$$\underbrace{f_{V|WZ}(v|w,z) = f_{V|WZ}(v|w,z') \quad \text{for all } z, z',\ w, \text{ and } v.}_{\text{within a coarse bin } (W_i = w),\ \text{valuations are locally invariant to age and income } (Z_i = z, z')} \tag{31}$$

In Section 4.6, we relax assumption (6)/(31) to a strictly weaker "imperfect instrument" assumption that allows for some local variation with age and income.

The other assumption we maintain is the support condition (7), which we use to exploit the vertical ordering of plans in terms of actuarial generosity. We specify the sets $\mathcal{V}^\bullet(w)$ as follows:

$$\mathcal{V}^\bullet(w) = \begin{cases} \{v \in \mathbb{R}^4 : v_2 \geq v_4 \geq v_3 \geq v_1\} & \text{if } w \text{ has income below 150\% FPL} \\ \{v \in \mathbb{R}^4 : v_2 \geq v_1,\ v_4 \geq v_3 \geq v_1\} & \text{if } w \text{ has income in 150–200\% FPL} \\ \{v \in \mathbb{R}^4 : v_4 \geq v_2 \geq v_1,\ v_4 \geq v_3 \geq v_1\}, & \text{if } w \text{ has income in 200–250\% FPL} \\ \{v \in \mathbb{R}^4 : v_4 \geq v_3 \geq v_2 \geq v_1\} & \text{if } w \text{ has income above 250\% FPL} \end{cases}$$

This specification requires consumers to always prefer a plan that dominates on all actuarial characteristics. The different cases are needed to account for the CSRs,

which change at 150, 200 and 250% of the FPL (see Table 1). Note that in no case do we assume that any of the plans are preferred to the outside option, that is, we allow for some or all of the components of v to be smaller than $v_0 = 0$.

4.3 Counterfactual Prices

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

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 individual's premium from the observed premium, $P_i \equiv \pi(M_i, X_i)$, to a counterfactual premium, $P_i^\star \equiv \pi(M_i, X_i) + \delta$. For each coarse bin (value of W_i), we construct an MRP using the set of premiums formed from all P_i and $P_i^\star$ within that bin.

Figure 3 illustrates by plotting observed and counterfactual Bronze and Silver premiums for three counterfactuals. In Figure 3b, the counterfactual is a \$10 increase in the premium of the Bronze plan for all consumers, while Figure 3c illustrates a \$10 increase for the Silver plan. The former corresponds to $\delta = (10, 0, 0, 0)$, while the latter corresponds to $\delta = (0, 10, 0, 0)$. In Figure 3d, both the Bronze and Silver plan premiums increase by \$10, which corresponds to $\delta = (10, 10, 0, 0)$.

4.4 Demand Responses

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

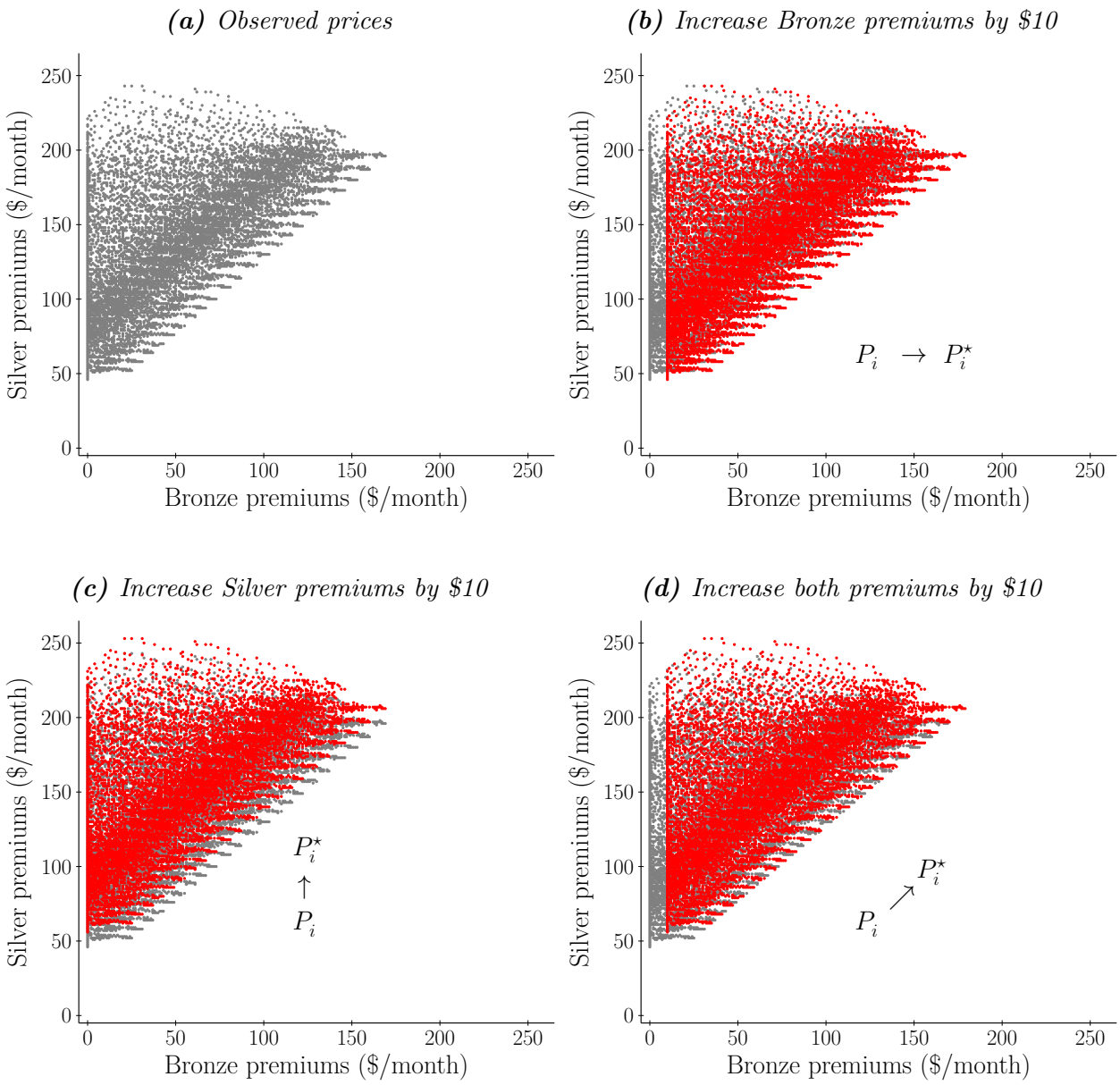
$$\Delta\text{Share}_j(m, x; f) \equiv \int_{\mathcal{V}_j(\pi(m,x)+\delta)} f(v|m, x)\, dv - \int_{\mathcal{V}_j(\pi(m,x))} f(v|m, x)\, dv, \tag{32}$$

where $\mathcal{V}_j(p)$ was defined in (9).[31] In order to aggregate (32) into a single measure, we average it over markets and demographics:

$$\Delta\text{Share}_j(f) \equiv \sum_{m,x} \Delta\text{Share}_j(m, x; f)\, \mathbb{P}[M_i = m, X_i = x]. \tag{33}$$

[31] Note that on the left-hand side of (32) we have omitted the dependence on the premium change, δ, since this will be clear from the way we present the results.

Figure 3: *Observed and Counterfactual Premiums*



Note: *The figure shows observed and counterfactual premiums of Bronze and Silver plans. Panel (a) plots the prices observed in the data in grey, where each observation is a unique region-age-income combination (N=30,027). 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, Upper and Lower Bounds*

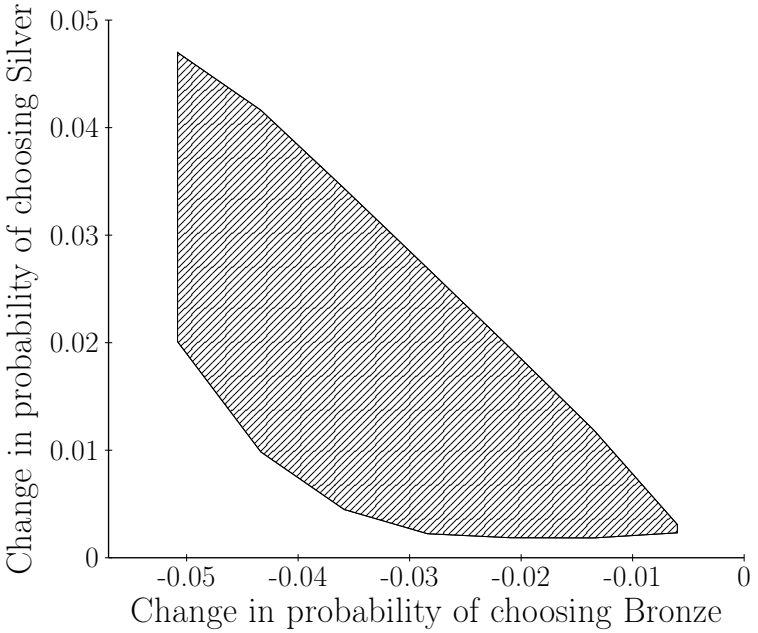
$10/month premium increase for	Change in probability of choosing: Bronze LB	Bronze UB	Silver LB	Silver UB	Gold LB	Gold UB	Platinum LB	Platinum UB	Any plan LB	Any plan UB
Panel (a): Full sample (140 - 400% FPL)										
Bronze	-0.051	-0.006	+0.002	+0.048	+0.000	+0.031	+0.000	+0.026	-0.013	-0.001
Silver	+0.000	+0.128	-0.170	-0.013	+0.000	+0.126	+0.000	+0.100	-0.052	-0.003
Gold	+0.000	+0.007	+0.000	+0.013	-0.016	-0.001	+0.000	+0.014	-0.004	-0.000
Platinum	+0.000	+0.005	+0.000	+0.008	+0.000	+0.012	-0.012	-0.001	-0.003	-0.000
All plans	-0.014	-0.003	-0.053	-0.010	-0.005	-0.001	-0.004	-0.000	-0.070	-0.016
Panel (b): Lower income (140 - 250% FPL)										
Bronze	-0.049	-0.006	+0.002	+0.047	+0.000	+0.030	+0.000	+0.025	-0.011	-0.001
Silver	+0.001	+0.184	-0.243	-0.017	+0.000	+0.178	+0.000	+0.144	-0.078	-0.004
Gold	+0.000	+0.006	+0.000	+0.011	-0.013	-0.001	+0.000	+0.012	-0.003	-0.000
Platinum	+0.000	+0.005	+0.000	+0.008	+0.000	+0.012	-0.012	-0.001	-0.003	-0.000
All plans	-0.012	-0.002	-0.080	-0.014	-0.004	-0.000	-0.004	-0.000	-0.093	-0.018
Panel (c): Higher income (250 - 400% FPL)										
Bronze	-0.053	-0.006	+0.001	+0.049	+0.000	+0.032	+0.000	+0.027	-0.015	-0.002
Silver	+0.000	+0.058	-0.077	-0.008	+0.000	+0.059	+0.000	+0.044	-0.019	-0.001
Gold	+0.000	+0.009	+0.000	+0.015	-0.019	-0.002	+0.000	+0.016	-0.005	-0.000
Platinum	+0.000	+0.005	+0.000	+0.008	+0.000	+0.012	-0.012	-0.001	-0.003	-0.000
All plans	-0.016	-0.004	-0.020	-0.005	-0.006	-0.001	-0.004	-0.000	-0.040	-0.014

In the notation of Section 3, ΔShare_j (either averaged or conditional) is an example of a target parameter, θ.

Table 3 reports estimated bounds for ΔShare_j across the four metal tiers together with bounds on overall participation, i.e. on $1 - \Delta\text{Share}_0$. 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, with the full sample in panel (a), we estimate that a simultaneous $10 increase in all premiums reduces the proportion of individuals who purchase coverage by between 1.6 and 7.0%. Panel (b) shows that these estimates are larger in magnitude for lower income individuals,

Figure 4: *Effect of Increasing Bronze Premiums by $10 on Bronze and Silver Choice Shares*



Note: *The figure shows the joint identified set for the effect of a $10 increase in Bronze monthly premiums on the choice probabilities of Bronze and Silver plans. To construct the set, we take a grid of equidistant points between the estimated upper and lower bounds for the change in Bronze choice shares. At each point in the grid, we find bounds on the change in Silver, while fixing the change in Bronze to be the value at the grid point.*

at between 1.8 and 9.3%, and panel (c) shows that they are smaller in magnitude for higher income individuals, who we estimate would reduce participation in Covered California by between 1.4 and 4.0%. Comparing panels (b) and (c) more generally, we find a pattern of higher price sensitivity for lower income enrollees.

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 0.6 and 5.1% in the share of consumers choosing Bronze coverage, and an increase in the share choosing Silver of between 0.2 and 4.8%. 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 in participation for a Bronze premium increase is between 0.1 and 1.3%, 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, which we measure to be at most 0.3%. Overall, Table 3 indicates substitution patterns inconsistent with the independence of irrelevant alternatives property of the logit model.

Table 3 reports estimated bounds obtained by considering ΔShare_j separately for each plan. However, changes in choice shares for different plans are tightly related to one another. For example, if the decrease in the share of Bronze in response to a Bronze premium increase is smaller, the increase in the share of other alternatives is likely to also be smaller, and vice versa. We can describe these patterns by plotting a joint identified set for the change in multiple choice shares in response to a given premium change, as in Figure 4. The shape of this set shows that among the range of responses that could result from an increase in Bronze premiums, larger decreases in the probability of choosing Bronze would be associated with larger increases in the probability of choosing Silver. For example, if the probability of choosing Bronze were to decrease by 5%, then the probability of choosing Silver would increase by at least 2%, implying that at least 40% of the individuals leaving Bronze would substitute to Silver.

In a partial identification framework, the width of the bounds reflects the amount of information that the data and assumptions yield about a specific counterfactual. The bounds for more ambitious counterfactuals will be wider than the bounds for more modest counterfactuals that are closer to what was observed in the data. This situation is evident in Figure 5, which plots the average extensive margin (enrollment) response as a function of a given increase or decrease in all premiums. The 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 the approach, since it reflects the increased 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.[32]

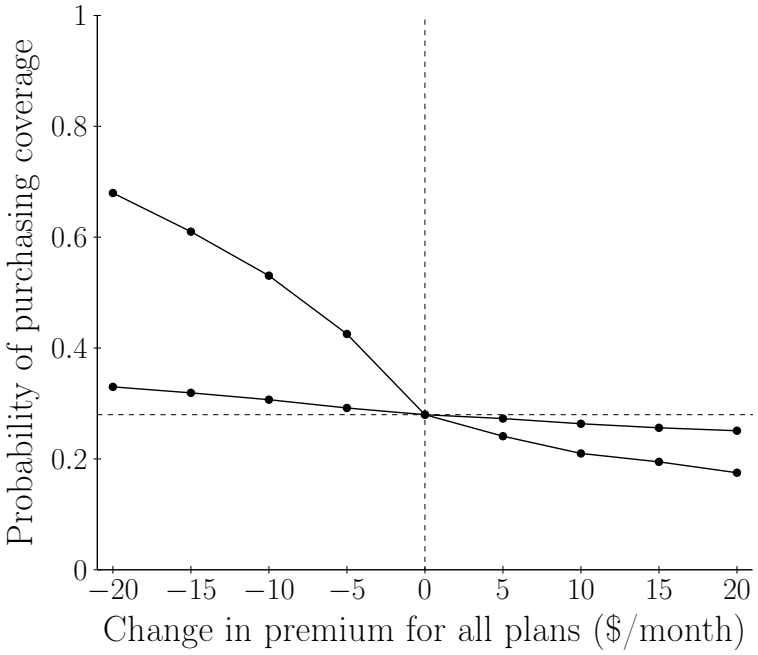
4.5 Consumer Surplus and Subsidy Expenditure

The second set of parameters we consider measure the effects of changing premium subsidies on consumer surplus and government spending. From the individual's perspective, a decrease in premium subsidies—which in terms of policy can be thought of as an increase in the ACA's "maximum affordable amount"—is the same as an increase in premiums faced.[33] Such a subsidy change generates an average change in consumer

[32] 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 effectively zero even for distant extrapolations.

[33] Our analysis here requires maintaining a partial equilibrium framework in which there are no other supply-side responses in base premiums 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.

Figure 5: *Extensive Margin Demand for Different Counterfactuals*



surplus for an individual in market m with characteristics x of

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

which we aggregate by averaging over markets and demographics into

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

We will contrast the change in consumer surplus to the change in government spending on premium subsidies. This is given by

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

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

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

Both ΔCS and ΔGS are examples of target parameters θ. Constructing bounds on ΔCS involves a two-step construction where we first determine bounds on the change in consumer surplus within each set in the MRP. We discuss this in detail in Appendix F.

Figure 6a depicts the bounds on ΔCS for a \$10 decrease in subsidies as the shaded areas to the left of the two demand curves. The lower bound on the decline 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 left of the steeper demand curve. Intuitively, the smallest decrease in consumer surplus is attained when price sensitivity is highest, while the largest decrease is attained when price sensitivity is lowest.[34] Figure 6b plots the joint identified set of the change in consumer surplus and the change in government spending for this same counterfactual of a \$10 decrease in subsidies. The fact that the joint set is not rectangular reflects the mutual dependence of the two parameters on the underlying price sensitivity; relatively large consumer surplus decreases will only happen under relatively small decreases in spending, and vice versa.

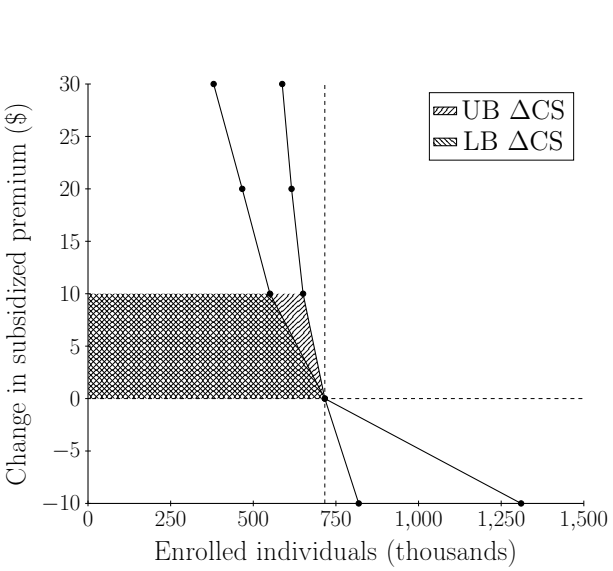
Table 4 summarizes the estimates illustrated in Figure 6. The first column shows estimated bounds using the entire sample, while the second and third columns split the estimates on income. In the fourth column of Table 4, we report estimated bounds on the corresponding reduction in government spending.

Our estimates imply that a \$10 decrease in monthly subsidies would lead to a reduction in average monthly consumer surplus of between \$1.99 and \$2.45 per person. The impacts for the lower-income sample (\$2.55–\$3.16) are estimated to be approximately twice as large as the impacts for the higher-income sample (\$1.27–\$1.55). This reflects the fact that individuals with income lower than 250% of the FPL have a higher uptake of insurance and are covered under more generous plans due to the CSRs.

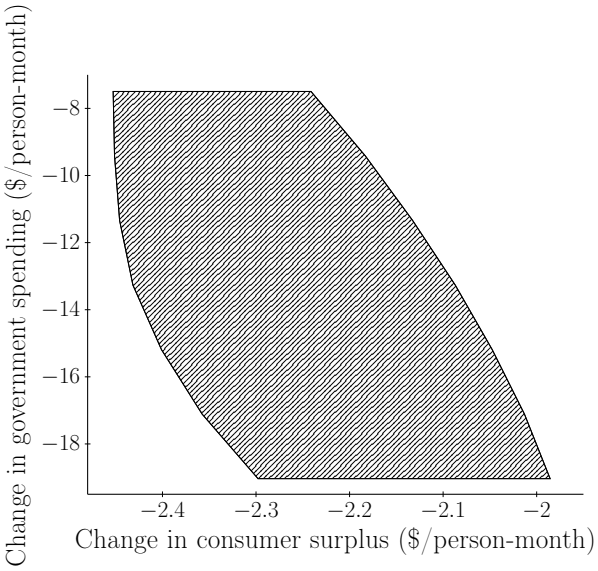
Our estimates of changes in consumer surplus are dwarfed by the corresponding change in government expenditure on premium subsidies, which we estimate to be between -\$7.50 and -\$19.03 per consumer, per month. The large magnitude of the expenditure savings is due to the marginal buyers who exit the market. When these

[34] Note that while the bounds on ΔCS shown here are sharp, the demand curves we have plotted are not unique, 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.

Figure 6: *Changes in Consumer Surplus and Government Spending*



(a) *Bounds on the change in consumer surplus.*



(b) *The joint identified set of consumer surplus and government spending.*

Table 4: *The Impacts of Reducing Premium Subsidies by $10 per Month*

	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 outlays	
	LB	UB	LB	UB	LB	UB	LB	UB
Average ($/person-month)	-2.45	-1.99	-3.16	-2.55	-1.55	-1.27	-19.03	-7.50
Aggregate ($ million/year)	-77.82	-62.99	-57.59	-46.48	-22.48	-18.33	-603.89	-237.80

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 of subsidies in Covered California. The total consumer surplus impact would be between $63 and $78 million, with the majority of the losses concentrated among individuals with income below 250% of the FPL. At the same time, government subsidy outlays would decline by between $238 and $604 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 is consistent with a growing number of empirical analyses, see e.g. Finkelstein et al.

(2019). One caveat is that our estimates do not account for potential sampling error. In Appendix H, we provide suggestive evidence that sampling error is unlikely to have a large impact on our bounds due to the large sample sizes we are considering. A more important caveat when interpreting our welfare estimates is that they 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 et al., 2018).

4.6 Relaxing the Instrumental Variable Assumptions

The assumption that drives our results is (31), which requires valuations to be independent of age and income within 5 year and 50% FPL bins. We view this local invariance assumption as reasonable for the relatively homogenous groups of individuals within these bins. However, it is unlikely to hold exactly. Local invariance to age will fail if valuations change with risk factors and risk factors change with age.[35] Local invariance to income gives the additive separability in (1) the interpretation of quasilinearity, but this will of course fail if there are general income effects.

In this section, we consider a strictly weaker version of (31) that allows for deviations away from perfect local invariance. We do this by relaxing (31) from an equality to two inequalities controlled by a slackness parameter. For age, the relaxed assumption is that

$$(1-\kappa_{\text{age}}(z,z'))f_{V|WZ}(v|w,z') \leq f_{V|WZ}(v|w,z) \leq (1+\kappa_{\text{age}}(z,z'))f_{V|WZ}(v|w,z') \\ \text{for all } z, z', w, \text{ and } v, \tag{34}$$

where $\kappa_{\text{age}}(z,z')$ is a function specified as

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

and $\kappa_{\text{age}} \geq 0$ is the slackness parameter. For income, we impose the analog of (34) with the roles of age and income swapped and a slackness parameter $\kappa_{\text{inc}} \geq 0$. We use these weaker forms to conduct a sensitivity analysis by varying κ_{age} and κ_{inc}, similar in spirit to Conley et al. (2010), Nevo and Rosen (2012), and Manski and Pepper (2017).

In words, (34) requires that within any coarse bin (i.e., conditional on $W_i = w$), the pointwise difference in conditional valuation densities for any two adjacent one-year age

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

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

Allowed variation in preference with age and income		Change in probability of purchasing coverage if all per-person premiums increase by $10/month		Change in consumer surplus ($/person-month) if per-person subsidies decrease by $10/month		Change in government spending ($/person-month) if per-person subsidies decrease by $10/month	
κ_{age}	κ_{inc}	LB	UB	LB	UB	LB	UB
0	0	-0.070	-0.016	-2.45	-1.99	-19.03	-7.50
0.2	0	-0.072	-0.017	-2.46	-1.98	-19.47	-7.48
0.6	0	-0.076	-0.019	-2.47	-1.96	-20.43	-7.70
$+\infty$	0	-0.089	-0.015	-2.51	-1.80	-23.92	-6.52
0	0.2	-0.075	-0.019	-2.47	-1.98	-20.22	-8.00
0	0.6	-0.089	-0.022	-2.48	-1.92	-23.36	-8.72
0	$+\infty$	-0.147	-0.021	-2.53	-1.44	-39.01	-8.26
0.2	0.2	-0.098	-0.023	-2.52	-1.92	-25.90	-9.35
0.6	0.6	-0.154	-0.015	-2.66	-1.65	-40.50	-7.71
$+\infty$	$+\infty$	-0.280	-0.000	-2.80	-0.00	-72.56	-2.70

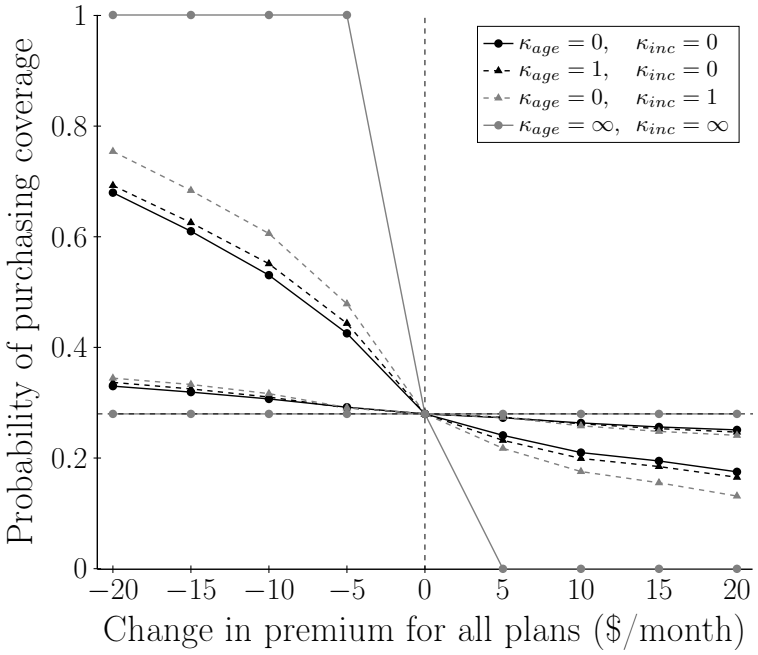
bins (with identical income) can be no greater than $(\kappa_{\text{age}} \times 100)\%$. Taking $\kappa_{\text{age}} = \kappa_{\text{inc}} = 0$ corresponds to the previous assumption of (31). Alternatively, taking $\kappa_{\text{age}} = +\infty$ and $\kappa_{\text{inc}} = 0$ completely relaxes the age invariance restriction while retaining income invariance.

Table 5 reports bounds on some key target parameters for different values of κ_{age} and κ_{inc}.[36] The top row with $\kappa_{\text{age}} = 0$ and $\kappa_{\text{inc}} = 0$ is the same as the estimates reported in the previous section. At the opposite extremes, the row with $\kappa_{\text{age}} = +\infty$ uses only variation in income, the row with $\kappa_{\text{inc}} = +\infty$ uses only variation in age, and the row with $\kappa_{\text{age}} = \kappa_{\text{inc}} = +\infty$ uses neither. Values of κ_{age} and κ_{inc} in between limit the amount by which adjacent bins can differ, with larger values of these parameters representing strictly weaker identifying assumptions.[37]

[36] Note that it is straightforward to modify the sharp characterization in Proposition 1 to allow for an assumption like (34) instead of (31)/(6). The difference in implementation just amounts to replacing (25) with two appropriate inequalities.

[37] The bounds tend to widen with increases in κ_{age} and κ_{inc}, 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 Section 3.10. That procedure works by restricting attention to densities that come closest to fitting the observed choice shares. This fit mechanically improves as κ_{age} or κ_{inc} increases, because a

Figure 7: *Extensive Margin Demand Relaxing Exclusion Restrictions*

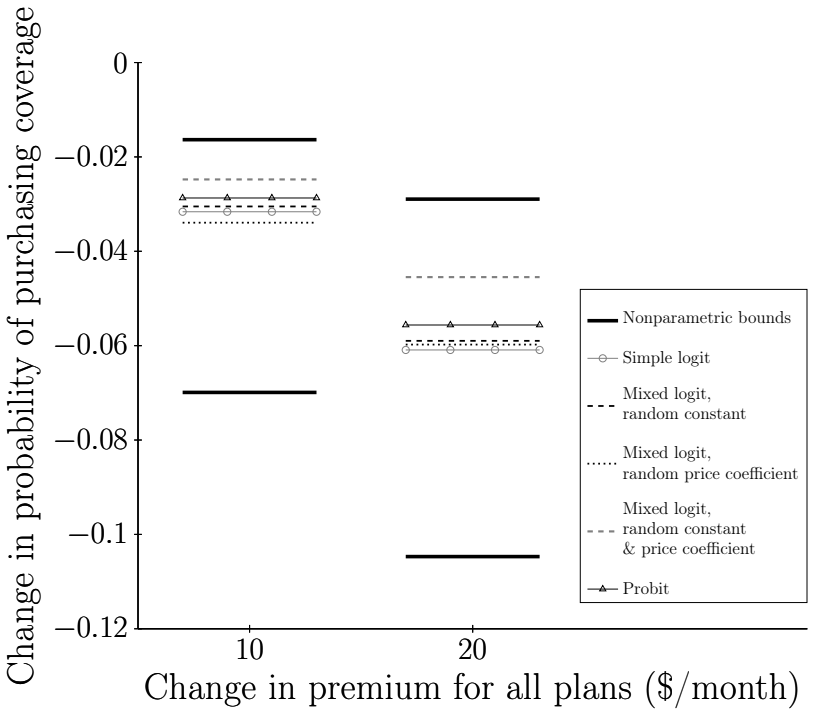


Overall, Table 5 suggests the estimated bounds in the previous section are quite robust to violations in either the age or income invariance assumption. If we completely drop age invariance, we obtain bounds that are not much wider. If we completely drop income invariance, the bounds widen significantly, but only at the upper end of the price sensitivity. As we will see in the next section, comparable parametric models produce estimates of price sensitivity that tend to be small in magnitude. Overall, our estimates of the impacts of a $10 decrease in subsidies remain qualitatively similar if we drop either age or income invariance separately. Figure 7 shows that extensive margin changes also remain similar under these specifications for different changes in subsidies.

The bounds widen more quickly when we relax both age and income invariance together in the third panel of Table 5. When both assumptions are completely removed, our bounds become completely uninformative: We cannot rule out that a $10 increase in monthly premiums causes all 28% of the population currently enrolled to exit the market. Setting $\kappa_{\text{age}} = \kappa_{\text{inc}} = .6$ allows for the density of valuations in adjacent age

larger class of valuation densities are considered. Densities that fit well for smaller values of the slackness parameters might be deemed to no longer fit well for larger values, since the best fit has improved. As a consequence, the best-fitting set of densities need not weakly increase, which can lead to non-monotonicity in the estimated bounds, even though monotonicity must hold for the population bounds.

Figure 8: *Extensive Margin: Nonparametric Bounds vs. Parametric Point Estimates*



and/or income bins to increase or decrease by 60%. This seems extremely conservative, and our bounds do widen significantly, yet they remain qualitatively similar to our estimates from the main specification. For a more modest relaxation, like $\kappa_{\text{age}} = \kappa_{\text{inc}} = .2$, our bounds are, for practical purposes, essentially unmoved from our baseline estimates.

5 Comparison to Estimates from Parametric Models

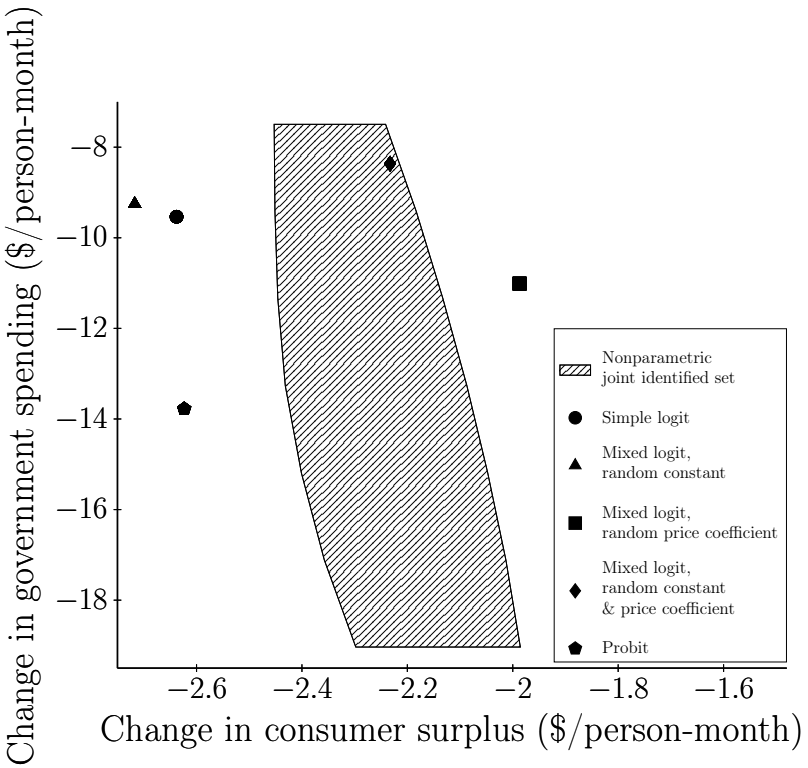
The motivation of this paper has been to construct estimates of key demand-side policy parameters using a model that does not impose parametric distributional assumptions. In this section, we compare the estimates of the nonparametric model to estimates from some fully parametric logit and probit models.

The models we consider all follow a specification similar to (2):

$$Y_i = \arg\max_{j \in \mathcal{J}} \mathbb{1}[j \geq 1] \left(\gamma_i + \beta_i \text{AV}_{ij} - \alpha_i P_{ij} + \xi_j\right) + \epsilon_{ij}, \tag{35}$$

where γ_i is an individual-specific intercept, AV_{ij} is the actuarial value of tier j for individual i (see Table 1), α_i and β_i are individual slope coefficients, and ξ_j are unobservable preference shifters for each tier. The indicator sets the contribution of these

Figure 9: *Consumer Surplus and Government Expenditure Changes from a $10 Decrease in Premium Subsidies: Nonparametric Bounds vs. Parametric Point Estimates*



terms to 0 for the outside option ($j = 0$). We consider logit models in which ϵ_{ij} is assumed to follow a type I extreme value distribution, independently across j, as well as a probit model in which this unobservable is assumed to follow a normal distribution. We always estimate (35) market-by-market, so that all parameters vary by market in an unrestricted way.

The first model we estimate is a logit in which the price parameter, α_i, is constant within markets, but both γ_i and β_i vary with observables in a rich way.[38] The second model is a probit with the same specification.[39] We then consider three mixed logit models. In each of these models, γ_i and β_i vary with observables in the same way as in the baseline model. The three models differ in whether γ_i, α_i, or both have an additional unobservable component that is normally distributed with unknown vari-

[38] In particular, the specification allows β_i 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_i to vary freely by market, and within each market restricts $\gamma_i = \gamma_i^{\text{Inc}} + \gamma_i^{\text{Age}}$, where γ_i^{Inc} varies in three FPL income bins {140–200, 200–250, 250–400}, and γ_i^{Age} varies in the same four age bins as β_i.

[39] The probit still has ϵ_{ij} independent across j. We had difficulty allowing for correlation across j because the likelihood is very flat.

ance. In the latter case, we also assume that the two unobservable components are uncorrelated.

Figure 8 illustrates how the nonparametric bounds for the extensive margin compare to the estimates one obtains from these five parametric models. The estimates shown are for \$10 and \$20 increases in monthly premiums. All of the point estimates are clustered together within the nonparametric bounds. For a \$10 increase, the estimates are towards the upper bound, where price sensitivity is the lowest, while for a \$20 increase they are more towards the center of the bounds. Our bounds are constructed so that any value within the upper and lower bound can be obtained by a distribution of valuations that fits the observed data equally well. Thus, one implication of Figure 8 is that distributional assumptions on ϵ_{ij} do in fact matter here, since substantially different conclusions could be obtained while fitting the data equally well.[40]

Figure 9 shows that the parametric models also make substantially different predictions for the consumer surplus and government spending impacts of a \$10 decrease in premium subsidies. Only the richest parametric model we consider (the mixed logit with random coefficients on the constant and price) falls within the nonparametric joint identified set. All of the parametric models estimate changes in government expenditure contained within the marginal nonparametric bounds.[41] However, only the two mixed logits with random price coefficients yield estimated consumer surplus changes within the marginal bounds, suggesting that non-random price coefficients lead to attenuated demand responses. Of these two models, the one without the random constant term (shown as a square in Figure 9) predicts a combination of changes in consumer surplus and government expenditure that is inconsistent with the nonparametric model.

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 policy-relevant target parameters.

[40] This conclusion appears to be quite robust to potential sampling error; see Appendix H.

[41] The marginal bounds can be seen by projecting the set in Figure 9 against the vertical axis. Alternatively, these bounds are reported in Table 4.

Our nonparametric estimates of demand point to the possibility of substantially greater price sensitivity than would be recognized using comparable parametric models. This is consistent with the folklore that logits are "flat" models. We showed that this has potentially important policy implications for the impact of decreasing subsidies on consumer surplus and government expenditure. 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 et al. (2006, 2015), Pakes and Porter (2016), Shi et al. (2016), and Khan et al. (2019). 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 include Thompson (1989), Manski (2007, 2014), Briesch et al. (2010), Chiong et al. (2017), and Allen and Rehbeck (2017). The importance of allowing for endogenous explanatory variables in discrete choice demand analysis was emphasized by Berry (1994) and Hausman et al.

[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 et al. (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). See also the discussion in Appendix C.

(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 correlated with unobserved components of a consumer's valuations.

Third, 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. Compiani (2019) uses these results to develop a nonparametric estimator and applies it to study the demand for strawberries in California supermarkets. 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 et al., 2013). Not maintaining these types of support and completeness conditions leads naturally to a partial identification framework (Santos, 2012).

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 et al. (2017) analyzed the problem of non-purchases in scanner data as one of partial identification.

Chesher et al. (2013) provide a general framework for deriving moment inequalities in partially identified discrete choice models.[44] They use random set theory to characterize identification, which leads to the concept of a core-determining class of sets. For the quasilinear utility models that we consider in this paper, the core-determining class is strictly larger than the collection of sets we call the minimal relevant partition (see footnote 17). However, the analysis of Chesher et al. (2013) also applies to

[44] Chesher and Rosen (2017) generalize this framework to an even broader class of models.

models that are not quasilinear in utility. Our methodology also differs from theirs in terms of computation. Whereas Chesher et al. (2013) provide moment inequalities that must be checked for *each* candidate distribution of valuations (f), our approach effectively profiles out this distribution in search of bounds on the finite-dimensional target parameter (θ). As a consequence, our approach can be implemented nonparametrically, whereas feasibly implementing the Chesher et al. (2013) approach requires parameterizing the distribution of valuations (see their Section 4.2).

More generally, our work is related to a literature on computational approaches to characterizing identified sets in the presence of partial identification.[45] In particular, linear programming has been used by many other authors in different contexts, see e.g. Balke and Pearl (1994, 1997) and Hansen et al. (1995) for early examples. Previous work that has used linear programming to characterize sharp identified sets includes Honoré and Tamer (2006), Honoré and Lleras-Muney (2006), Manski (2007, 2014), Lafférs (2013), Freyberger and Horowitz (2015), Demuynck (2015), Kline and Tartari (2016), Torgovitsky (2016, 2018), Kamat (2017), and Mogstad et al. (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 that 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 premiums (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

$$Y_i = \arg\max_{j \in \mathcal{J}} \int U_{ij}(e)\, dG_{ij}(e), \tag{36}$$

where $U_{ij}(e)$ is consumer i's ex-post utility from choosing plan j given realized expenditures of e, and G_{ij} is the distribution of these expenditures, which varies both by

[45] In addition to the series of papers by Chesher and Rosen (2013, 2014, 2017) and Chesher et al. (2013), this also includes work by Beresteanu et al. (2011), Galichon and Henry (2011), and Schennach (2014).

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

$$U_{ij}(e) = -\frac{1}{A_i}\exp(-A_i C_{ij}(e)), \tag{37}$$

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

$$C_{ij}(e) = \text{Inc}_i - P_{ij} - e + \widetilde{V}_{ij}, \tag{38}$$

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

Substituting (38) into (37) and then into (36), we obtain

$$Y_i = \underset{j\in\mathcal{J}}{\arg\max} -\frac{1}{A_i}\left[\exp(A_i(P_{ij} - \text{Inc}_i - \widetilde{V}_{ij}))\int \exp(A_i e)\, dG_{ij}(e)\right]$$

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

$$\begin{aligned} Y_i &= \underset{j\in\mathcal{J}}{\arg\max} -\log\left(\frac{1}{A_i}\left[\exp(A_i(P_{ij} - \text{Inc}_i - \widetilde{V}_{ij}))\int \exp(A_i e)\, dG_{ij}(e)\right]\right) \\ &= \underset{j\in\mathcal{J}}{\arg\max} -\log\left(\frac{1}{A_i}\right) + A_i\left(\text{Inc}_i - P_{ij} + \widetilde{V}_{ij}\right) + \log\left(\int \exp(A_i e)\, dG_{ij}(e)\right). \end{aligned}$$

Eliminating additive terms that do not depend on plan choice yields

$$Y_i = \underset{j\in\mathcal{J}}{\arg\max} -A_i P_{ij} + A_i\widetilde{V}_{ij} + \log\left(\int \exp(A_i e)\, dG_{ij}(e)\right).$$

Suppose that $A_i > 0$, so that all consumers are risk averse.[46] Then we can express the consumer's choice as

$$Y_i = \underset{j\in\mathcal{J}}{\arg\max}\left[\widetilde{V}_{ij} + \frac{1}{A_i}\log\left(\int \exp(A_i e)\, dG_{ij}(e)\right)\right] - P_{ij},$$

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

which takes the form of (1) with

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

Examining the components of V_{ij} reveals the factors that contribute to heterogeneous valuations in this model. Heterogeneity across i can come from variation in risk aversion (A_i), from differences in risk factors or beliefs (G_{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 corresponding distribution of expenditures (G_{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 (37) and the quasilinearity of ex-post consumption in (38). However, as noted in the main text, these restrictions do not have empirical content until they are combined with an assumption about the dependence between income (here called Inc_i) and the preference parameters, A_i and $\widetilde{V}_{ij}$.

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 factors that the researcher does not observe. In this case, our methodology can be applied with little more than notational changes. In addition to (M_i, X_i), one would also need to condition on P_i when defining the primitive distribution of valuations, f. 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 3.5. Condition (11) would be modified so that it is defined for all (p, m, x) in the support of (P_i, M_i, X_i).

The less straightforward (and more interesting) case is when one observes only a single price for each market, as in Berry et al. (1995) and 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 and can be made comparable to prices.[47] This is implicit in standard discrete choice models like (2). Consider modifying (1) to

$$Y_i = \arg\max_{j \in \mathcal{J}} V_{ij} + X_i' \beta_j - P_{ij}, \tag{39}$$

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, $\tilde{P}_{ij}(\beta_j)$ can 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 is a common assumption in empirical implementations of (2). In our framework, this independence can be incorporated by modifying the instrumental variable assumptions in Section 3.5.1.

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

$$Y_i = \arg\max_{j \in \mathcal{J}} V_{ij} + \xi_j(M_i) - P_{ij}, \tag{40}$$

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 product-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 variable assumptions in Section 3.5.1. While there is still only a single price per market, (40) together with such an independence assumption enables aggregation across markets by requiring the distribution of valuations to be the same up to a location shift.

Implementing either (39) or (40) 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 (39) or (40) are needed for our application, we leave fuller

[47] Berry and Haile (2010) show how such variables can be used to relax assumptions used in the nonparametric point identification arguments in Berry and Haile (2014).

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 subsets like $\{v \in \mathbb{R}^J : v_j - p_j = v_k - p_k\}$ 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 ASP. *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

$$\left\{\widetilde{\mathcal{V}}_y : y \in \mathcal{Y}\right\} \quad \text{where} \quad \widetilde{\mathcal{V}}_y \equiv \bigcap_{l=1}^{L} \mathcal{V}_{y_l}(p_l) \tag{41}$$

also constitutes an a.s. partition of $\mathbb{R}^J$.[48] Intuitively, each vector $y \equiv (y_1, \ldots, y_L)$ is a profile of L choices made under the price vectors $(p_1, \ldots, p_L)$ that comprise $\mathcal{P}$. Each set $\widetilde{\mathcal{V}}_y$ in the a.s. partition (41) 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 MRP by construction. To see this, note that if $v, v' \in \widetilde{\mathcal{V}}_y$ for some y, then by (41), $v, v' \in \mathcal{V}_{y_l}(p_l)$ for all $l = 1, \ldots, L$, at least up to collections of v, v' that have Lebesgue measure zero. Recalling (9) and the notation of Definition MRP, this implies 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

$$y \equiv (Y(v, p_1), \ldots, Y(v, p_L)) = (Y(v', p_1), \ldots, Y(v', p_L)), \tag{42}$$

[48] 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 $\mathcal{V}_j(p)$.

yields an L–tuple $y \in \mathcal{Y}$ such that $v, v' \in \mathcal{V}_{y_l}(p_l)$ for every 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 implementation. 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 Definition MRP.

The algorithm works as follows.[49] 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 quasilinear choice model in the sense that $y^t \in \overline{\mathcal{Y}}_t$ if and only if the set

$$\left\{v \in \mathbb{R}^J : v_{y_l^t} - p_{y_l^t} \geq v_j - p_j \text{ for all } j \in \mathcal{J} \text{ and } p \in \mathcal{P}_t\right\} \tag{43}$$

is non-empty. In practice, we do this by sequentially checking the feasibility of a linear program with (43) as the constraint set. The sense in which we do this sequentially is that instead of checking (43) 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 (43) 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 (43) 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 (43) 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

[49] 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.

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 (43) 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 $\overline{\phi}(f)$ be defined as in (20), which we reproduce here for convenience:

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

Note that $\overline{\phi}(f) \in \Phi$, 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 $\overline{\theta}(\overline{\phi}(f)) = \theta(f) = t$, so that (23) is satisfied. To see that $\overline{\phi}(f)$ satisfies (24), observe that

$$\sum_{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))} \overline{\phi}(f)(\mathcal{V}|m,x) \equiv \sum_{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))} \int_{\mathcal{V}} f(v|m,x)\,dv = s_j(m,x;f) = s_j(m,x),$$

where the first equality follows by definition (20), the second follows from (10) and (18), and the third follows from the definition of $\mathcal{F}^\star$. Similarly, $\overline{\phi}(f)$ satisfies (25) because

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

where the third equality follows by Tonelli's Theorem (e.g. Shorack, 2000, pg. 82), the fourth uses (6), which holds (by definition) for all $f \in \mathcal{F}^\star$, and the final equality

reverses the steps of the first three. That $\overline{\phi}(f)$ also satisfies (26) follows using a similar argument since $f \in \mathcal{F}^\star$ satisfies (7), i.e.

$$\begin{aligned}\sum_{\mathcal{V}\in\mathbb{V}^\bullet(w)} \overline{\phi}(f)_{\mathbb{V}|WZ}(\mathcal{V}|w,z) &= \sum_{\mathcal{V}\in\mathbb{V}^\bullet(w)} \int_{\mathcal{V}} \mathbb{E}\left[f(v|M_i,X_i)|W_i=w,Z_i=z\right] dv \\ &= \int_{\cup\{\mathcal{V}:\mathcal{V}\in\mathbb{V}^\bullet(w)\}} f_{V|WZ}(v|w,z)\,dv \\ &\geq \int_{\mathcal{V}^\bullet(w)} f_{V|WZ}(v|w,z)\,dv = 1. \end{aligned} \tag{44}$$

The inequality in (44) follows because the definition of $\mathbb{V}^\bullet(w)$, together with the fact that $\mathbb{V}$ is an a.s. partition of $\mathbb{R}^J$, implies that $\mathcal{V}^\bullet(w)$ is contained in the union of sets in $\mathbb{V}^\bullet(w)$. This inequality implies that $\overline{\phi}(f)$ satisfies (26), because

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

as a result of $\overline{\phi}(f)$ being an element of Φ. We have now established that if $t \in \Theta^\star$, then there exists a $\phi \in \Phi$ satisfying (23)–(26) 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

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

noting that the definition of $\mathbb{V}^\bullet(w)$ ensures that the summands are well-defined. The function $\overline{f}(\phi)(\cdot|m,x)$ places total mass of $\phi(\mathcal{V}|m,x)$ on sets $\mathcal{V} \in \mathbb{V}^\bullet(\omega(m,x))$, and distributes this mass uniformly across each set. We will show that $t \in \Theta^\star$ by establishing that $\overline{f}(\phi) \in \mathcal{F}^\star$ and $\theta(\overline{f}(\phi)) = t$.

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

$$\begin{aligned}\int_{\mathcal{V}} \overline{f}(\phi)(v|m,x)\,dv &\equiv \sum_{\mathcal{V}'\in\mathbb{V}^\bullet(\omega(m,x))} \int_{\mathcal{V}} \frac{\mathbb{1}\left[v \in \mathcal{V}' \cap \mathcal{V}^\bullet(\omega(m,x))\right]}{\lambda\left(\mathcal{V}' \cap \mathcal{V}^\bullet(\omega(m,x))\right)} \phi(\mathcal{V}'|m,x)\,dv \\ &= \mathbb{1}[\mathcal{V} \in \mathbb{V}^\bullet(\omega(m,x))]\phi(\mathcal{V}|m,x),\end{aligned} \tag{45}$$

since the sets in $\mathbb{V}$ and thus $\mathbb{V}^\bullet(\omega(m,x))$ are disjoint (almost surely). Using (45), we

have that

$$\int_{\mathbb{R}^J} \overline{f}(\phi)(v|m,x)\,dv = \sum_{\mathcal{V}\in\mathbb{V}} \int_{\mathcal{V}} \overline{f}(\phi)(v|m,x)\,dv = \sum_{\mathcal{V}\in\mathbb{V}^\bullet(\omega(m,x))} \phi(\mathcal{V}|m,x) = 1, \qquad (46)$$

where the first equality uses the fact that $\mathbb{V}$ is an (a.s.) partition of $\mathbb{R}^J$, and the final equality is implied by the hypothesis that ϕ satisfies (26), since

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

and every $\phi \in \Phi$ satisfies

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

Thus, from (46), and since $\overline{f}(\phi)$ inherits non-negativity from $\phi \in \Phi$, we conclude that $\overline{f}(\phi)$ is a conditional density, i.e. $\overline{f}(\phi) \in \mathcal{F}$.

To see that $\overline{f}(\phi)$ satisfies (6), notice that

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

where the fourth equality uses (25), and the final equality reverses the steps of the first three. The satisfaction of the support condition, (7), follows in a similar way from (26) and Tonelli's Theorem, since

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

That $\overline{f}(\phi)$ satisfies (11) follows from (10), (18), (45), and (24) via

$$\begin{aligned} s_j(m,x;\overline{f}(\phi)) &\equiv \sum_{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))} \int_{\mathcal{V}} \overline{f}(\phi)(v|m,x)\,dv \\ &= \sum_{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))} \mathbb{1}[\mathcal{V}\in\mathbb{V}^{\bullet}(\omega(m,x))]\phi(\mathcal{V}|m,x) \\ &= \sum_{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))} \phi(\mathcal{V}|m,x) - \sum_{\mathcal{V}\in\mathbb{V}_j(\pi(m,x))} \mathbb{1}[\mathcal{V}\notin\mathbb{V}^{\bullet}(\omega(m,x))]\phi(\mathcal{V}|m,x) \\ &= s_j(m,x), \end{aligned}$$

for all $j\in\mathcal{J}$ and $(p,x)\in\text{supp}(P_i,X_i)$. The last equality here uses the implication of (46) that $\phi(\mathcal{V}|m,x)=0$ for any $\mathcal{V}\notin\mathbb{V}^{\bullet}(\omega(m,x))$.

Finally, note that in the notation of (20), (45) says

$$(\overline{\phi}\circ\overline{f}(\phi))(\mathcal{V}|m,x) = \mathbb{1}[\mathcal{V}\in\mathbb{V}^{\bullet}(\omega(m,x))]\phi(\mathcal{V}|m,x).$$

This equality implies that $(\overline{\phi}\circ\overline{f}(\phi))(\mathcal{V}|m,x)=\phi(\mathcal{V}|m,x)$ for all $\mathcal{V}$, since (46) implies that $\phi(\mathcal{V}|m,x)=0$ for $\mathcal{V}\notin\mathbb{V}^{\bullet}(\omega(m,x))$. Thus,

$$\theta(\overline{f}(\phi)) = \overline{\theta}(\overline{\phi}\circ\overline{f}(\phi)) = \overline{\theta}(\phi) = t,$$

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 (24)–(26) are linear equalities, the subset of Φ that satisfies them is also compact and connected. Thus, if $\overline{\theta}$ is continuous on this subset, 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 $t^{\star}_{\text{lb}}$ and $t^{\star}_{\text{ub}}$. *Q.E.D.*

F Implementing Bounds on Consumer Surplus

In this section, we show how sharp bounds on changes in average consumer surplus can be found using Propositions 1 and 2 by constructing appropriate concentrated target parameter functions, $\overline{\theta}$. The function used for the upper bound is different from that used to find the lower bound. Both functions are linear in ϕ.

For shorthand, we denote average consumer surplus at premium $p^{\star}$, conditional on

$(M_i, X_i) = (m, x)$ under valuation density f as

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

Suppose that $\mathbb{V}$ is the MRP constructed from a set of premiums $\mathcal{P}$ that contains the two premiums, p and $p^\star$, at which average consumer surplus is to be contrasted. Then

$$\mathrm{CS}^{p^\star}(m, x; f) = \sum_{\mathcal{V} \in \mathbb{V}} \int_{\mathcal{V}} \left\{ \max_{j \in \mathcal{J}} v_j - p_j^\star \right\} f(v|m, x)\, dv, \tag{47}$$

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 MRP, $\arg\max_{j \in \mathcal{J}} v_j - p_j \equiv Y(v, p) = Y(v', p) \equiv Y(\mathcal{V}, p)$ for all $v, v' \in \mathcal{V}$ and any $p \in \mathcal{P}$. Consequently, we can write (47) as

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

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

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

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

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

These quantities can be computed in an initial step with linear programming. Since we do not restrict the distribution of valuations within each MRP set, the sharp lower bound on a change in average consumer surplus is attained when this distribution

concentrates all of its mass on $v_{\text{lb}}^{p\to p^\star}(\mathcal{V})$ in every $\mathcal{V}\in\mathbb{V}$. That is,

$$\begin{aligned}\Delta\text{CS}^{p\to p^\star}(m,x;f) &\geq \sum_{\mathcal{V}\in\mathbb{V}} p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)} + v_{\text{lb}}^{p\to p^\star}(\mathcal{V})\int_{\mathcal{V}} f(v|m,x)\,dv \\ &= \sum_{\mathcal{V}\in\mathbb{V}} p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)} + v_{\text{lb}}^{p\to p^\star}(\mathcal{V})\left[\overline{\phi}(f)(\mathcal{V}|m,x)\right] \\ &\equiv \Delta\text{CS}_{\text{lb}}^{p\to p^\star}(m,x;f). \end{aligned} \tag{48}$$

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

$$\Delta\text{CS}_{\text{ub}}^{p\to p^\star}(m,x;f) \equiv \sum_{\mathcal{V}\in\mathbb{V}} p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)} + v_{\text{ub}}^{p\to p^\star}(\mathcal{V})\left[\overline{\phi}(f)(\mathcal{V}|m,x)\right].$$

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

$$\overline{\theta}(\phi) \equiv \sum_{\mathcal{V}\in\mathbb{V}} p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)} + v_{\text{lb}}^{p\to p^\star}(\mathcal{V})\phi(\mathcal{V}|m,x), \tag{49}$$

and applying Propositions 1 or 2. The requirement that $\theta(f)=\overline{\theta}(\overline{\phi}(f))$ can be seen to be satisfied here by comparing (48) and (49). The sharp upper bound is found analogously.

G 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, that is, 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 construct conditional choice probability (market shares) by estimating the number of potential buyers and dividing the quantity purchased of the inside choices ($j \geq 1$) by this estimate. This gives us estimated choice shares for the inside choices; the estimated choice share for the outside choice ($j = 0$) is just the difference between 1 and the sum of the estimated inside shares.

The key step here is estimating the number of potential buyers (market size) for

[50] This is common in discrete choice contexts, see e.g. Berry (1994, pg. 247).

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.

Our estimator is constructed using a flexible linear regression. The outcome variable is the indicator I_i. The main regressors are the X_i bins, that is, age in years and income in FPL (taken at the lower endpoint of the bin). We include a full set of interactions between these variables and indicators for the coarse age and income bins described in Section 4.2 (called W_i there). We also include a full set of market indicators (M_i), and interactions between these indicators and both age and income. This regression yields estimated potential buyer probabilities for each (m, x) pair. We convert these probabilities into an estimate of the total number of buyers in each (m, x) pair by using the individual sampling weights provided in the ACS.

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 individuals 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.

H Statistical Uncertainty

One concern in interpreting our estimated bounds is that they may be estimated with statistical uncertainty due to noise in the estimated choice shares. In this section, we examine the extent to which this might be the case through a simulation exercise. For each consumer, we redraw their plan choice from a multinomial distribution with probabilities given by the estimated choice shares in their fine bin. We use these new choices to form new choice shares, and then we run these new choice shares through

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

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

Table 6: *Simulated Distributions of Bounds and Point Estimates*

	Change in probability of purchasing coverage if all per-person premiums increase by $10/month		Change in consumer surplus ($/person-month) if per-person subsidies decrease by $10/month		Change in government spending ($/person-month) if per-person subsidies decrease by $10/month	
Nonparametric bounds	LB	UB	LB	UB	LB	UB
5th percentile	-0.0705	-0.0176	-2.4231	-1.9539	-19.2117	-7.8223
95th percentile	-0.0700	-0.0171	-2.4094	-1.9419	-19.0483	-7.6602
Simple logit	Point estimate		Point estimate		Point estimate	
5th percentile	-0.0317		-2.6418		-9.5548	
95th percentile	-0.0315		-2.6339		-9.5140	
Mixed logit random constant	Point estimate		Point estimate		Point estimate	
5th percentile	-0.0306		-2.7216		-9.3041	
95th percentile	-0.0302		-2.7121		-9.1439	
Mixed logit random price coefficient	Point estimate		Point estimate		Point estimate	
5th percentile	-0.0340		-1.9859		-11.0390	
95th percentile	-0.0305		-1.6172		-10.1852	
Mixed logit random constant & price coefficient	Point estimate		Point estimate		Point estimate	
5th percentile	-0.0264		-2.2268		-9.0732	
95th percentile	-0.0219		-1.7602		-7.9383	
Probit	Point estimate		Point estimate		Point estimate	
5th percentile	-0.0288		-2.7176		-13.7916	
95th percentile	-0.0286		-2.5284		-13.7642	

the same estimators that we used for the actual data, obtaining a new set of bounds (for our procedure) and point estimates (for the parametric models in Section 5). We repeat this procedure 100 times and then look at the distribution of the simulated bounds across these 100 replications.

Table 6 reports the 5th and 95th percentile of both the upper and lower bounds for our primary target parameters under a $10 increase in all premiums. The distribution suggests that neither our bounds nor the point estimates would be very different if the data were realized again under the same distribution. While reassuring, we emphasize that this is a simulation exercise that redraws from an estimated distribution; these are

not confidence regions. Unfortunately, constructing uniformly valid confidence regions for estimators defined by large-scale linear programs remains both theoretically and computationally challenging, especially in problems the size of ours.[53] However, the results of the simulation do suggest that our sample size is large enough such that valid confidence regions both for our model and the parametric models would be quite tight.

[53] Of course, we can construct confidence intervals for the parametric models, but to make a fair comparison we use the same exercise.

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*. 42, 43

Armstrong, T. B. (2016): "Large Market Asymptotics for Differentiated Product Demand Estimators With Economic Models of Supply," *Econometrica*, 84, 1961–1980. 8

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. 44

——— (1997): "Bounds on Treatment Effects From Studies With Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176. 44

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models With Convex Moment Predictions," *Econometrica*, 79, 1785–1821. 44

Berry, S. and P. Haile (2015): "Identification in Differentiated Products Markets," *Annu. Rev. Econ.*, –. 7, 9

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890. 1, 7, 8, 43, 46

——— (2004a): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112, 68–105. 8, 43

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. 8

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25, 242–262. 1, 8, 42, 55

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, 8, 47

——— (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797. 2, 3, 8, 9, 13, 43, 46, 47

Bertsimas, D. and J. N. Tsitsiklis (1997): *Introduction to linear optimization*, vol. 6, Athena Scientific Belmont, MA. 22

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. 42

——— (2010): "Nonparametric Discrete Choice Models With Unobserved Heterogeneity," *Journal of Business & Economic Statistics*, 28, 291–307. 42

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. 43

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

Chesher, A. and A. Rosen (2014): "An Instrumental Variable Random Coefficients Model for Binary Outcomes," *Econometrics Journal*. 44

Chesher, A. and A. M. Rosen (2013): "What Do Instrumental Variable Models Deliver with Discrete Dependent Variables?" *American Economic Review*, 103, 557–62. 44

——— (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85, 959–989. 43, 44

Chesher, A., A. M. Rosen, and K. Smolinski (2013): "An instrumental variable model of multiple discrete choice," *Quantitative Economics*, 4, 157–196. 3, 13, 43, 44

Chiappori, P.-A. and I. Komunjer (2009): "On the Nonparametric Identification of Multiple Choice Models," *Working paper*. 2, 43

Chiong, K. X., Y.-W. Hsieh, and M. Shum (2017): "Counterfactual Estimation in Semiparametric Multinomial Choice Models," *SSRN Electronic Journal*. 42, 43

Compiani, G. (2019): "Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach," *Working paper*. 3, 43

Congressional Budget Office (2017): "Federal Subsidies Under the Affordable Care Act for Health Insurance Coverage," https://www.cbo.gov/sites/default/files/recurringdata/51298-2017-01-healthinsurance.pdf, accessed: January 25, 2019. 1

Conley, T. G., C. B. Hansen, and P. E. Rossi (2010): "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260–272. 35

Demuynck, T. (2015): "Bounding average treatment effects: A linear programming approach," *Economics Letters*, 137, 75–77. 44

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, 35

Finkelstein, A., N. Hendren, and M. Shepard (2019): "Subsidizing Health Insurance for Low-Income Adults: Evidence from Massachusetts," *American Economic Review*, forthcoming. 3, 24, 34

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. 35

Fourer, R., D. M. Gay, and B. W. Kernighan (2002): *AMPL: A Modeling Language for Mathematical Programming*, Cengage Learning. 17

Fox, J. T. (2007): "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *The RAND Journal of Economics*, 38, 1002–1019. 42

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. 11, 43

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 nonparametric instrumental variables estimation," *Journal of Econometrics*, 189, 41–53. 44

Galichon, A. and M. Henry (2011): "Set Identification in Models with Multiple Equilibria," *The Review of Economic Studies*, 78, 1264–1298. 44

Gandhi, A., Z. Lu, and X. Shi (2017): "Estimating Demand for Differentiated Products with Zeroes in Market Share Data," *Working paper*. 43

Garthwaite, C., T. Gross, and M. J. Notowidigdo (2018): "Hospitals as Insurers of Last Resort," *American Economic Journal: Applied Economics*, 10, 1–39. 35

Geruso, M. (2017): "Demand Heterogeneity in Insurance Markets: Implications for Equity and Efficiency," *Quantitative Economics*, 8, 929–975. 35

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," . 17

Handel, B., I. Hendel, and M. D. Whinston (2015): "Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk," *Econometrica*, 83, 1261–1313. 6, 44

Handel, B. R. (2013): "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103, 2643–82. 6, 44

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. 44

Hausman, Leonard, and Zona (1994): "Competitive Analysis with Differenciated Products," *Annales d'Économie et de Statistique*, 159. 42

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. 11

Heckman, J. J. and S. Navarro (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136, 341–396. 11

Ho, K. and A. Pakes (2014): "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, 104, 3841–84. 3, 42, 43

Ho, K. and A. M. Rosen (2017): "Partial Identification in Applied Research: Benefits and Challenges," in *Advances in Economics and Econometrics*, ed. by B. Honore, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge University Press, 307–359. 2

Honoré, B. E. and A. Lleras-Muney (2006): "Bounds in Competing Risks Models and the War on Cancer," *Econometrica*, 74, 1675–1698. 44

Honoré, B. E. and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629. 44

IBM (2010): *IBM ILOG AMPL Version 12.2*, International Business Machines Corporation. 17

Kamat, V. (2017): "Identification with Latent Choice Sets: The Case of the Head Start Impact Study," *Working paper*. 44

Khan, S., F. Ouyang, and E. Tamer (2019): "Inference on Semiparametric Multinomial Response Models," *Working paper*. 42

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. 44

Lafférs, L. (2013): "A note on bounding average treatment effects," *Economics Letters*, 120, 424–428. 44

Lewbel, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177. 11, 42, 43

Mahoney, N. (2015): "Bankruptcy as Implicit Health Insurance," *American Economic Review*, 105, 710–46. 35

Manski, C. F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3, 205–228. 42

——— (1985): "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27, 313–333. 11

——— (2007): "Partial Indentification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393–1410. 42, 43, 44

——— (2014): "Identification of income-leisure preferences and evaluation of income tax policy," *Quantitative Economics*, 5, 145–174. 42, 44

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*. 35

Matzkin, R. L. (1991): "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica*, 59, 1315–1327. 42

——— (1993): "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 58, 137–168. 2, 42

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

Miller, T. P. (2017): "Examining the Effectiveness of the Individual Mandate under the Affordable Care Act," American Enterprise Institute, statement before the House Committee on Ways and Means Subcommittee on Oversight. 6

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters," *Econometrica*, 86, 1589–1619. 22, 44

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342. 7, 8

——— (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. 35, 43

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. 42, 43

Pakes, A. and J. Porter (2016): "Moment Inequalities for Multinomial Choice with Fixed Effects," Tech. rep. 42, 43

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006): "Moment Inequalities and Their Application," *Working paper*. 42, 43

——— (2015): "Moment Inequalities and Their Application," *Econometrica*, 83, 315–334. 42, 43

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, 56

Saltzman, E. (2019): "Demand for health insurance: Evidence from the California and Washington ACA exchanges," *Journal of Health Economics*, 63, 197 – 222. 1

Santos, A. (2012): "Inference in Nonparametric Instrumental Variables With Partial Identification," *Econometrica*, 80, 213–275. 43

Schennach, S. M. (2014): "Entropic Latent Variable Integration via Simulation," *Econometrica*, 82, 345–385. 44

Shi, X., M. Shum, and W. Song (2016): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *SSRN Electronic Journal*. 42

Shorack, G. R. (2000): *Probability for Statisticians*, Springer-Verlag. 50

Tebaldi, P. (2017): "Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA," *Working paper*. 1, 3, 24, 35

Thompson, T. S. (1989): "Identification of Semiparametric Discrete Choice Models," *Discussion paper 249, Center for Economics Research, Department of Economics, University of Minnesota*. 2, 11, 13, 42

Torgovitsky, A. (2016): "Nonparametric Inference on State Dependence with Applications to Employment Dynamics," *SSRN Working Paper 2564305*. 44

——— (2018): "Partial Identification by Extending Subdistributions," *Quantitative Economics*, Forthcoming. 42, 44