NONPARAMETRIC ESTIMATES OF DEMAND IN THE CALIFORNIA HEALTH INSURANCE EXCHANGE

Pietro Tebaldi
Alexander Torgovitsky
Hanbin Yang

WORKING PAPER 25827

NBER WORKING PAPER SERIES

NONPARAMETRIC ESTIMATES OF DEMAND IN THE CALIFORNIA HEALTH INSURANCE EXCHANGE

Pietro Tebaldi
Alexander Torgovitsky
Hanbin Yang

Working Paper 25827
http://www.nber.org/papers/w25827

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2019

We thank Nikhil Agarwal, Stéphane Bonhomme, Steve Berry, Øystein Daljord, Michael Dinerstein, Liran Einav, Phil Haile, Kate Ho, Ali Hortaçsu, Simon Lee, Chuck Manski, Sanjog Misra, Magne Mogstad, Adam Rosen, Ariel Pakes, Áureo de Paula, Jack Porter, Bernard Salanié, Thomas Wollman, and participants in many conferences and seminars for helpful comments and feedback. Five anonymous referees and the co-editor provided comments that led to substantial improvements in the paper. Research supported in part by the Becker Friedman Institute Health Economics Initiative and by National Science Foundation grant SES-1426882. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w25827.ack

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Pietro Tebaldi, Alexander Torgovitsky, and Hanbin Yang. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Nonparametric Estimates of Demand in the California Health Insurance Exchange
Pietro Tebaldi, Alexander Torgovitsky, and Hanbin Yang
NBER Working Paper No. 25827
May 2019
JEL No. C14,C3,C5,I13

ABSTRACT

We develop a new nonparametric discrete choice methodology and use it to analyze the demand for health insurance in the California Affordable Care Act marketplace. The methodology allows for endogenous prices and instrumental variables, while avoiding parametric functional form assumptions about the unobserved components of utility. We use the methodology to estimate bounds on the effects of changing premiums or subsidies on coverage choices, consumer surplus, and government spending on subsidies. We find that a $10 decrease in monthly premium subsidies would cause a decline between 1.8% and 6.7% in the proportion of subsidized adults with coverage. The reduction in total annual consumer surplus would be between $63 and $74 million, while the savings in yearly subsidy outlays would be between $209 and $601 million. We estimate the demand impacts of linking subsidies to age, finding that shifting subsidies from older to younger buyers would increase average consumer surplus, with potentially large impacts on enrollment. We also estimate the consumer surplus impact of removing the highly-subsidized plans in the Silver metal tier, where we find that a nonparametric model is consistent with a wide range of possibilities. We find that comparable mixed logit models tend to yield price sensitivity estimates towards the lower end of the nonparametric bounds, while producing consumer surplus impacts that can be both higher and lower than the nonparametric bounds depending on the specification of random coefficients.

Pietro Tebaldi
Department of Economics
Columbia University
IAB, MC 3308
New York, NY 10027
and NBER
pt2571@columbia.edu

Alexander Torgovitsky
University of Chicago
1126 E. 59th Street
Chicago, IL, 60637
atorgovitsky@gmail.com

Hanbin Yang
Harvard University
hanbin.v.yang@gmail.com

An online appendix is available at http://www.nber.org/data-appendix/w25827

1 Introduction

1.1 Motivation and Overview

Under the Patient Protection and Affordable Care Act of 2010 ("ACA"), the United States federal government spends over $40 billion annually on subsidizing health insurance premiums (Congressional Budget Office, 2017). The design of the ACA, and more generally the regulation of non-group, publicly-sponsored, privately-provided health insurance remain objects of intense debate among policy makers across the developed world (e.g. Einav and Levin, 2015). Addressing key design issues, such as the structure of premium subsidies or the types of plans to offer, requires estimating demand.

Recent research has filled this need using parametric discrete choice models. 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 marketplace, Tebaldi (2017) estimated demand in the California ACA marketplace with a variety of logit, nested logit, and mixed (random coefficient) logit models, and Polyakova and Ryan (2019) used mixed logit models to estimate demand in states with federally-facilitated marketplaces.

The different logit models differ in how they deal with the independence of irrelevant alternatives (e.g. Goldberg, 1995; McFadden and Train, 2000), and in how they deal with the potential endogeneity of prices (e.g. Berry, 1994; Berry, Levinsohn, and Pakes, 1995). However, they are all fully parametric, with the logistic distribution playing a key role. The richer mixed logit models also typically rely on normal distributions for additional unobserved heterogeneity. The concerning possibility is that counterfactual predictions generated by these models are significantly driven by these functional forms.

In Section 2 of this paper we develop a new nonparametric methodology for discrete choice. We analyze a distribution-free counterpart of a standard discrete choice model that allows for endogenous prices. We show how to compute, estimate, and conduct statistical inference on sharp identified sets for policy-relevant target parameters without imposing parametric assumptions. In Sections 3–5, we then apply the methodology at scale to estimate the demand impacts of various regulatory counterfactuals in the California ACA marketplace. We contrast our findings with logit, mixed logit, and probit models, which we find sometimes produce quite different predictions, especially in regards to consumer surplus.

1.2 Methodological Contribution

The discrete choice model we analyze in Section 2 is

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

where Y_i is individual i's choice from the discrete set of options $j \in \mathcal{J} = \{0, 1, \ldots, J\}$ with prices P_{ij} and latent valuations V_{ij}. We do not require valuations to follow a specific distribution, such as normal (probit) or type I extreme value (logit), and we allow them to be arbitrarily dependent across options for each individual. We allow for prices and valuations to be correlated, and use instrumental variables to address this endogeneity. The main restriction embedded in (1) is that prices and valuations are additively separable, which can be viewed as requiring quasilinearity.

Nonparametric point identification arguments for discrete choice models often require a large amount of variation in prices (e.g. Manski, 1975; Thompson, 1989; Matzkin, 1993; Fox and Gandhi, 2016). Yet variation in prices is limited in many applications, including ours. Nonparametric point identification arguments with endogenous prices and instrumental variables also typically use an additional "completeness" condition (Chiappori and Komunjer, 2009; Berry and Haile, 2010, 2014; Compiani, 2019), which in some cases is also necessary for point identification (Newey and Powell, 2003; Santos, 2012).[1] Completeness can be difficult to interpret, and has been shown to be untestable (Canay, Santos, and Shaikh, 2013). Not assuming either large variation or completeness leads naturally to partial identification (Tamer, 2010; Ho and Rosen, 2017; Molinari, 2020).

The primary challenge with partial identification is computing sharp bounds, which can be difficult in complicated models. We develop a strategy that builds on Manski (2007), who analyzed discrete choice under only the basic assumptions of rationality. Manski's insight was that individuals facing a discrete choice can be characterized by a discrete set of latent types based on what their choices would be under any of a finite number of potential choice scenarios, such as different prices. The data provides some information on the relative frequency of the latent types, while assumptions such as rationality eliminate other types altogether. Manski showed how to use linear programming to combine these sources of information and produce bounds on choice probabilities.

The number of types in Manski's framework grows extremely quickly with the

[1] Berry and Haile (2014, Section 5) also provide nonparametric identification results that do not require completeness and instead leverage both demand and supply under a shape restriction on the distribution of unobservables; see also Berry and Haile (2018).

number of choice sets. This is because rationality is a weak assumption: the number of rational preference configurations compatible with even a small number of price vectors is astronomically large even when the number of choices (J) is small. We eliminate many of these types by assuming not only that individuals are rational, but also that preferences are consistent with (1). We show that under (1) each type can be interpreted as a subset of valuations. These subsets divide the space of valuations into a partition that we call the minimal relevant partition (MRP).

Using the MRP, we then follow Manski in developing a linear programming approach for computing sharp bounds on various target parameters. Our approach has several features that Manski's does not. Our bounds incorporate the identifying content of quasilinearity in (1). We show how to use instruments when prices are endogenous, as well as how to further restrict the number of types to reflect known vertical orderings. In addition to bounding choice probabilities, we also show how to bound elasticities, as well as changes in consumer surplus resulting from price changes or the removal of a choice option. We use a constructive argument to prove that our bounds remain sharp even after adding all of these features. Then, we implement our method at scale, making use of recent advances in estimation and inference under partial identification.

Our paper is also related to Chesher, Rosen, and Smolinski (2013). Those authors demonstrate that endogenous prices lead to partial identification in a class of discrete choice models that includes (1), even if one maintains the usual logistic parametric distribution (Chesher et al., 2013, Section 4.2). Using random set theory, Chesher et al. (2013) show that whether a candidate value of the model parameters belongs to the sharp identified set can be determined by checking a system of moment inequalities. The system has a finite number of moment inequalities when the number of prices is finite, as in our setting. Each moment inequality corresponds to a "core-determining set" of valuations, which we show can be a much larger collection of sets than the MRP (see Section 2.7.1).

The Chesher et al. (2013) identification analysis requires checking whether the moment inequalities hold *for every candidate value of the model parameters*. If the model is nonparametric, as in our setting, then one component of the model parameters is the distribution of valuations. Implementing the Chesher et al. (2013) analysis would require checking whether *every* distribution of valuations satisfies a set of moment inequalities. This is computationally intractable: iterating over *every* distribution—even approximately, using a grid—is simply not possible because the space of distributions for continuous random variables is infinite-dimensional. Chesher et al. (2013, pg. 160) also derive other nonparametric outer (non-sharp) bounds that are straightforward to implement, but these do not exploit the assumption that choices were generated ac-

cording to (1). In Appendix S5, we replicate a simulation in Chesher et al. (2013), demonstrating that our sharp nonparametric bounds lie strictly between their non-sharp nonparametric bounds and their sharp parametric bounds.

Our methodology is computationally tractable because it shifts focus towards specific scalar target parameters, such as choice probabilities, elasticities, and changes in consumer surplus. Instead of trying to iterate over the space of distributions, we check whether *there exists* a distribution that is consistent with the data and reproduces a candidate value of the target parameter. We prove that one can determine whether such a distribution exists by solving a finite system of linear equations, thereby providing the basis of our linear programming procedure. The strategy makes our nonparametric methodology feasible to implement at scale, as showcased by our application.

There are a number of papers on semi- and nonparametric discrete choice that parameterize the valuations in (1) using a linear index of choice and/or individual characteristics, and then focus on identifying and estimating these index parameters, while treating the distribution of unobservables as a nuisance parameter. Examples include Manski (1975), Matzkin (1993), Lewbel (2000), Fox (2007), Pakes (2010), Ho and Pakes (2014), Pakes, Porter, Ho, and Ishii (2006, 2015), Pakes and Porter (2016), Shi, Shum, and Song (2016), and Khan, Ouyang, and Tamer (2019). In contrast, we do not have index parameters in our model, and we do not treat the distribution of unobservables as a nuisance parameter because it determines the counterfactuals we want to infer. Other semi- and nonparametric papers that focus on counterfactuals but do not allow for endogeneity and instruments include Thompson (1989), Manski (2007), Briesch, Chintagunta, and Matzkin (2010), Chiong, Hsieh, and Shum (2017), Allen and Rehbeck (2019), and Fosgerau and Kristensen (2020). Manski (2014), Kline and Tartari (2016), and Kamat (2020) apply approaches similar to Manski (2007) with choice models different from (1).

In our application we observe individual choices and have price variation across individuals within each market. Our analysis of (1) is tailored to this setting, which is different in important ways from settings with a single price per market, as considered by Berry et al. (1995), Nevo (2001), and many subsequent authors. The methodology we develop in the main text is not immediately useful for this case, although in Appendix S2 we suggest ways in which our analysis could be extended to it.

1.3 Empirical Results

We use our methodology to estimate demand counterfactuals in the California ACA marketplace (Covered California). Our estimates are based on administrative records

on all purchases through Covered California in 2014. We focus on the choice of coverage level (metal tier) for low-income individuals who are not covered under employer-sponsored insurance or public programs. In addition to the nonparametric model, we also report estimates from comparable parametric logit, probit, and mixed logit models.

Our main empirical strategy leverages institutionally-induced variation in post-subsidy premiums across age and income, similar to Polyakova and Ryan (2019). We cast this as an instrumental variable strategy by dividing individuals into relatively homogenous groups by age and income, then using residual variation in age and income within that group to instrument for prices (post-subsidy premiums) while assuming that latent variations remain stable. For all of our results we also conduct a sensitivity analysis that relaxes or drops either invariance to age or invariance to income (or both). The sensitivity analysis provides a transparent connection between the price variation in the data and the estimated results.

We estimate target parameters related to choice shares and welfare for several different counterfactuals. The main counterfactual we consider—which is simple but particularly policy-relevant—is a uniform increase in all post-subsidy premiums, which from the consumer's perspective is the same as a decrease in premium subsidies. We also estimate substitution patterns by considering counterfactuals in which premiums are increased for different metal tiers separately. The other counterfactuals we consider concern regulatory design. In one, we consider the impacts of shifting premium subsidies towards younger or older consumers. In another, we evaluate the consumer surplus impact of removing Silver plans.

One theme that emerges throughout our results is that parametric models tend to produce price sensitivity estimates that are attenuated compared to the nonparametric bounds. For example, we estimate that a uniform \$120 increase in yearly premiums for all plans would cause between a 1.8% and 6.7% decline in the proportion of low-income adults who purchase insurance, off of a base of 28%. The parametric models, by contrast, produce point estimates that are all between 2.5% and 3.4%, suggesting that they could be understating price sensitivity. Other logit-based estimates for Covered California have been similar, such as Tebaldi (2017) who estimates between a 1.5% and 3.5% decline in the proportion enrolled in response to a \$100 increase in yearly premiums, Saltzman (2019), who estimates a 3.3% decline, and Saltzman (2021), who estimates a 1.2% decline. The findings are consistent with the idea that the logit model has a "flat" functional form. Ho and Pakes (2014) and Compiani (2019) reached similar conclusions, albeit using different methods in different empirical settings.

The differences between the nonparametric estimates and the comparable parametric models become more nuanced when considering consumer surplus. The nonpara-

metric estimate of the consumer surplus impact of decreasing subsidies by \$120 annually is between \$63 and \$74 million annually when aggregated, with poorer individuals incurring the bulk of the loss. Total annual savings on premium subsidies would be between \$209 and \$601 million per year. The various mixed logit models—which differ only by which coefficients are random—produce consumer surplus estimates between \$61 and \$83 million, with estimates of subsidy savings towards the lower end of our bounds. For a shift in subsidies towards younger buyers, the parametric models predict an average consumer surplus impact of less than \$1 per person, per month, while the nonparametric bounds include impacts that could be almost three times as large. The parametric models predict that removing Silver plans would create an aggregate consumer surplus impact of anywhere between \$162 to \$290 million per year, whereas the nonparametric model produces a lower bound of \$13 million, while recognizing that there is no logical upper bound. The results are consistent with long-standing concerns about the use of logit models for welfare analysis, especially when the counterfactual involves adding or removing a product (e.g. Hausman, 1996; Petrin, 2002; Ackerberg and Rysman, 2005; Berry and Pakes, 2007)

Our empirical analysis is related to a large literature on the demand for subsidized health insurance, including Chan and Gruber (2010), Krueger and Kuziemko (2013), Ericson and Starc (2015), Hackmann, Kolstad, and Kowalski (2015), Shepard (2016), Tebaldi (2017), DeLeire, Chappel, Finegold, and Gee (2017), Finkelstein, Hendren, and Shepard (2019), Saltzman (2019), Drake (2019), Polyakova and Ryan (2019), and Jaffe and Shepard (2020). The estimates on alternative subsidy schemes is related to the design of premium subsidies, as studied by Decarolis (2015), Einav, Finkelstein, and Tebaldi (2019), Polyakova and Ryan (2019), and Decarolis, Polyakova, and Ryan (2020). The effects of removing certain coverage options has been studied in different contexts by Dafny, Ho, and Varela (2013) and Marone and Sabety (2019). We contribute to the literature by providing nonparametric estimates that do not depend on functional form choices: our estimates depend only on the structure of the choice model (1) together with instrumental variable assumptions and limited vertical orderings. These assumptions have straightforward economic interpretations.

The primary drawback of our empirical analysis is that we do not model supply, so all of our estimates must be interpreted as holding insurers' decisions fixed. This is a clear limitation, since one might expect potentially strong supply-side responses in the counterfactuals we consider. Integrating our nonparametric methodology for estimating demand with a model of supply is an interesting avenue for further research.

2 Methodology

2.1 Nonparametric Discrete Choice Model

Individual i makes a single choice Y_i from a set $\mathcal{J} \equiv \{0, 1, \ldots, J\}$ of $J+1$ choices. Each choice j has a potentially endogenous characteristic called price, P_{ij}, which we model as discretely distributed, and collect into the vector $P_i \equiv (P_{i0}, P_{i1}, \ldots, P_{iJ})$. Choice $j = 0$ represents the outside option of not choosing any of the inside choices $j \geq 1$, and has its price normalized to $P_{i0} = 0$. When we apply the model to Covered California, we will have five choices ($J = 4$) with options $1, 2, 3$, and 4 representing Bronze, Silver, Gold, and Platinum plans.

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

The main economic restriction in (1) is the additive separability between valuations and prices, which imposes restrictions on substitution patterns consistent with quasilinear utility. If all prices were to increase by the same amount, then an individual who chooses $j \geq 1$ before the increase will either continue to choose j after the increase, or will switch to the outside option ($j = 0$), but they will not switch to a different inside choice $k \geq 1$, $k \neq j$. This limits the role of income effects to the extensive margin of choosing any inside choice versus taking the outside option. In Appendix S1, we derive (1) from an insurance choice model similar to the ones in Handel (2013) and Handel, Hendel, and Whinston (2015), in which individuals have quasilinear utility and constant absolute risk aversion. Variation in V_i across individuals arises from heterogeneity in unobserved preferences, risk factors, and risk aversion.

The limited role of income effects in (1) pertain to conjectured variations for a single individual i. When we take (1) to the data we combine observations on many individuals, so in practice we can allow for income effects by allowing for dependence between an individual's income and their valuations. We formalize this by treating an individual's income and other observed characteristics as part of a vector, X_i, and then restricting the dependence between V_i and the various components of X_i. Throughout

the paper, we model X_i as discretely distributed with finite support.

One observable characteristic that is particularly important is the individual's market. When we estimate demand we will do so conditional on a market, so that market-level characteristics—both observable and unobservables—are held fixed in the counterfactual (see e.g. Berry and Haile, 2010, pg. 5). To emphasize this we let M_i denote individual i's market, and we treat M_i as separate from X_i in the notation.

Our empirical 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). These settings would have P_i constant conditional on M_i. The methodology we develop in the main text is not immediately useful for this case. In Appendix S2 we propose two ways in which our methodology could be extended to handle more aggregated price variation.

2.2 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 individuals, 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. See, for example, equation (6) of Nevo (2011), or equation (1) of Berry and Haile (2015). 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 (or log-normally) distributed. Our motivation for considering (1) is to preserve the utility maximization structure in (2), while avoiding these types of parametric assumptions.

The three indices in (2) reflect different possible levels of data aggregation. If only market-level data is available, as in Berry et al. (1995) or Nevo (2001), then (2) is aggregated to the (j, m) level, and the data is viewed as drawn from a population of markets and/or products (Berry, Linton, and Pakes, 2004b; Armstrong, 2016). Our analysis presumes richer individual-level choice data as in Berry, Levinsohn, and Pakes (2004a) or Berry and Haile (2010), but the number of markets we study is small and fixed. To emphasize this, we index the nonparametric model (1) only over i and j, and we record individual 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 $V_{ij} \equiv \alpha_i^{-1}(X_{ij}'\beta_i + \xi_{ij} + \epsilon_{ij})$.[2]

2.3 The Density of Valuations

The primitive object in model (1) is the distribution of valuations, V_i, conditional on prices P_i, market, M_i, and other covariates, X_i. We will assume throughout the paper that this distribution is continuous so that ties between choices in (1) occur with zero probability. More formally, we assume that the distribution of V_i is absolutely continuously distributed with respect to Lebesgue measure on $\mathbb{R}^J$, conditional on $(P_i, M_i, X_i) = (p, m, x)$ for every (p, m, x) in the support of (P_i, M_i, X_i). In addition to ensuring no ties, absolute continuity means we can associate the conditional distribution of valuations with a conditional density function $f(\cdot|p, m, x)$ for each realization $P_i = p$, $M_i = m$, and $X_i = x$. Let $\mathcal{F}$ denote the set of all such conditional density functions.

2.4 Target Parameters

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 individuals who would choose j at a new price vector, $p^\star$. The proportion can be written in terms of f as

$$\int \underbrace{\mathbb{1}[v_j - p_j^\star \geq v_k - p_k^\star \text{ for all } k]}_{\text{choose } j \text{ if prices were } p^\star} f(v|m, x)\, dv,$$
$$\text{where} \quad f(v|m, x) \equiv \mathbb{E}\left[f(v|P_i, m, x)|M_i = m, X_i = x\right] \tag{3}$$

is the density of valuations conditional on market, m and other characteristics x, averaged over the remaining variation in observed prices (if any). Another natural counterfactual quantity is the impact on average consumer surplus caused by changing prices from p to $p^\star$. The impact 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.

[2] This requires the mild assumption that $\alpha_i > 0$. As suggested by a referee, an advantage of normalizing in this way is that V_{ij} can be directly interpreted as willingness-to-pay relative to the outside option (see e.g. Lewbel, 2019, Section 6.3).

We view (3) and (4) as examples of *target parameters*; that is, functions $\theta : \mathcal{F} \to \mathbb{R}^{d_\theta}$ that map the collection $\mathcal{F}$ of all conditional density functions on $\mathbb{R}^J$ into real vectors. The goal is to infer the values of $\theta(f)$ that are consistent with both the observed data and the assumptions.

2.5 Assumptions

We augment the choice model (1) with two types of assumptions.

2.5.1 Instrumental Variables

Let W_i and Z_i be two subvectors (or more general functions) of prices, market and covariates, (P_i, M_i, X_i). Define the density of valuations conditional on W_i and Z_i as

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

Assumption IV. *$f(v|w,z) = f(v|w,z')$ for all z, z', w and v.*

Assumption IV says that the distribution of valuations is invariant to shifts in Z_i, conditional on W_i. That is, Z_i are exogenous instruments and W_i are control variables. Assumption IV nests exogenous prices as a special case by taking $Z_i = P_i$ and W_i to be deterministic (e.g. $W_i = 1$). It also nests the unconditional instrumental variable assumption used by Chesher et al. (2013, Restriction A.6) by taking W_i to be deterministic.

For Assumption IV to be useful, prices should shift with the instruments Z_i while conditioning on the controls W_i. This follows the usual intuition: if Z_i is exogenous, then changes in observed choice shares as Z_i varies reflect changes in prices, rather than changes in valuations. The more that prices 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. Our methodology does not require the instruments to have any particular amount of variation, but greater variation produces more informative bounds.

In the application we also use the following generalization of Assumption IV, which allows the instruments to be "imperfect" in the terminology of Nevo and Rosen (2012).

Assumption IV. ***(Generalized)*** *For every z, z', w and v,*

$$(1 - \kappa(z,z',w))f(v|w,z') \leq f(v|w,z) \leq (1 + \kappa(z,z',w))f(v|w,z'),$$

where $\kappa(z,z',w) \geq 0$ is chosen by the researcher.

The generalization allows for a pointwise difference in the density of valuations between z and z' of up to $(100 \times \kappa(z, z', w))\%$. Similar approaches to relaxing assumptions involving exact equalities have also been used in different contexts by Conley, Hansen, and Rossi (2010), Manski and Pepper (2017), Torgovitsky (2019), and Rambachan and Roth (2020). Taking $\kappa(z, z', w) = 0$ returns the generalized assumption back to the baseline case of an exogenous instrument.

2.5.2 Support

The second assumption is that the support of f is contained in a known subset of $\mathbb{R}^J$.

Assumption SP. $\int_{\mathcal{V}^\bullet(w)} f(v|w, z)\, dv = 1$ *for each* w *and* z*, where* $\mathcal{V}^\bullet(w)$ *is a known subset of* $\mathbb{R}^J$*.*

In the application we use Assumption SP to exploit the fact that certain plans in the ACA are unambiguously more attractive at equal prices than other plans, at least for some individuals.

2.6 The Identified Set

Suppose that we know the observed choice shares

$$s_j(p, m, x) \equiv \mathbb{P}[Y_i = j | P_i = p, M_i = m, X_i = x]. \tag{6}$$

Each density of valuations $f \in \mathcal{F}$ also implies a set of choice shares

$$s_j(p, m, x; f) \equiv \int_{\mathcal{V}_j(p)} f(v|p, m, x)\, dv, \tag{7}$$

$$\text{where} \quad \mathcal{V}_j(p) \equiv \left\{(v_1, \ldots, v_J) \in \mathbb{R}^J : v_j - p_j \geq v_k - p_k \ \text{ for all } k\right\}. \tag{8}$$

A density f is consistent with the observed choice shares ("matches the data") if

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

The sharp identified set of valuation densities is defined as the subset of $f \in \mathcal{F}$ that both match the data and satisfy Assumptions IV and SP. We call this subset $\mathcal{F}^\star$:

$$\mathcal{F}^\star \equiv \{f \in \mathcal{F} : \ f \text{ satisfies Assumptions IV, SP, and equation (MD)}\}. \tag{9}$$

The object of interest is the sharp identified set for the target parameter, which is defined as the image of $\mathcal{F}^\star$ under θ, and denoted as

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

2.7 Identification Analysis

In this section we show how to compute $\Theta^\star$. If $\mathcal{F}$ were a parametric class of densities, then the results of Chesher et al. (2013) could be used to compute $\Theta^\star$ (approximately) by iterating over a grid of $f \in \mathcal{F}$ that satisfy Assumptions IV and SP, checking whether each such f satisfies a finite set of moment inequalities, and adding f to $\mathcal{F}^\star$ and $\theta(f)$ to $\Theta^\star$ if and only if they are satisfied. In our model, f is nonparametric, so $\mathcal{F}$ is an infinite-dimensional set. Iterating over even a grid of $f \in \mathcal{F}$ is not feasible in this case.

We develop a different strategy. Instead of trying to check a set of conditions for every $f \in \mathcal{F}$, we check whether there exists *some* $f \in \mathcal{F}^\star$ that would generate a given candidate value $t = \theta(f)$ for the target parameter. As we prove ahead in Proposition 1, the existence of such an f is equivalent to the existence of a solution to a carefully-constructed finite-dimensional system of linear equations. The finite-dimensional system is constructed by first partitioning the space of valuations into the smallest number of sets needed to describe choice behavior under all relevant prices. We call this partition the minimal relevant partition (MRP).

2.7.1 The Minimal Relevant Partition of Valuations

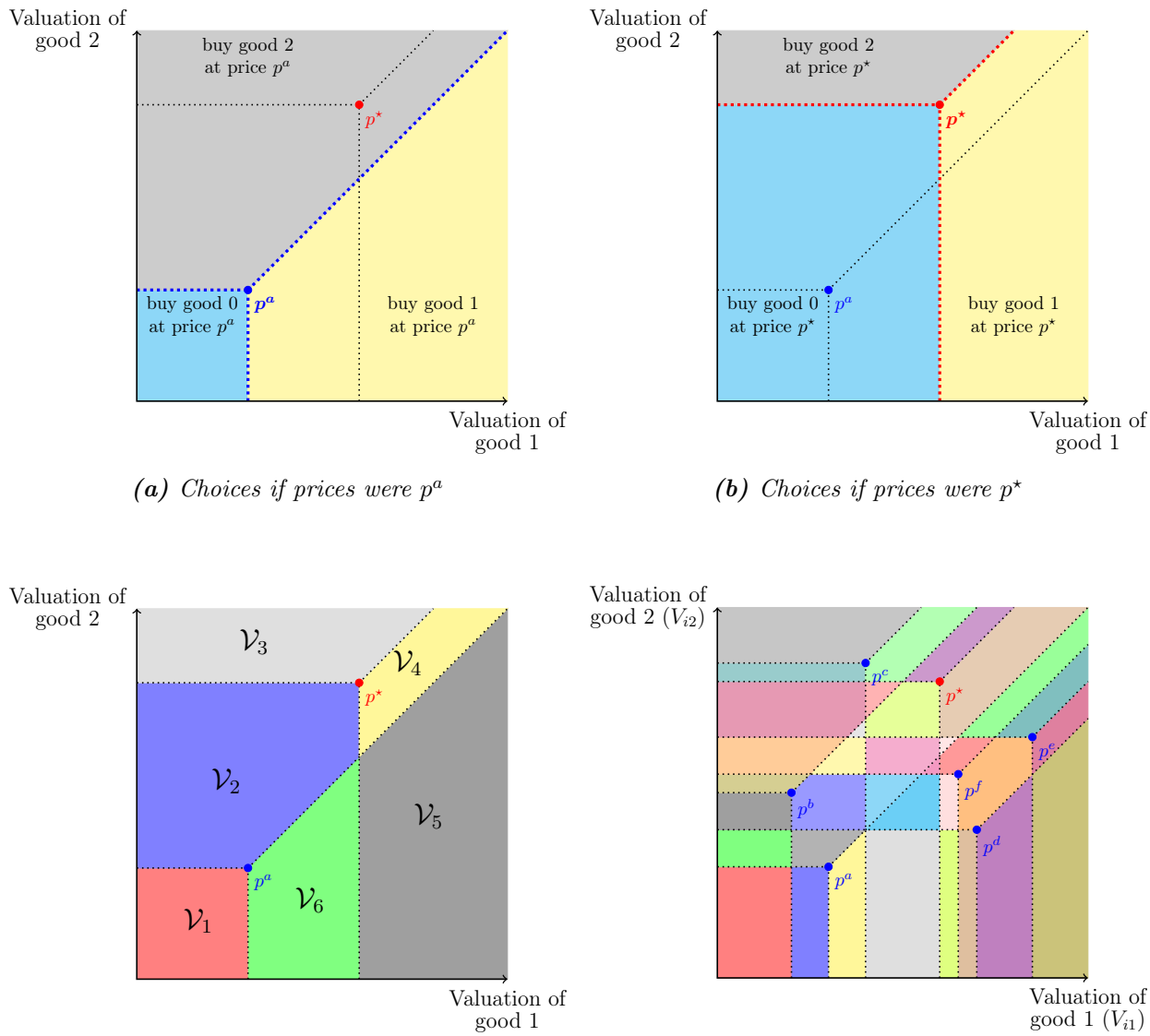
Let $\mathcal{P}$ denote a finite set of prices that contains the observed prices, as well as any additional prices relevant for evaluating the target parameter, θ. The MRP is the smallest (coarsest) partition of valuation space ($\mathbb{R}^J$) with the following property: any two valuations in the same subset produce the same choice behavior for all $p \in \mathcal{P}$, while any two valuations in different subsets produce different choice behavior for at least one $p \in \mathcal{P}$. The formal definition is 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):*

$$v, v' \in \mathcal{V} \text{ for some } \mathcal{V} \in \mathbb{V} \quad \Leftrightarrow \quad Y(v,p) = Y(v',p) \text{ for all } p \in \mathcal{P}. \tag{10}$$

Suppose that the data consists of a single observed price vector, p^a, and that we are concerned with behavior under a counterfactual price vector, $p^\star$. Take $\mathcal{P} = \{p^a, p^\star\}$.

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 seven prices*

Figure 1 illustrates the MRP in this case with $J = 2$. Panel (a) shows that considering choice behavior under price p^a divides $\mathbb{R}^2$ into three sets depending on whether an individual would choose options 0, 1, or 2 when faced with p^a. Panel (b) shows the analogous division under price $p^\star$. See e.g Thompson (1989, Figure 1), Chesher et al. (2013, Figure 1), or Berry and Haile (2014, Figure 1) for similar diagrams. Intersecting these two three-set collections creates the collection of six sets shown in panel (c). The collection of six sets is the MRP generated by $\mathcal{P} = \{p^a, p^\star\}$. Figure 1d shows the MRP constructed from a set $\mathcal{P}$ with seven prices. We describe an algorithm for constructing the MRP in Appendix S3.

The MRP is the coarsest division of $\mathbb{R}^J$ that captures all choice behavior relevant for prices in $\mathcal{P}$. The six sets in Figure 1c correspond to the six types of individuals that could exist under indirect utility specification (1), where a type is defined as a pair of choices under $(p^a, p^\star)$. The characterization by types follows Manski (2007), but here we adapt it to the quasilinear choice model (1).

The MRP is related to the class of core-determining sets (CDS) derived by Chesher et al. (2013), but there are fewer sets in the MRP than the CDS. With two prices and $J = 2$, the CDS consists of 12 sets (Chesher et al., 2013, Figures 2–3) compared to 6 sets in the MRP. With seven prices and $J = 2$, the CDS consists of 842 sets (Chesher, Rosen, and Smolinski, 2011, Table 1, pg. 28) compared to 35 sets in the MRP (Figure 1d). The MRPs in our application have between 195,000 and 900,000 sets, depending on the region and target parameter.

2.7.2 Equivalent Finite-Dimensional Problem

The MRP is formed by organizing the continuum of valuations in $\mathbb{R}^J$ into sets that yield identical choice behavior at all prices in $\mathcal{P}$. If our concern is behavior that occurs (or would occur) under these prices, then we can shift our attention from densities in $\mathbb{R}^J$ to mass functions over the sets in the MRP. For example, in Figure 1c the share of individuals who would choose 1 if prices were p^a can be written as

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

which depends not on f per se, but only on the mass that f places on $\mathcal{V}_5$ and $\mathcal{V}_6$. Focusing on these masses replaces the intractable infinite-dimensional problem of searching over all $f \in \mathcal{F}$ into a tractable finite-dimensional problem of searching over a finite set of non-negative numbers that sum to unity. As we now show, the replacement can be done while preserving all of the information provided by the choice model, the data, and Assumptions IV and SP.

We denote the set of conditional mass functions supported on $\mathbb{V}$ as

$$\Phi \equiv \left\{ \phi \in \mathbb{R}_+^{d_\phi} : \sum_{\mathcal{V} \in \mathbb{V}} \phi(\mathcal{V}|p, m, x) = 1 \quad \text{for all } (p, m, x) \in \text{supp}(P_i, M_i, X_i) \right\}, \tag{11}$$

where d_ϕ is the cardinality of $\mathbb{V} \times \text{supp}(P_i, M_i, X_i)$. Let $\mathbb{V}_j(p) \equiv \{\mathcal{V} \in \mathbb{V} : Y(v, p) = j \text{ for almost every } v \in \mathcal{V}\}$ denote the sets in the MRP associated with choosing j under price $p \in \mathcal{P}$. Then $\phi \in \Phi$ is consistent with an f that satisfies (MD) only if

$$\sum_{\mathcal{V} \in \mathbb{V}_j(p)} \phi(\mathcal{V}|p, m, x) = s_j(p, m, x) \quad \text{for all } j \in \mathcal{J} \text{ and } (p, m, x). \tag{MD'}$$

Each ϕ generates the conditional-on-(W_i, Z_i) probability mass function

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

To be consistent with an f that satisfies Assumption IV, ϕ should satisfy

$$(1 - \kappa(z, z', w))\phi(\mathcal{V}|w, z') \leq \phi(\mathcal{V}|w, z) \leq (1 + \kappa(z, z', w))\phi(\mathcal{V}|w, z') \quad \text{for all } z, z', w, \text{ and } \mathcal{V}. \tag{IV'}$$

To be consistent with an f that satisfies Assumption SP, ϕ should satisfy

$$\sum_{\mathcal{V} \in \mathbb{V}^\bullet(w)} \phi(\mathcal{V}|w, z) = 1 \quad \text{for all } w \text{ and } z, \tag{SP'}$$

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

Since our focus is not on f but on the target parameter, θ, we assume that it is possible to represent θ in terms of ϕ.

Condition TP. *For any $f \in \mathcal{F}$, define $\overline{\phi}(f) \in \Phi$ as*

$$\overline{\phi}(f)(\mathcal{V}|p, m, x) \equiv \int_{\mathcal{V}} f(v|p, m, x)\, dv \quad \textit{for all } \mathcal{V} \in \mathbb{V}. \tag{12}$$

Then there exists a known function $\overline{\theta} : \Phi \to \mathbb{R}^{d_\theta}$ such that

$$\theta(f) = \overline{\theta}(\overline{\phi}(f)) \quad \textit{for every } f \in \mathcal{F}.$$

Since Φ depends on the MRP, and the MRP depends on $\mathcal{P}$, Condition TP can always be satisfied by choosing $\mathcal{P}$ to include all prices relevant for evaluating the target parameter,

θ. For example, if the target parameter is demand at the new premium vector $p^\star$ in Figure 1c, then Condition TP simply requires choosing $\mathcal{P}$ to include $p^\star$.

Under Condition TP, $\Theta^\star$ can be characterized in terms of the mass function ϕ without having to consider the full density f. If $t \in \Theta^\star$, then (by definition) there exists an $f \in \mathcal{F}^\star$ with $\theta(f) = t$. Since $f \in \mathcal{F}^\star$ satisfies (MD), Assumption IV, and Assumption SP (again all by definition), taking the mass that any such f places on the MRP yields a $\phi \in \Phi$ that satisfies (MD'), (IV'), (SP'), and $\overline{\theta}(\phi) = t$; for example, take $\overline{\phi}(f)$ defined in (12). The opposite direction is more delicate: if $\overline{\theta}(\phi) = t$ for some $\phi \in \Phi$ that satisfies (MD'), (IV'), and (SP'), then is $t \in \Theta^\star$? The following proposition shows that the answer is yes. The idea is to use such a ϕ to construct an $f \in \mathcal{F}^\star$ that yields $\theta(f) = t$. The proof is in Appendix S4.

Proposition 1. *Suppose that Condition TP is satisfied. Let*

$$\Phi^\star \equiv \{\phi \in \Phi : \ \phi \textit{ satisfies } \text{(MD')}, \text{(IV')}, \text{(SP')}\}, \tag{13}$$

and for any $t \in \mathbb{R}^{d_\theta}$, let $\Phi^\star(t) \equiv \Phi^\star \cap \{\phi \in \Phi : \overline{\theta}(\phi) = t\}$. Then $t \in \Theta^\star$ if and only if $\Phi^\star(t)$ is non-empty.

Proposition 1 shows that $\Theta^\star$ can be exactly characterized by solving a finite-dimensional system of equations. If the target parameter is scalar, then $\Theta^\star$ can be computed more directly by solving two optimization problems.

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*

$$t^\star_{lb} \equiv \min_{\phi \in \Phi^\star} \ \overline{\theta}(\phi) \quad \textit{and} \quad t^\star_{ub} \equiv \max_{\phi \in \Phi^\star} \ \overline{\theta}(\phi). \tag{14}$$

Each condition that defines $\Phi^\star$—(MD'), (IV'), and (SP')—is linear in ϕ. If $\overline{\theta}$ is also linear in ϕ, then (14) is a linear program. A change in choice shares, (3), yields a $\overline{\theta}$ function that is linear in ϕ. A change in consumer surplus, (4), also does, although constructing $\overline{\theta}$ is less obvious (see Appendix S6). Elasticities produce $\overline{\theta}$ functions that are nonlinear, but the resulting optimization problem is linear-fractional, so can be reformulated as a linear program using the Charnes and Cooper (1962) transformation; see Appendix S7 for more detail, and Kamat (2020) for a similar observation in an instrumental variables model. Linearity ensures that both the minimization and maximization problems in (14) can be solved reliably and relatively quickly, even when the dimension of Φ is quite large.[3]

[3] All of the linear programs ahead were computed using Gurobi (Gurobi Optimization, 2015), and we

2.8 Estimation and Statistical Inference

The choice shares $s_j(p, m, x)$ are in practice replaced by estimates $\hat{s}_j(p, m, x)$. Using the estimated choice shares in (MD') can lead to empty feasible sets in (14), even when the model is correctly specified so that the feasible sets are non-empty with the true choice shares. We solve this problem by applying an estimator of $[t^\star_{\text{lb}}, t^\star_{\text{ub}}]$ developed by Mogstad, Santos, and Torgovitsky (2018). The estimator proceeds in two steps.

In the first step, we find the best fit to the estimated choice shares by solving

$$\widehat{Q}^\star \equiv \min_{\phi \in \Phi} \widehat{Q}(\phi) \qquad \text{subject to (IV') and (SP')}, \tag{15}$$

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

and where $\widehat{\mathbb{P}}[P_i = p, M_i = m, X_i = x]$ are estimated probabilities. Defining $\widehat{Q}$ with absolute deviations means that (15) can be reformulated as a linear program by replacing terms in absolute values by the sum of their positive and negative parts (e.g. Bertsimas and Tsitsiklis, 1997, pp. 19–20). The absolute deviations are weighted so that $\widehat{Q}(\phi)$ is an estimate of the average absolute deviation in choice shares under ϕ.

In the second step, we collect values of $\overline{\theta}(\phi)$ for ϕ that come close to minimizing (15). That is, we construct the set:

$$\widehat{\Theta}^\star \equiv \left\{ \overline{\theta}(\phi) : \phi \in \Phi,\ \phi \text{ satisfies (IV'), (SP'), and } \widehat{Q}(\phi) \leq (1+\eta)\widehat{Q}^\star, \right\}. \tag{16}$$

The qualifier "close" here reflects the tuning parameter η, which is a small positive constant that must converge to zero at an appropriate rate with the sample size. The purpose of η is to smooth out potential discontinuities that could conceivably arise in the convergence of a set estimator. In our empirical estimates we set $\eta = 10^{-4}$ and found 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. Notice that the way $\widehat{Q}^\star$ is defined ensures that $\widehat{\Theta}^\star$ is never empty.

We compute $\widehat{\Theta}^\star$ by solving

$$\widehat{t}^\star_{\text{lb}} \equiv \min_{\phi \in \Phi} \overline{\theta}(\phi) \quad \text{subject to (IV'), (SP'), and } \widehat{Q}(\phi) \leq \widehat{Q}^\star(1+\eta), \tag{17}$$

and an analogous maximization problem for $\widehat{t}^\star_{\text{ub}}$. The set estimator for $\Theta^\star$ is $\widehat{\Theta}^\star \equiv [\widehat{t}^\star_{\text{lb}}, \widehat{t}^\star_{\text{ub}}]$. When $\overline{\theta}$ is linear, (17) can be reformulated as a linear program, again by

checked a subset of the results using CPLEX (IBM, 2010).

appropriately reformulating the absolute value terms in $\widehat{Q}(\phi)$. The overall estimation procedure requires solving three linear programs: one for (15), one for (17), and one for the analogous maximization problem.

We conduct statistical inference using the test developed by Deb, Kitamura, Quah, and Stoye (2021); see also Kitamura and Stoye (2018). The specifics of the test are notationally involved, so here we provide a high-level overview of what it entails. Appendix S8 contains a detailed discussion of how we implement the test in our application.

The null hypothesis of the test is $H_0 : t \in \Theta^\star$, where t is a candidate value of the target parameter. The test statistic is the minimum weighted sum of squares between the estimated choice shares and the shares predicted by a $\phi \in \Phi$ that satisfies (IV'), (SP'), and $\overline{\theta}(\phi) = t$. The test statistic is thus the optimal value of a quadratic program. A critical value is found by bootstrapping the choice shares and, for each bootstrap replication, solving a tightened version of the same quadratic program. The tightening effectively provides the type of generalized moment selection used in the moment inequality literature (e.g. Andrews and Soares, 2010). The null hypothesis is rejected at the 5% level if the test statistic exceeds the .95 bootstrapped quantile of the tightened problem.

We construct a 95% confidence interval by inverting 5% tests. That is, the confidence interval contains all target parameter candidate values t for which the null hypothesis $H_0 : t \in \Theta^\star$ is not rejected. The confidence interval can be constructed relatively efficiently by bisecting its endpoints. In our application, the overall procedure of building confidence intervals is still computationally demanding, since it requires solving an extremely large quadratic program for each bootstrap replication and each trial value of t. For this reason, and because our sample size is quite large, we report confidence intervals for only a few parameters.

3 Covered California

3.1 Institutional Details

Covered California is one of the largest state-level ACA marketplaces, accounting for more than 10% of national enrollment. The marketplace offers health insurance plans directed at individuals not covered by an employer or by a public program such as Medicaid or Medicare.

The basic structure of Covered California is determined by federal regulation 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 rating regions. Insurers are allowed to vary premiums across (but not

Table 1: *Standardized plan characteristics in Covered California*

Panel (a): Characteristics by metal tier before cost-sharing reductions (CSRs)							
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 (CSRs)							
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

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, then undergo state certification. Consumers purchase insurance for the subsequent year during an open enrollment period at the end of the year.

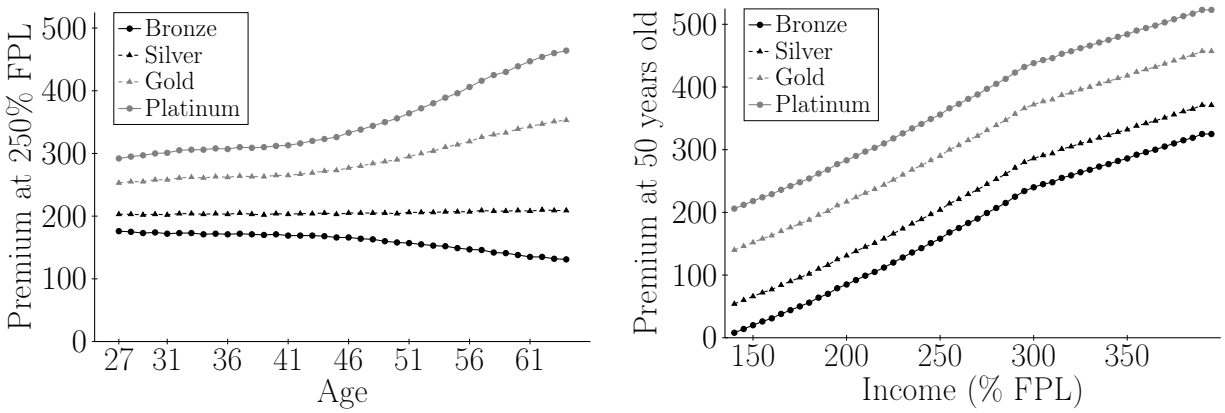
Covered California differs from other ACA marketplaces in important ways. An insurer who intends to participate in a rating region is required to offer a menu of four plans classified into metal tiers of increasing actuarial value: Bronze, Silver, Gold, and Platinum.[4] Unlike other marketplaces, where insurers do not need to offer Bronze or Platinum plans, in California an insurer must provide the entire menu of four plans in any region it enters. Moreover, the actuarial features of the plans are required to have the standardized characteristics shown in Table 1 (among others not shown).

Insurers are also regulated on how they can set premiums. Each insurer chooses a base premium for each metal tier in each rating region. The base premium is multiplied by a federally-determined adjustment factor that increases with a consumer's age from 1 at age 21 to 3 at age 64 (see Orsini and Tebaldi, 2017, for further detail). The insurer is not permitted to adjust premiums based on any non-age characteristics of the consumer. Insurer premiums are thus a deterministic function of a consumer's age and resident rating region.

Individuals with household income below 400% of the Federal Poverty Level (FPL)

[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.

Figure 2: *Post-subsidy premium variation by age and income*



receive premium subsidies. The size of the premium subsidy is set so that the subsidized premium of the second-cheapest Silver plan is lower than a so-called "maximum affordable amount" that varies with income. Post-subsidy premiums are thus a deterministic function of a consumer's age, resident rating region, and household income.

Figure 2 illustrates how post-subsidy premiums vary as a function of both age and income. Holding age fixed, post-subsidy premiums increase with income across all plans equally due to lower subsidies. Holding income fixed, post-subsidy premiums increase differentially for different plans due to the adjustment factor.

In addition to premium subsidies, the ACA also provides cost-sharing reductions (CSRs) for individuals with household income lower than 250% of the FPL. In Covered California, CSRs are implemented by changing the actuarial terms of Silver plans by income, with discrete changes at 150%, 200%, and 250% of the FPL; see Table 1.

The ACA had a universal coverage mandate that specified an income tax penalty for remaining uninsured. We treat the tax penalty as affecting the value of the outside option of not purchasing any Covered California plan. The universal mandate was generally unenforced between 2014–2017 (Miller, 2017).

3.2 Data

Our primary data are administrative records on all individuals who purchased a plan through Covered California in 2014. The records contain unique individual and household identifiers, as well as 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, we can also calculate premiums for plans a consumer did not choose. We focus on adults aged 27–64 with household income between 140% and 400% of the FPL, which is 73% of the roughly 1.29 million enrollees.

The remaining 27% of enrollees are either ineligible for subsidies (11%), or are younger than 26, so considered dependents under the ACA (16%).

We characterize each individual i by their resident rating region (market), M_i, and a vector X_i 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, running from 140% to 400%. When crossed with the 19 rating regions in Covered California, we obtain 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. To transform quantities purchased into shares, we estimate the number of potential buyers using data from the 2011–2013 American Community Survey public use file (via IPUMS, Ruggles et al., 2015). Our potential buyer estimates use a flexible linear regression, similar to Finkelstein et al. (2019) and Tebaldi (2017). More detail is provided in Appendix S9.

We combine potential buyer estimates with the administrative data to construct choice shares for each of the region $\times$ age $\times$ income bins. To avoid excessive extrapolation, we drop bins that are empty in the ACS sample. We also drop a few small bins for which we estimate fewer potential buyers than there are enrollees in the administrative data. We are left with 30,027 bins that we use as the main estimation sample. 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.

We focus on an individual's choice of coverage level (metal tier), so that $J = 4$, 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 is that the choice of coverage level is separable from the choice of insurer. We view this assumption as reasonable 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 for each tier in each bin as the median post-subsidy premium across insurers. Because of ACA regulations, $P_i = \pi(M_i, X_i)$ is a deterministic function of a consumer's region, M_i, and demographic characteristics, X_i. We reflect this in notation ahead by writing $f(v|p, m, x) = f(v|m, x)$.

Table 2 provides some summary statistics. Each (M_i, X_i) bin contains 85 potential buyers on average. The average participation rate in Covered California is 28%, but participation varies widely across demographics and rating regions. Older and poorer buyers in particular are more likely to purchase coverage. The impact of the CSRs is evident in panel (b): buyers with income below 200% face monthly premiums under \$100 for a Silver plan with actuarial value of 88% or more (see Table 1). Over 30%

Table 2: *Summary statistics*

Panel (a): Distribution of bin characteristics	Mean	St. Dev.	P-10	Median	P-90
Number of buyers	85.27	90.86	14	55	194
Age	43.411	10.696	29	43	59
Income (FPL%)	243.985	72.049	155	230	355
Takeup rate	0.280	0.208	0.053	0.235	0.576
Average premium paid	175.510	89.058	69	163	298
Share choosing Bronze	0.065	0.073	0	0	0
Share choosing Silver	0.188	0.173	0	0	0
Share choosing Gold	0.015	0.021	0	0	0
Share choosing Platinum	0.012	0.018	0	0	0

Panel (b): Premiums and choice shares 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

Notes: *Panel (a) reports statistic taken across the 30,027 bins in our main estimation sample. All statistics are weighted by number of potential buyers. For income, standard deviation means the standard deviation of the within-bin medians of income and average premium paid. In panel (b), premium is the average premium paid for buyers of a given age/income group.*

of such consumers purchase a Silver plan, whereas among consumers with income over 250% of the FPL, fewer than 9% purchase a more expensive and less generous non-CSR Silver plan.

3.3 Identifying Assumptions

Insurers in Covered California choose the base premium for each rating region and coverage level. This choice 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 networks. Using variation across regions risks confounding premiums with these other unobservable differences. For this reason, our primary empirical strategy uses only within-region variation, and we do not impose any restriction on how preferences (the density of

valuations f) vary across regions, M_i.[5]

Instead, we assume—in a limited way—that valuations are locally invariant to age and income. Within a region, post-subsidy premiums vary with age and income due to ACA regulations. The variation is outside of the control of insurers, who only choose base premiums for each region. The main concern with this strategy is that valuations also change with age or income due to changes in latent risk factors. We defend against this concern by only using local variation in age and income and then conducting a sensitivity analysis using the imperfect instrument generalization of Assumption IV.

In the notation of Section 2, we let W_i denote a coarse aggregate of (M_i, X_i). The aggregates are constructed by grouping 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 region indicator M_i crossed with all possibilities of these coarser age-income bins. Conditional on W_i, we still observe variation in premiums due to variation in age and income within the W_i bin. The assumption is that the distribution of latent valuations does not change as X_i varies within this bin. In terms of Assumption IV, this means taking $Z_i = X_i$ as the instrument, while conditioning on W_i.

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 there are 50 values of X_i comprised of the ages 36, 37, 38, 39, and 40 crossed with the 10 income bins between 150% and 200% in steps of 5%. We observe a different premium vector for each of these 50 values, and we assume that observed choices are generated by the same distribution of valuations.

We evaluate sensitivity to this assumption by using the imperfect instrument generalization of Assumption IV. We specify κ as

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

where $\kappa_{\text{age}}, \kappa_{\text{inc}} \geq 0$ are parameters that control how much the density of valuations is allowed to vary. For example, taking $\kappa_{\text{age}} = .2$ and $\kappa_{\text{inc}} = 0$ allows for a 20% difference in valuations for any two adjacent one-year age bins with the same income, while still requiring adjacent income bins with the same age to have identical valuations.

[5] As a consequence, we can estimate bounds region-by-region, which helps with computation. For statistical inference, we need to consider all regions simultaneously, which is the main reason that inference is so much more computationally intensive; see Appendix S8 for more detail.

In Appendix S10, we also implement an alternative empirical strategy that allows for valuations to change arbitrarily across age and instead uses limited cross-region variation in premiums, while still maintaining invariance across income. The demand estimates from that strategy are broadly similar to the main results that use only within-region variation in age and income.

Throughout all of the analysis we use Assumption SP to exploit unambiguous vertical orderings between plans. We specify the support sets $\mathcal{V}^{\bullet}(w)$ as:

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

This specification requires consumers to always have higher valuations for plans that dominate on all actuarial characteristics. The different cases account for the CSRs, which change at 150%, 200% and 250% of the FPL (see Table 1). We do not restrict the ordering of Silver relative to Gold and Platinum for the 150–200% FPL bracket, or relative to Gold for the 200–250% FPL bracket, because the CSRs make these plans differ in ways that may be more or less attractive for different consumers. We also do not assume that any of the plans are valued more than the outside option, that is, we allow valuations for Covered California plans to be negative.

3.4 Parametric Models

For comparison, we also consider some fully parametric models, all of which follow a specification similar to (2):

$$Y_i = \arg\max_{j \in \mathcal{J}} \mathbb{1}[j \neq 0] \left(\gamma_i - \alpha_i P_{ij} + \beta_i \text{AV}_{ij}\right) + \epsilon_{ij}, \tag{18}$$

where γ_i is an individual-specific intercept, AV_{ij} is the actuarial value of tier j for individual i (see Table 1), and α_i and β_i are individual slope coefficients. The indicator normalizes the contribution of these 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 ϵ_{ij} follows a normal distribution. We always estimate (18) region-by-region, so that all parameters vary by region in an unrestricted way.

The first model we estimate is a logit in which the price parameter, α_i, is constant

within regions, but both γ_i and β_i vary with age and income in coarse bins. In particular, the specification allows β_i to vary freely by region 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 region, and within each region 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. The second model is a probit with the same specification.[6]

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 first logit model. The three models differ in whether γ_i ("mixed logit I"), α_i ("mixed logit II"), or both ("mixed logit III") have an additional unobservable component that is normally distributed with unknown variance. In the latter case, we also assume that the two unobservable components are uncorrelated.

We use these five parametric models to contextualize the nonparametric results and demonstrate some of the benefits—and limitations—of the nonparametric methodology. The comparison should be qualified by pointing out that the parametric methodologies also differ in other ways from the nonparametric methodology. The limited verticality we impose on the nonparametric model through Assumption SP cannot be imposed in the parametric models because the distribution of ϵ_{ij} has full support. Using actuarial value (AV_{ij}) in (18) adds a notion of verticality that should make the comparison closer.[7] We also use coarser age and income bins together with some additive separability because including more parameters created convergence problems in some regions. Perhaps most substantively, the parametric models are estimated through maximum likelihood, while our nonparametric bounds are estimated using the absolute deviations approach in Section 2.8. As a consequence of these differences, the parametric estimates need not lie inside the estimated nonparametric bounds.

3.5 Demand Only

We do not model supply, so all of the results should be interpreted as holding supply fixed. This is an important caveat to all of our counterfactual estimates, both parametric and nonparametric. Integrating the nonparametric demand methodology with a supply-side model is an interesting avenue for future research, but beyond the scope of the current paper.

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

[7] We also estimated specifications where AV_{ij} was replaced by product–specific intercepts. These specifications produced less price sensitivity than (18).

4 Demand

We begin by estimating demand responses to uniform changes in monthly, per-person premiums. The premium vectors in these counterfactuals take the form $\pi(M_i, X_i) + \delta$ for various choices of δ. The first class of target parameters we consider are changes in choice shares. For good j, region (market) m, and consumer characteristics x, these can be written as

$$\Delta\text{Share}_j^\delta(m, x; f) \equiv \int_{\mathcal{V}_j(\pi(m,x)+\delta)} f(v|m, x)\, dv - \int_{\mathcal{V}_j(\pi(m,x))} f(v|m, x)\, dv, \tag{19}$$

where $\mathcal{V}_j(p)$ was defined in (8). To aggregate (19) into a single measure, we average it over regions and characteristics:

$$\Delta\text{Share}_j^\delta(f) \equiv \sum_{m,x} \Delta\text{Share}_j^\delta(m, x; f)\, \mathbb{P}[M_i = m, X_i = x]. \tag{20}$$

Table 3 reports estimated bounds for $\Delta\text{Share}_j^\delta$ across the four metal tiers together with bounds on overall participation, $\sum_{j\geq 1} \Delta\text{Share}_j^\delta = 1 - \Delta\text{Share}_0^\delta$. In the first row of each panel of Table 3, δ is set to a \$10 increase for all plans together. From the consumer's perspective, this is the same as a \$10 decrease in premium subsidies or, alternatively, the same as a \$10 increase in the ACA's "maximum affordable amount." In the next four rows of each panel of Table 3, δ is set to a \$10 increase in per-person monthly premiums in each of the four metal tiers alone.

4.1 Participation and Substitution

Estimated impacts on overall participation are shown in the first column of Table 3. A \$10 increase in all premiums reduces the proportion of individuals who purchase coverage by between 1.8% and 6.7%. Panel (b) shows that the bounds are larger in magnitude for lower income individuals, at between 2.0% and 9.1%, and panel (c) shows that they are smaller in magnitude for higher income individuals, at between 1.6% and 3.7%. Comparing panels (b) and (c) more generally, we find a pattern consistent with higher price sensitivity for lower income enrollees, which is consistent with the literature (e.g. Abraham, Drake, Sacks, and Simon, 2017; Finkelstein et al., 2019).

Premium increases for Gold or Platinum plans would have little impact on overall enrollment, which makes sense since there are relatively few buyers for these plans to begin with, and \$10 is a proportionally smaller increase in premiums for these plans. For higher-income consumers, we estimate similar bounds on participation from increasing premiums for either Bronze or Silver. For low-income consumers, increasing

Table 3: *Nonparametric bounds on changes in choice shares*

$10/month premium increase for	Change in probability of choosing: Any plan LB	Any plan UB	Bronze LB	Bronze UB	Silver LB	Silver UB	Gold LB	Gold UB	Platinum LB	Platinum UB
	Panel (a): Full sample (140 - 400% FPL)									
All plans	-0.067	-0.018	-0.012	-0.004	-0.051	-0.011	-0.004	-0.001	-0.003	-0.001
Bronze	-0.011	-0.002	-0.047	-0.009	+0.004	+0.044	+0.000	+0.028	+0.000	+0.023
Silver	-0.050	-0.003	+0.001	+0.124	-0.165	-0.017	+0.001	+0.122	+0.000	+0.096
Gold	-0.003	-0.000	+0.000	+0.005	+0.000	+0.010	-0.013	-0.003	+0.000	+0.011
Platinum	-0.002	-0.000	+0.000	+0.003	+0.000	+0.006	+0.001	+0.009	-0.010	-0.002
	Panel (b): Lower income (140 - 250% FPL)									
All plans	-0.091	-0.020	-0.011	-0.003	-0.077	-0.015	-0.003	-0.001	-0.003	-0.001
Bronze	-0.009	-0.001	-0.046	-0.008	+0.004	+0.043	+0.000	+0.027	+0.000	+0.023
Silver	-0.076	-0.005	+0.001	+0.179	-0.238	-0.021	+0.001	+0.174	+0.000	+0.140
Gold	-0.002	-0.000	+0.000	+0.004	+0.000	+0.009	-0.011	-0.002	+0.000	+0.010
Platinum	-0.002	-0.000	+0.000	+0.004	+0.000	+0.006	+0.001	+0.010	-0.010	-0.002
	Panel (c): Higher income (250 - 400% FPL)									
All plans	-0.037	-0.016	-0.015	-0.006	-0.018	-0.007	-0.004	-0.001	-0.003	-0.001
Bronze	-0.013	-0.003	-0.049	-0.009	+0.003	+0.045	+0.000	+0.029	+0.000	+0.023
Silver	-0.016	-0.001	+0.001	+0.053	-0.072	-0.012	+0.001	+0.055	+0.000	+0.040
Gold	-0.003	-0.000	+0.000	+0.006	+0.000	+0.012	-0.016	-0.004	+0.000	+0.013
Platinum	-0.002	-0.000	+0.000	+0.003	+0.000	+0.005	+0.001	+0.009	-0.010	-0.003

Silver premiums could decrease participation by as much as 7.6%, while an increase in the Bronze premium would cause a decrease of at most 0.9%.

The estimated bounds on substitution patterns within and between coverage tiers are also informative in many cases. For example, panel (a) shows that an increase in Bronze premiums by $10 would lead to a decrease of between 0.9% and 4.7% in the share of consumers choosing Bronze coverage, and an increase in the share choosing Silver of between 0.4% and 4.4%.[8] The upper bound on the increase in the share choosing Gold or Platinum is significantly smaller, reflecting the closer substitutability of the Bronze and Silver plans. The change in participation from a Bronze premium increase is between 0.2% and 1.1%, 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 0.2%.

[8] Figure S2 shows that the joint identified set for the two choice shares is far from rectangular. For example, if the effect on Bronze is a decrease of 4.7%, then the effect on Silver must be an increase of between roughly 2% and 4.4%, narrower than the marginal bounds of 0.4% to 4.4%.

Table 4: *Sensitivity to Assumption IV*

	Change in probability of purchasing coverage if all per-person premiums increase by \$10/month					
	$\kappa_{\text{age}} = \kappa$, $\kappa_{\text{inc}} = 0$		$\kappa_{\text{age}} = 0$, $\kappa_{\text{inc}} = \kappa$		$\kappa_{\text{age}} = \kappa_{\text{inc}} = \kappa$	
κ	LB	UB	LB	UB	LB	UB
0	-0.0674	-0.0184	-0.0674	-0.0184	-0.0674	-0.0184
0.2	-0.0689	-0.0193	-0.1066	-0.0353	-0.0931	-0.0291
0.3	-0.0697	-0.0197	-0.1217	-0.0404	-0.1050	-0.0315
0.4	-0.0704	-0.0200	-0.1346	-0.0445	-0.1170	-0.0329
0.6	-0.0718	-0.0206	-0.1545	-0.0494	-0.1401	-0.0320
$+\infty$	-0.0866	-0.0158	-0.2599	-0.0293	-0.2798	-0.0000

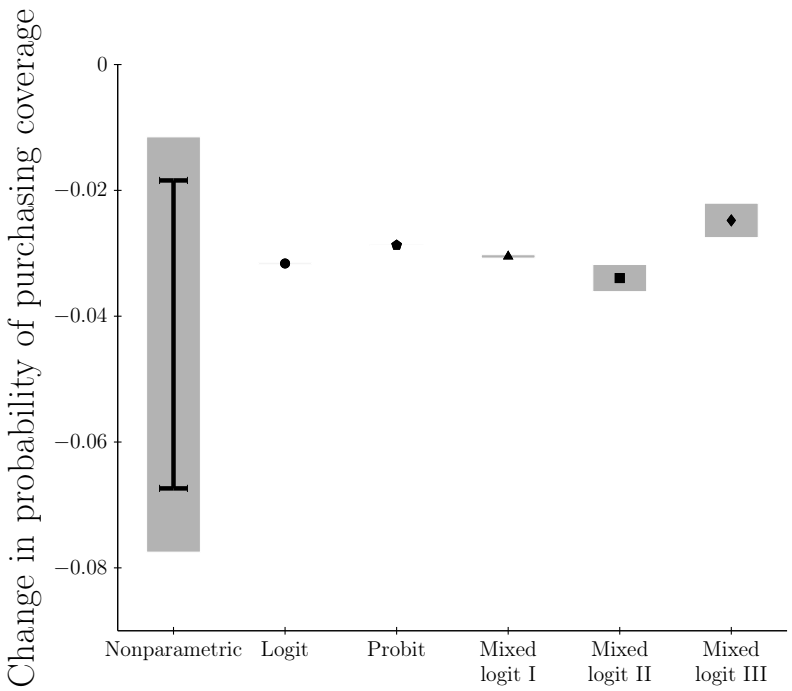
4.2 Sensitivity

Table 4 reports sensitivity to Assumption IV. The reported bounds are for the change in participation due to a \$10 decrease in subsidies (increase in all premiums). The three pairs of columns correspond to allowing for variation across age but not income bins ($\kappa_{\text{age}} > 0$, $\kappa_{\text{inc}} = 0$), allowing for variation across income but not age bins ($\kappa_{\text{age}} = 0$, $\kappa_{\text{inc}} > 0$), and allowing for variation across both income and age bins ($\kappa_{\text{age}} = \kappa_{\text{inc}} > 0$). The top row with $\kappa_{\text{age}} = 0$ and $\kappa_{\text{inc}} = 0$ is the same for each of the three pairs of columns, and the same as reported in Table 3. At the opposite extremes, the bottom row in the first pair of columns uses only premium variation in income, the bottom row in the second pair uses only premium variation in age, and the bottom row in the third pair uses neither. Intermediate values of κ_{age} and κ_{inc} limit the amount by which adjacent bins can differ, with larger values representing weaker identifying assumptions.[9]

The bottom row of the first pair of columns ($\kappa_{\text{age}} = +\infty$ and $\kappa_{\text{inc}} = 0$) shows that if we completely drop age invariance, the estimated bounds are 1.6% to 8.7%, which is only modestly wider than the baseline estimates of 1.8% to 6.7%. The bottom row of the second pair ($\kappa_{\text{age}} = 0$ and $\kappa_{\text{inc}} = +\infty$) has much wider bounds, implying that the estimates depend more on the income invariance assumption. When both income and age invariance are completely removed, the bounds become completely uninformative and we cannot rule out that a \$10 decrease in subsidies causes all 28% of the population currently enrolled to stop participating. Setting $\kappa_{\text{age}} = \kappa_{\text{inc}} = .6$ allows for the density

[9] The bounds tend to be monotonic in κ_{age} and κ_{inc}, but this is not always the case. The population bounds would necessarily be monotonic, since larger values of κ_{age} and κ_{inc} correspond to weaker assumptions. However, this does not need to be the case when estimating the bounds using the procedure discussed in Section 2.8. The reason is that as κ_{age} or κ_{inc} increases, the value of $\widehat{Q}^{\star}$ always mechanically decreases, and thus the set $\widehat{\Theta}^{\star}$ over which the bounds are taken in the second step also changes, potentially excluding densities that fit the observed choice shares sufficiently well for smaller values of κ_{age} or κ_{inc}.

Figure 3: *Change in participation in response to a $10 decrease in subsidies*



Notes: *Bound and point estimates are shown in solid black, and 95% confidence intervals are indicated with grey shading. The confidence interval for the logit and probit models are too narrow to be visible.*

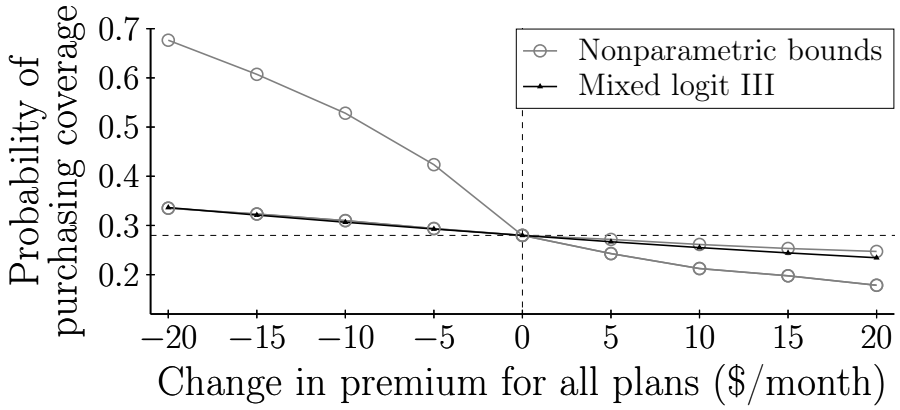
of valuations in adjacent age and/or income bins to increase or decrease by 60%. The bounds do widen considerably, but remain informative. For a more modest relaxation, like $\kappa_{\text{age}} = \kappa_{\text{inc}} = .2$, the bounds are still close to the baseline estimates.

Overall, Table 4 shows that the participation estimates primarily rely on the assumption that consumers in narrow income bins have similar preferences. This makes sense given Figure 2, since post-subsidy premiums vary more with income than with age. Assuming valuations are invariant to income is natural, since it gives the additive separability in (1) the interpretation of quasilinearity (within coarse bins). While we view local invariance of valuations to age as reasonable for the relatively homogenous groups of individuals within each 5 year bin, the assumption is unlikely to hold exactly since risk factors change with age (see e.g. Ericson and Starc, 2015; Geruso, 2017; Orsini and Tebaldi, 2017). However, as Table 4 demonstrates, age invariance plays a secondary role in some of our estimates.

4.3 Comparison to Parametric Models

Figure 3 plots the baseline bounds on change in participation due to a $10 decrease in subsidies (1.8% to 6.7%) together with 95% confidence intervals. For comparison, it

Figure 4: *Change in participation in response to different sized changes in subsidies*



also plots point estimates and confidence intervals for the parametric models discussed in Section 3.4. All of the point estimates lie within the nonparametric bounds, towards the upper bound, where price sensitivity is lowest.

By definition, any value within the nonparametric bounds can be produced by a distribution of valuations that matches the observed choice shares equally well. The parametric models produce a single point estimate by further requiring the distribution of valuations to also have a particular shape. The conclusion we take from Figure 3 is therefore not that the parameterizations used in the parametric models do not matter. Instead, it's that the five parametric models we consider provide similar conclusions, and that there are other distributions of valuations that match the data equally well while leading to much different conclusions.

Figure 4 explores this comparison for subsidy increases and decreases of different sizes. Larger changes reflect more ambitious counterfactuals that are more dissimilar from what was observed in the data. The nonparametric bounds widen to reflect the increasing distance of the counterfactual. They are also wider for premium decreases than for increases, intuitively because we have less information on price sensitivity for the 72% of potential buyers who did not participate under the observed premiums. In contrast, the richest mixed logit model produces a point prediction for both increases and decreases, regardless of how dissimilar the counterfactual is to the observed data.[10] The point prediction hugs the lower bound for premium decreases and the upper bound for premium increases, suggesting that it could be underestimating the degree of price sensitivity, potentially by a considerable amount.

[10] Confidence intervals for the mixed logit predictions in Figure 4 are not much different than those in Figure 3, even for the most distant counterfactuals.

Table 5: *Elasticities*

1% premium increase for			% change in probability of choosing: Outside Bounds	Outside Point estimate	High deductible Bounds	High deductible Point estimate	Low deductible Bounds	Low deductible Point estimate
High deductible	Nonparametric		+0.026	+0.169	-9.809	-1.691	+0.250	+2.703
	$\kappa_{\text{age}} = 0.4$	$\kappa_{\text{inc}} = 0$	+0.024	+0.189	-10.440	-1.873	+0.348	+2.927
	$\kappa_{\text{age}} = \infty$	$\kappa_{\text{inc}} = 0$	+0.000	+0.282	-11.413	-1.041	+0.070	+3.468
	$\kappa_{\text{age}} = 0$	$\kappa_{\text{inc}} = 0.4$	+0.084	+0.379	-10.097	-2.705	+0.149	+2.723
	$\kappa_{\text{age}} = 0$	$\kappa_{\text{inc}} = \infty$	+0.116	+0.903	-10.686	-2.308	+0.073	+2.718
	Logit		+0.145		-2.220		+0.157	
	Probit		+0.136		-2.377		+0.229	
	Mixed Logit I		+0.141		-2.272		+0.188	
	Mixed Logit II		+0.278		-4.854		+0.410	
	Mixed Logit III		+0.222		-5.160		+0.700	
Low deductible	Nonparametric		+0.213	+1.531	+1.350	+54.705	-15.615	-1.978
	$\kappa_{\text{age}} = 0.4$	$\kappa_{\text{inc}} = 0$	+0.209	+1.570	+1.918	+59.257	-16.265	-2.326
	$\kappa_{\text{age}} = \infty$	$\kappa_{\text{inc}} = 0$	+0.052	+1.920	+0.230	+68.381	-18.607	-1.367
	$\kappa_{\text{age}} = 0$	$\kappa_{\text{inc}} = 0.4$	+0.471	+3.065	+1.046	+53.090	-17.454	-3.725
	$\kappa_{\text{age}} = 0$	$\kappa_{\text{inc}} = \infty$	+0.458	+5.853	+0.198	+63.813	-20.496	-2.307
	Logit		+0.630		+1.021		-2.405	
	Probit		+0.563		+1.486		-2.309	
	Mixed Logit I		+0.606		+1.220		-2.379	
	Mixed Logit II		+0.431		+2.394		-2.074	
	Mixed Logit III		+0.272		+4.020		-1.992	

Notes: *High deductible is Bronze and low deductible is a bundle consisting of Silver, Gold, and Platinum. See Appendix S7 for further details on implementation and computation.*

4.4 Elasticities

Table 5 reports estimated bounds on average region-level elasticities. To avoid dividing by zero choice shares, we compute these elasticities at an aggregated level by grouping Silver, Gold, and Platinum into a single low deductible category, with the other option being high deductible Bronze plans. Appendix S7 contains more details on implementation.

The Lerner index computed at the endpoints of the bounds implies a markup for marginal buyers of between 10% and 59% for high-deductible Bronze plans, and between 6% and 51% for low-deductible plans.[11]

Own-price elasticity estimates from the parametric models tend to be towards the nonparametric upper bounds, although there is considerable variation for the high deductible Bronze plans. The same is true for cross-price elasticities with the exception of the elasticity of the outside option to the price of the high deductible Bronze plans, which the parametric models estimate to be at or above the nonparametric upper bound.

The sensitivity analysis in Table 5 aligns with the price variation in the data (Figure

[11] Markups for inframarginal buyers can be much smaller (possibly negative) if there is adverse selection and higher risk buyers are less price sensitive (e.g. Einav, Finkelstein, and Cullen, 2010; Tebaldi, 2017; Einav et al., 2019; Polyakova and Ryan, 2019).

2). The elasticities of choosing the outside option rely more on income variation than age variation, with the bounds widening only somewhat with $\kappa_{\text{age}} > 0$, but considerably more with $\kappa_{\text{inc}} > 0$. On the other hand, cross-tier elasticities rely more on age variation, which "rotates" the tier prices in a way that income variation does not. Own-tier elasticities seem to use both sources of variation in equal measure.

5 Welfare and Market Design

5.1 Consumer Surplus and Government Spending

The change in consumer surplus from a change in premiums to $\pi(M_i, X_i) + \delta$ is

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

The associated change in government spending is

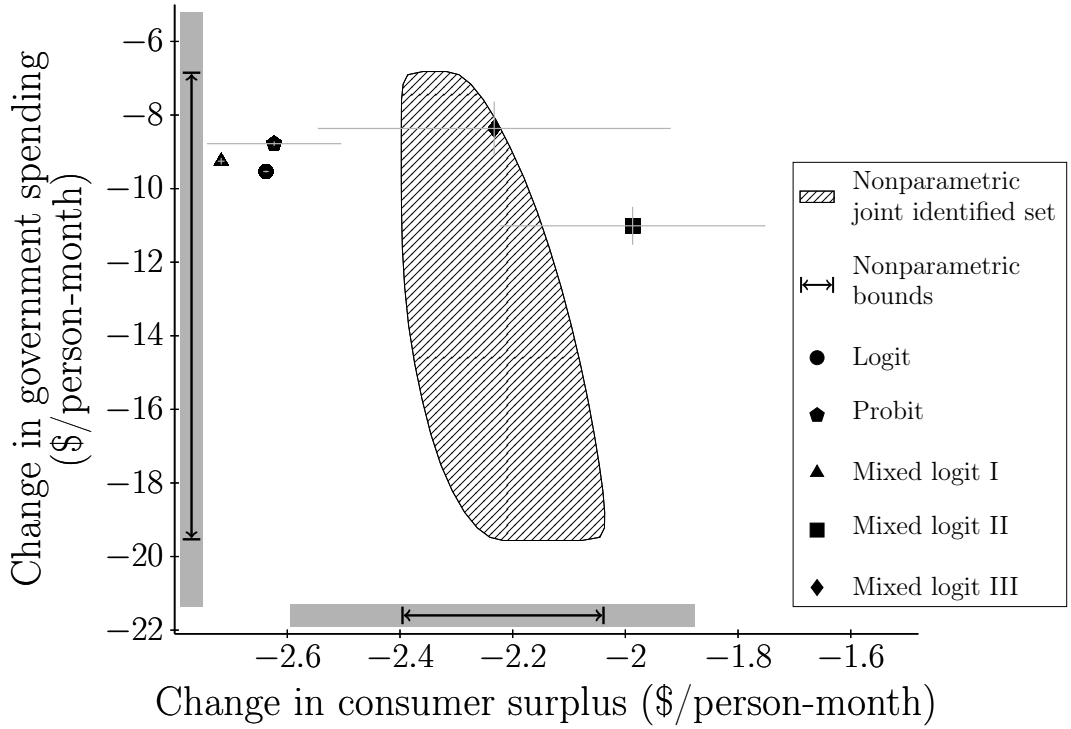
$$\begin{aligned}\Delta\text{GS}^{\delta}(m, x; f) \equiv \sum_{j \geq 1} (\text{Sub}_j(m, x) - \delta_j) \times \left[\int_{\mathcal{V}_j(\pi(m,x)+\delta)} f(v|m, x)\, dv\right] \\ - \sum_{j \geq 1} \text{Sub}_j(m, x) \times \left[\int_{\mathcal{V}_j(\pi(m,x))} f(v|m, x)\, dv\right],\end{aligned}$$

where $\text{Sub}_j(m, x)$ denotes the baseline premium subsidy for purchasing plan j. We aggregate both $\Delta\text{CS}^{\delta}(m, x; f)$ and $\Delta\text{GS}^{\delta}(m, x; f)$ into single measures $\Delta\text{CS}^{\delta}(f)$ and $\Delta\text{GS}^{\delta}(f)$ by averaging over regions and demographics, the same as in (20).

Figure 5 shows the estimated joint identified set for ΔCS^{δ} and ΔGS^{δ} when δ corresponds to a \$10 decrease in subsidies. The subsidy decrease would lead to a reduction in average monthly consumer surplus of between \$2.04 and \$2.40 per person, or between \$63 and \$74 million yearly when aggregated, and a reduction in government spending of between \$6.82 and \$19.56 per person, for a total of between \$209 and \$601 million per year. The joint identified set is not rectangular: larger consumer surplus declines would be accompanied by smaller declines in government spending. The large spending declines are due to the marginal buyers who exit the market and relinquish their entire premium subsidy, which in most cases is considerably greater than \$10.

Figure 5 also plots the five point predictions from the parametric models introduced in Section 3.4. The models yield similar predictions for the decline in government spending, all of which are towards the nonparametric upper bound. This is not surprising in light of Figures 3 and 4. Lower price sensitivity means fewer consumers leave the market when subsidies decrease, and all of the parametric models produced

Figure 5: *Changes in consumer surplus and government spending from a $10 decrease in subsidies*



Notes: *Bound and point estimates are shown in solid black. One-dimensional 95% confidence intervals are shown in grey vertical and horizontal bars.*

similarly insensitive participation responses.

More surprising is that the parametric models produce consumer surplus predictions that are both larger and smaller than the nonparametric bounds. Only the richest model (mixed logit III) makes a consumer surplus prediction within the nonparametric bounds. The usual intuition is that lower price sensitivity would lead to larger consumer surplus declines. However, the results in Figures 3 and 5 show that consumer surplus predictions in the parametric models are also driven by specific modeling decisions about the functional form of unobserved heterogeneity, even ones that yield similar estimates of price sensitivity.

Table 6 reports nonparametric consumer surplus changes by income. Changes for the lower-income sample ($2.60–$3.10) are estimated to be roughly twice as large as for the higher-income sample ($1.32–$1.50). This is likely due to the higher participation rate among the lower-income sample. Table 6 also reports sensitivity analysis for the consumer surplus and government spending estimates. As with the participation estimates in the previous section, the results are quite robust to relaxing the age invariance assumption and depend more on the income invariance assumption.

Table 6: *Aggregate impacts from reducing premium subsidies by $10 per month*

	140 - 400% FPL Change in consumer surplus ($ million/year) Bounds/Point estimate		140 - 250% FPL Change in consumer surplus ($ million/year) Bounds/Point estimate		250 - 400% FPL Change in consumer surplus ($ million/year) Bounds/Point estimate		140 - 400% FPL Change in government spending ($ million/year) Bounds/Point estimate	
Nonparametric	-73.67	-62.55	-53.48	-44.78	-20.19	-17.77	-601.13	-209.48
$\kappa_{\text{age}} = 0.4$ $\kappa_{\text{inc}} = 0$	-73.92	-62.22	-53.53	-44.62	-20.39	-17.60	-620.30	-219.06
$\kappa_{\text{age}} = \infty$ $\kappa_{\text{inc}} = 0$	-75.74	-56.68	-54.67	-41.17	-21.07	-15.52	-751.93	-188.15
$\kappa_{\text{age}} = 0$ $\kappa_{\text{inc}} = 0.4$	-71.77	-51.13	-51.71	-35.36	-20.06	-15.78	-1124.60	-402.24
$\kappa_{\text{age}} = 0$ $\kappa_{\text{inc}} = \infty$	-74.46	-11.88	-52.99	-5.97	-21.47	-5.91	-2087.52	-280.70
Logit	-81.05		-57.14		-23.91		-292.97	
Probit	-80.29		-56.23		-24.06		-269.72	
Mixed Logit I	-83.49		-58.23		-25.26		-284.37	
Mixed Logit II	-61.06		-40.85		-20.21		-338.32	
Mixed Logit III	-68.61		-46.74		-21.87		-257.02	

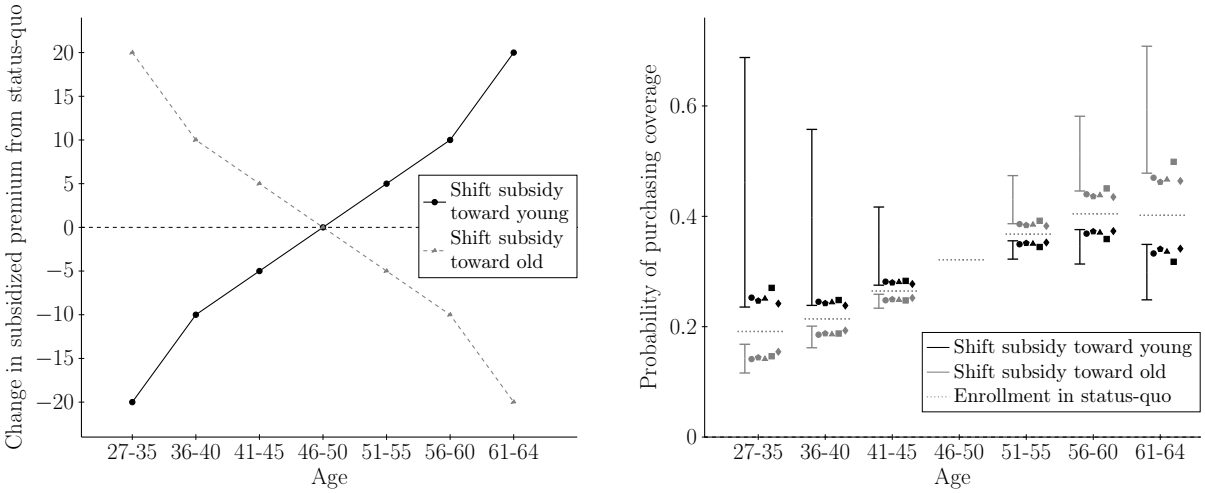
These results are consistent with a growing number of empirical analyses showing that consumers value individual health insurance significantly less than it costs in subsidies to induce them to purchase a plan (e.g. Finkelstein et al., 2019). An important caveat when interpreting welfare estimates in this context is that they do not account for the existence of potentially large healthcare externalities such as the cost of uncompensated care, debt delinquency, or bankruptcy (Finkelstein et al., 2012; Mahoney, 2015; Garthwaite, Gross, and Notowidigdo, 2018).

5.2 Linking Subsidies to Age

Regulated health insurance exchanges like Covered California rely on the participation of healthy consumers. A growing body of research considers linking subsidies to demographics as a tool for encouraging healthy consumers to participate (e.g. Tebaldi, 2017; Polyakova and Ryan, 2019; Decarolis et al., 2020). Figure 6a depicts two counterfactual regulatory changes that directly link premium subsidies to age. In one counterfactual, subsidies are shifted from older buyers to younger buyers, while in the other the shift goes the opposite way.

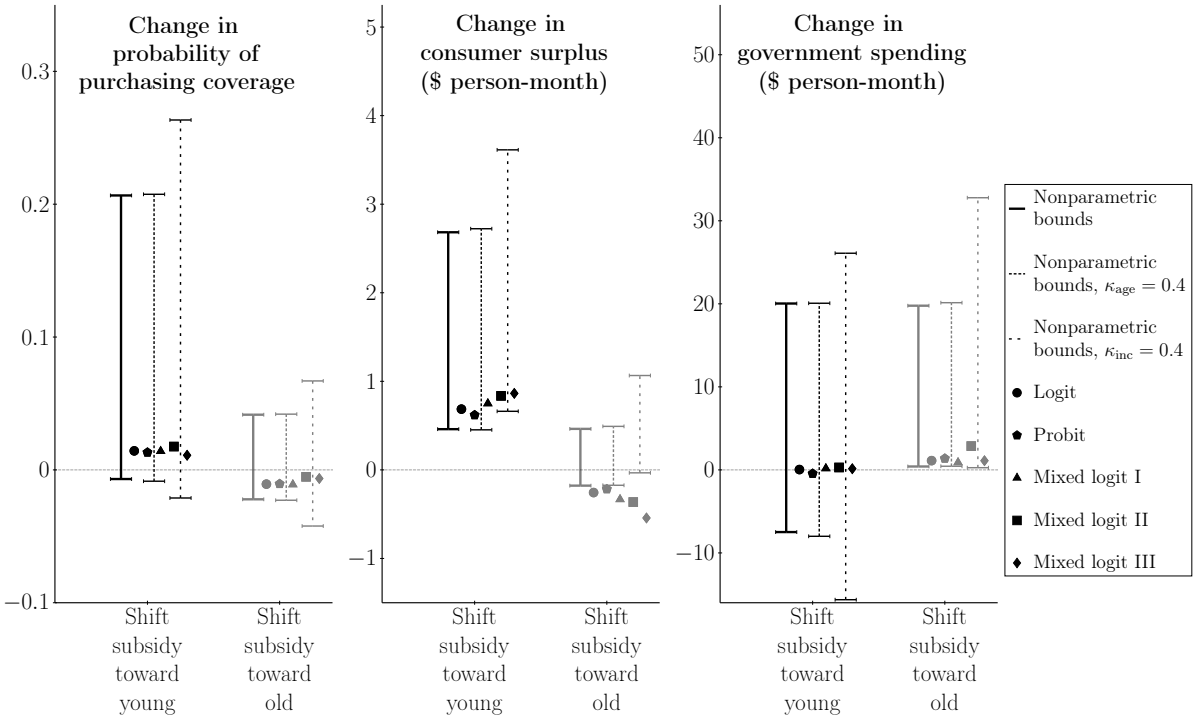
Figures 6b and 6c report the estimated effects of the two counterfactuals. Shifting subsidies toward younger buyers could have a potentially large positive impact on both their participation and on aggregate participation, while decreasing the participation of older buyers comparatively less, even in the worst case. While the change in consumer surplus would naturally be positive for younger buyers and negative for older buyers, we find that average consumer surplus would unambiguously increase by between $0.46 and $2.68 per person, per month. The impacts on government spending could be either positive or negative, depending on exactly which sets of buyers change their purchase decisions. In contrast, shifting subsidies toward older buyers would increase government spending without necessarily increasing participation or average consumer surplus.

Figure 6: *Linking subsidies to age*



(a) *Counterfactual changes to premiums*

(b) *Changes in enrollment by age group*



(c) *Effects on aggregate outcomes*

Figures 6b and 6c also include estimated effects for the five parametric models. As expected, they indicate price sensitivity toward the lower end of the nonparametric bounds. The parametric models all predict positive aggregate consumer surplus impacts from shifting subsidies toward younger buyers—like the nonparametric model—but potentially understate the magnitude of these impacts by a large amount. The parametric models predict no government spending impact from shifting subsidies towards younger buyers, whereas the nonparametric model says the data is consistent with either moderate decreases or large increases. Similarly, the parametric models unambiguously predict that the average consumer surplus impact of shifting subsidies towards older buyers would be negative, whereas the nonparametric model implies it could be either positive or negative.

5.3 Removing Silver Plans

The CSRs in Covered California make Silver plans especially attractive for low-income consumers (Tables 1 and 2). Without Silver plans, subsidized consumers would face a more standard trade-off between premiums and actuarial value. If the option corresponding to Silver plans ($j = 2$) were removed, the consumer surplus impact would be

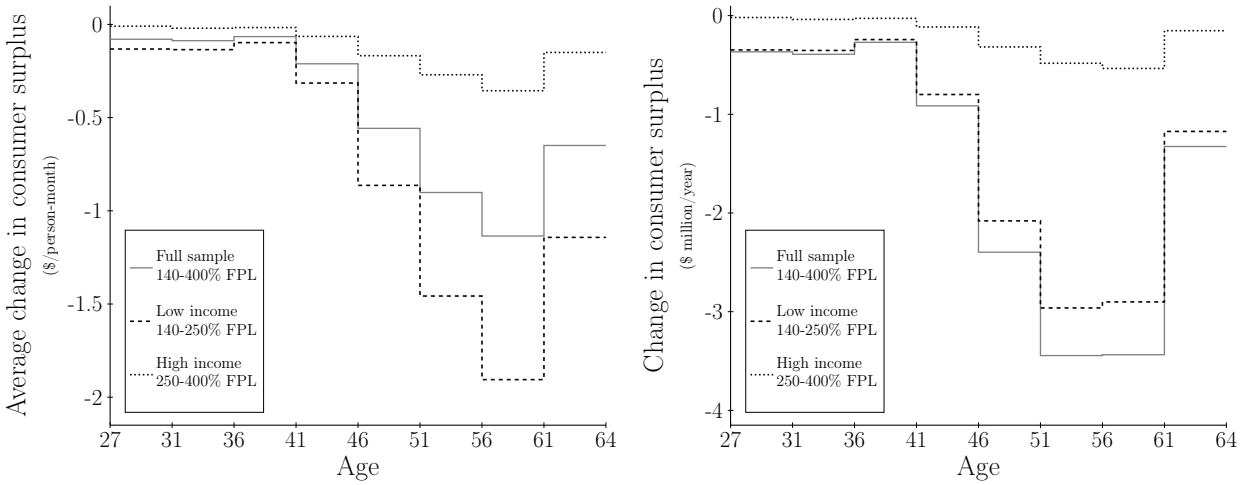
$$\Delta\text{RS}(m, x; f) \equiv \int \left[\max_{j \in (\mathcal{J} \setminus \{2\})} \{v_j - \pi_j(m, x)\} - \max_{j \in \mathcal{J}} \{v_j - \pi_j(m, x)\} \right] f(v|m, x) dv.$$

As with the other parameters, we aggregate $\Delta\text{RS}(m, x; f)$ into $\Delta\text{RS}(f)$ by averaging over regions and demographics.

An interesting property of the nonparametric model is that the sharp lower bound on ΔRS is infinite ($-\infty$), at least unless we observe a choke price for Silver plans in the data (which we do not). The explanation is intuitive: without a parametric form for f, there is nothing to restrict the tails of the valuation for Silver, allowing for the possibility that some consumers have an unbounded preference for Silver plans. We view this property as a transparent benefit of the nonparametric model. It implies that parametric models rely on functional form to pick out a single consumer surplus impact from an unbounded set of possibilities. While an arbitrarily large consumer surplus impact is implausible, the fact that neither nonparametric assumptions nor the data rule it out highlights the role played by the specific choice of parameterization.

We focus instead on estimating the sharp upper bound on the decrease in consumer surplus from removing Silver plans. Figure 7 shows that younger consumers would experience less surplus loss in the best-case scenario. The finding makes sense because these consumers also face a smaller premium differences between Bronze, Silver, and

Figure 7: *Upper bounds on the change in consumer surplus from removing Silver plans*



(a) *Average impacts per person/month* ***(b)*** *Aggregate impacts*

Gold plans. Figure 7 also shows that low income consumers would bear the brunt of the surplus loss, which also makes sense because it is these consumers who receive the CSRs incorporated into the terms of Silver plans (Table 1). Table 7 shows that nearly $11 million of the aggregate best-case surplus change of $12.6 million would be borne by consumers with income less than 250% of the FPL.

Estimates from the parametric models are also reported in Table 7. The parametric estimates all indicate consumer surplus impacts to be 8–10 times as large for lower income consumers, similar to the ratio of the nonparametric upper bounds. The magnitude of the estimates range from a decline of $162 to a decline of $290 million, which is between 2.5–4 times as large as the estimated impact from increasing all premiums by $10 per month. The sensitivity of the estimates to the precise type of logit model used is reminiscent of findings by Petrin (2002). According to the nonparametric model, any figure between $11 million and infinity is equally well-supported by the observed choice data.

The sensitivity estimates in Table 7 show that these conclusions rely on both invariance to age and invariance to income. Fully removing either assumption renders the bounds essentially uninformative. The intuition can again be seen from Figure 7. Fixing income, post-subsidy premiums for Silver do not vary in age, making it difficult to pin down substitution between Silver and non-participation. Fixing age, post-subsidy premiums for Silver increase in lock-step with income, making it difficult to pin down substitution between Silver and other tiers. Information on both types of substitution are needed to infer the welfare impacts of removing Silver plans from the choice set.

Table 7: *Aggregate impacts from removing Silver plans*

		140 - 400% FPL Change in consumer surplus ($ million/year) Bounds/Point estimate		140 - 250% FPL Change in consumer surplus ($ million/year) Bounds/Point estimate		250 - 400% FPL Change in consumer surplus ($ million/year) Bounds/Point estimate	
Nonparametric		$-\infty$	-12.55	$-\infty$	-10.86	$-\infty$	-1.69
$\kappa_{\text{age}} = 0.4$	$\kappa_{\text{inc}} = 0$	$-\infty$	-11.73	$-\infty$	-10.34	$-\infty$	-1.39
$\kappa_{\text{age}} = \infty$	$\kappa_{\text{inc}} = 0$	$-\infty$	-0.68	$-\infty$	-0.68	$-\infty$	-0.00
$\kappa_{\text{age}} = 0$	$\kappa_{\text{inc}} = 0.4$	$-\infty$	-9.71	$-\infty$	-8.31	$-\infty$	-1.40
$\kappa_{\text{age}} = 0$	$\kappa_{\text{inc}} = \infty$	$-\infty$	-1.72	$-\infty$	-1.40	$-\infty$	-0.32
Logit		-279.43		-243.63		-35.80	
Probit		-287.09		-257.69		-29.40	
Mixed Logit I		-289.75		-252.08		-37.67	
Mixed Logit II		-162.44		-146.85		-15.59	
Mixed Logit III		-171.30		-162.55		-8.75	

6 Conclusion

We estimated the demand for health insurance in California's ACA marketplace using a new nonparametric methodology. The central idea of the method is to divide consumer valuations into a minimal relevant partition (MRP) of sets for which behavior remains constant under all considered prices. Using the MRP, we developed a scalable linear programming procedure for consistently estimating sharp identified sets for policy-relevant target parameters. We believe the method should be useful for other discrete choice problems as well. For example, it could be used to remove the large support assumption in the ideological voting model analyzed by Merlo and de Paula (2016).

The nonparametric estimates point to the possibility of substantially greater price sensitivity than would be recognized using comparable parametric models. This finding is consistent with the folklore that logits are "flat" models. The greater price sensitivity estimates in turn have important welfare implications for counterfactual policy changes to subsidies or plan offerings. However, we also found direct evidence that specific parametric functional forms have first-order impacts on consumer surplus estimates that operate through channels other than price sensitivity. The results provide a clear example in which functional form assumptions about the distribution of unobserved heterogeneity are not innocuous, and actually play a leading role in driving empirical conclusions.

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. 27

Ackerberg, D. A. and M. Rysman (2005): "Unobserved Product Differentiation in Discrete-Choice Models: Estimating Price Elasticities and Welfare Effects," *The RAND Journal of Economics*, 36, 771–788. 7

Allen, R. and J. Rehbeck (2019): "Identification With Additively Separable Heterogeneity," *Econometrica*, 87, 1021–1054. 5

Andrews, D. W. K. and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157. 19

Armstrong, T. B. (2016): "Large Market Asymptotics for Differentiated Product Demand Estimators With Economic Models of Supply," *Econometrica*, 84, 1961–1980. 9

Berry, S. and P. Haile (2015): "Identification in Differentiated Products Markets," *Annu. Rev. Econ.*, –. 9, 10

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890. 2, 5, 9

——— (2004a): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112, 68–105. 9

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. 9

Berry, S. and A. Pakes (2007): "THE PURE CHARACTERISTICS DEMAND MODEL*," *International Economic Review*, 48, 1193–1225. 7

Berry, S. T. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25, 242–262. 2

Berry, S. T. and P. A. Haile (2010): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," *Cowles Foundation Discussion Paper 1718*. 3, 9

——— (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797. 3, 10, 15

——— (2018): "Identification of Nonparametric Simultaneous Equations Models With a Residual Index Structure," *Econometrica*, 86, 289–315. 3

Bertsimas, D. and J. N. Tsitsiklis (1997): *Introduction to linear optimization*, vol. 6, Athena Scientific Belmont, MA. 18

Briesch, R. A., P. K. Chintagunta, and R. L. Matzkin (2010): "Nonparametric Discrete Choice Models With Unobserved Heterogeneity," *Journal of Business & Economic Statistics*, 28, 291–307. 5

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. 3

Chan, D. and J. Gruber (2010): "How Sensitive are Low Income Families to Health Plan Prices?" *American Economic Review*, 100, 292–296. 2, 7

Charnes, A. and W. W. Cooper (1962): "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, 9, 181–186. 17

Chesher, A., A. Rosen, and K. Smolinski (2011): "An Instrumental Variable Model of Multiple Discrete Choice," Tech. rep. 15

Chesher, A., A. M. Rosen, and K. Smolinski (2013): "An instrumental variable model of multiple discrete choice," *Quantitative Economics*, 4, 157–196. 4, 5, 11, 13, 15

Chiappori, P.-A. and I. Komunjer (2009): "On the Nonparametric Identification of Multiple Choice Models," *Working paper*. 3

Chiong, K. X., Y.-W. Hsieh, and M. Shum (2017): "Counterfactual Estimation in Semiparametric Multinomial Choice Models," *SSRN Electronic Journal*. 5

Compiani, G. (2019): "Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach," *Working paper*. 3, 6

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. 2

Conley, T. G., C. B. Hansen, and P. E. Rossi (2010): "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260–272. 12

Dafny, L., K. Ho, and M. Varela (2013): "Let them have choice: Gains from shifting away from employer-sponsored health insurance and toward an individual exchange," *American Economic Journal: Economic Policy*, 5, 32–58. 7

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2021): "Revealed Price Preference: Theory and Empirical Analysis," *arXiv:1801.02702 [econ]*. 19

Decarolis, F. (2015): "Medicare part d: Are insurers gaming the low income subsidy design?" *American Economic Review*, 105, 1547–80. 7

Decarolis, F., M. Polyakova, and S. P. Ryan (2020): "Subsidy Design in Privately Provided Social Insurance: Lessons from Medicare Part D," *Journal of Political Economy*, 128, 1712–1752. 7, 35

DeLeire, T., A. Chappel, K. Finegold, and E. Gee (2017): "Do individuals respond to cost-sharing subsidies in their selections of marketplace health insurance plans?" *Journal of Health Economics*, 56, 71 – 86. 7

Drake, C. (2019): "What Are Consumers Willing to Pay for a Broad Network Health Plan?: Evidence from Covered California," *Journal of Health Economics*, 65, 63–77. 7

Einav, L., A. Finkelstein, and M. R. Cullen (2010): "Estimating Welfare in Insurance Markets Using Variation in Prices," *Quarterly Journal of Economics*, 125, 877–921. 32

Einav, L., A. Finkelstein, and P. Tebaldi (2019): "Market design in regulated health insurance markets: Risk adjustment vs. subsidies," *Unpublished mimeo, Stanford University, MIT, and University of Chicago*. 7, 32

Einav, L. and J. Levin (2015): "Managed competition in health insurance," *Journal of the European Economic Association*, 13, 998–1021. 2

Ericson, K. M. and A. Starc (2015): "Pricing Regulation and Imperfect Competition on the Massachusetts Health Insurance Exchange," *The Review of Economics and Statistics*, 97, 667–682. 2, 7, 30

Finkelstein, A., N. Hendren, and M. Shepard (2019): "Subsidizing Health Insurance for Low-Income Adults: Evidence from Massachusetts," *American Economic Review*, forthcoming. 7, 22, 27, 35

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

Fosgerau, M. and D. Kristensen (2020): "Identification of a Class of Index Models: A Topological Approach," *arXiv:2004.07900 [econ, math, stat]*. 5

Fox, J. T. (2007): "Semiparametric estimation of multinomial discrete-choice models using a subset of choices," *The RAND Journal of Economics*, 38, 1002–1019. 5

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. 3

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. 30

Goldberg, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63, 891–951. 2

Gurobi Optimization, I. (2015): "Gurobi Optimizer Reference Manual," . 17

Hackmann, M. B., J. T. Kolstad, and A. E. Kowalski (2015): "Adverse selection and an individual mandate: When theory meets practice," *American Economic Review*, 105, 1030–66. 7

Handel, B., I. Hendel, and M. D. Whinston (2015): "Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk," *Econometrica*, 83, 1261–1313. 8

Handel, B. R. (2013): "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103, 2643–82. 8

Hausman, J. A. (1996): *Valuation of New Goods under Perfect and Imperfect Competition*, University of Chicago Press, 207–248. 7

Ho, K. and A. Pakes (2014): "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, 104, 3841–84. 5, 6

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. 3

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

Jaffe, S. and M. Shepard (2020): "Price-linked subsidies and imperfect competition in health insurance," *American Economic Journal: Economic Policy*, 12, 279–311. 7

Kamat, V. (2020): "Identification of Program Access Effects with an Application to Head Start," *arXiv:1711.02048 [econ]*. 5, 17

Khan, S., F. Ouyang, and E. Tamer (2019): "Inference on Semiparametric Multinomial Response Models," *Working paper*. 5

Kitamura, Y. and J. Stoye (2018): "Nonparametric Analysis of Random Utility Models," *Econometrica*, 86, 1883–1909. 19

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. 5

Krueger, A. B. and I. Kuziemko (2013): "The demand for health insurance among uninsured Americans: Results of a survey experiment and implications for policy," *Journal of health economics*, 32, 780–793. 7

Lewbel, A. (2000): "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177. 5

——— (2019): "The Identification Zoo: Meanings of Identification in Econometrics," *Journal of Economic Literature*, 57, 835–903. 10

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. 3, 5

——— (2007): "Partial Indentification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393–1410. 3, 5, 15

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

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*. 12

Marone, V. R. and A. Sabety (2019): "Should There be Vertical Choice in Health Insurance Markets?" Tech. rep., Working Paper. 7

Matzkin, R. L. (1993): "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 58, 137–168. 3, 5

McFadden, D. and K. Train (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447–470. 2

Merlo, A. and Á. de Paula (2016): "Identification and Estimation of Preference Distributions When Voters Are Ideological," *The Review of Economic Studies*, rdw046. 39

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. 21

Mogstad, M., A. Santos, and A. Torgovitsky (2018): "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters," *Econometrica*, 86, 1589–1619. 18

Molinari, F. (2020): "Microeconometrics with Partial Identification," in *Handbook of Econometrics*, ed. by S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, Elsevier, vol. 7 of *Handbook of Econometrics, Volume 7A*, 355–486. 3

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–342. 5, 9

——— (2011): "Empirical Models of Consumer Behavior," *Annual Review of Economics*, 3, 51–75. 9

Nevo, A. and A. M. Rosen (2012): "Identification With Imperfect Instruments," *Review of Economics and Statistics*, 94, 659–671. 11

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578. 3

Orsini, J. and P. Tebaldi (2017): "Regulated Age-based Pricing in Subsidized Health Insurance: Evidence from the Affordable Care Act," *Available at SSRN 2464725*. 20, 30

Pakes, A. (2010): "Alternative Models for Moment Inequalities," *Econometrica*, 78, 1783–1822. 5

Pakes, A. and J. Porter (2016): "Moment Inequalities for Multinomial Choice with Fixed Effects," Tech. rep. 5

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006): "Moment Inequalities and Their Application," *Working paper*. 5

——— (2015): "Moment Inequalities and Their Application," *Econometrica*, 83, 315–334. 5

Petrin, A. (2002): "Quantifying the benefits of new products: The case of the minivan," *Journal of political Economy*, 110, 705–729. 7, 38

Polyakova, M. and S. P. Ryan (2019): "Subsidy targeting with market power," Tech. rep., National Bureau of Economic Research. 2, 6, 7, 32, 35

Rambachan, A. and J. Roth (2020): "An Honest Approach to Parallel Trends," *Working paper*. 12

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*. 22

Saltzman, E. (2019): "Demand for health insurance: Evidence from the California and Washington ACA exchanges," *Journal of Health Economics*, 63, 197 – 222. 2, 6, 7

——— (2021): "Managing Adverse Selection: Underinsurance versus Underenrollment," *The RAND Journal of Economics*, 52, 359–381. 6

Santos, A. (2012): "Inference in Nonparametric Instrumental Variables With Partial Identification," *Econometrica*, 80, 213–275. 3

Shepard, M. (2016): "Hospital network competition and adverse selection: evidence from the Massachusetts health insurance exchange," Tech. rep., National Bureau of Economic Research. 7

Shi, X., M. Shum, and W. Song (2016): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *SSRN Electronic Journal*. 5

Tamer, E. (2010): "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167–195. 3

Tebaldi, P. (2017): "Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA," *Working paper*. 2, 6, 7, 22, 32, 35

Thompson, T. S. (1989): "Identification of Semiparametric Discrete Choice Models," *Discussion paper 249, Center for Economics Research, Department of Economics, University of Minnesota*. 3, 5, 15

Torgovitsky, A. (2019): "Nonparametric Inference on State Dependence in Unemployment," *Econometrica*, 87, 1475–1505. 12