Online Appendix for "Nonparametric Estimates of Demand in the California Health Insurance Exchange"

Pietro Tebaldi* Alexander Torgovitsky† Hanbin Yang‡

July 22, 2021

S1 A Model of Insurance Choice

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

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

$$Y_i = \underset{j \in \mathcal{J}}{\arg\max} \int U_{ij}(e) \, dG_{ij}(e), \tag{S1}$$

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

$$U_{ij}(e) = -\frac{1}{A_i} \exp(-A_i C_{ij}(e)), \tag{S2}$$

*Department of Economics, Columbia University, and NBER. Research supported by the Becker Friedman Institute Health Economics Initiative.

†Kenneth C. Griffin Department of Economics, University of Chicago. Research supported by National Science Foundation grants SES-1530538 and SES-1846832.

‡Harvard Business School.

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

$$C_{ij}(e) = \text{Inc}_i - P_{ij} - e + \widetilde{V}_{ij}, \tag{S3}$$

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

Substituting (S3) into (S2) and then into (S1), we obtain

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

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

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

Eliminating additive terms that do not depend on plan choice yields

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

Suppose that $A_i > 0$, so that all consumers are risk averse.[1] Then we can express the consumer's choice as

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

which takes the form of (1) with

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

Examining the components of V_{ij} reveals the factors that contribute to heterogeneous

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

valuations in this model. Heterogeneity across i can come from variation in risk aversion (A_i), from differences in risk factors or beliefs (G_{ij}), and from idiosyncratic differences in the valuation of health insurance ($\widetilde{V}_{ij}$). Differences in valuations across j arise from the interaction between risk factors and the corresponding distribution of expenditures (G_{ij}), as well as from idiosyncratic differences in valuations across plans ($\widetilde{V}_{ij}$). The main restrictions in this model are the assumption of CARA preferences in (S2) and the quasilinearity of ex-post consumption in (S3). However, as noted in the main text, these restrictions do not have empirical content until they are combined with an assumption about the dependence between income (here called Inc_i) and the preference parameters, A_i and $\widetilde{V}_{ij}$.

S2 Modifications for Less Price Variation

The discussion in the main text is tailored to the situation in which P_i still varies conditional on M_i. This is the case in the application to Covered California. In this section, we discuss how to modify our approach to settings in which prices do not vary within markets, as in the "market-level" data setting considered by Berry, Levinsohn, and Pakes (1995), Nevo (2001), and Berry and Haile (2014). As a technical matter, our methodology applies exactly as before to this case. However, since there is only a single price per market, and since we are not assuming anything about how demand varies across markets, the resulting bounds will be uninformative. Here we suggest two additional assumptions that could potentially be used to compensate for limited price variation.

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

$$Y_i = \underset{j \in \mathcal{J}}{\arg\max}\; V_{ij} + X_i'\beta_j - P_{ij}, \tag{S4}$$

where $\beta \equiv (\beta_1, \ldots, \beta_J)$ are unknown parameter vectors. For each fixed β, this model is like (1) but with "prices" given by $\tilde{P}_{ij}(\beta) \equiv P_{ij} - X_i'\beta_j$. While P_{ij} does not vary within markets, $\tilde{P}_{ij}(\beta)$ can if a component of X_i does. In order to make use of this variation, that component of X_i needs to be independent of V_i, which is a common assumption in empirical implementations of (2). In our framework, this independence can be incorporated by modifying the instrumental variable assumptions in Section 2.5.1.

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

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

$$Y_i = \arg\max_{j \in \mathcal{J}} V_{ij} + \xi_j(M_i) - P_{ij}, \tag{S5}$$

where ξ_j is an unknown function of the consumer's market. For each fixed ξ, this model is like (1) but with valuations given by $\tilde{V}_{ij}(\xi) \equiv V_{ij} + \xi_j(M_i)$. After incorporating unobserved product-market effects in this way, one may be willing to assume that V_{ij} is independent of P_{ij} (perhaps conditional on M_i), as is common in implementations of (2). This can be imposed with Assumption IV. While there is still only a single price per market, (S5) together with such an independence assumption enables aggregation across markets by requiring the distribution of valuations to be the same up to a location shift.

Implementing either (S4) or (S5) requires looping over possible parameter values β or ξ. For each candidate β and ξ, one can characterize and compute an identified set exactly as before, so such a procedure will still be sharp. We apply this procedure in Appendix S5 to a small-scale simulation considered by Chesher, Rosen, and Smolinski (2013). Developing a computational strategy that is feasible at scale is more challenging, but not impossible. Since neither (S4) or (S5) are needed for our application, we leave fuller investigations of these extensions to future work.

S3 Construction of the Minimal Relevant Partition (MRP)

We first observe that any price (premium) vector $p \in \mathbb{R}^J$ divides $\mathbb{R}^J$ into the sets $\{\mathcal{V}_j(p)\}_{j=0}^J$, as shown in Figures 1a and 1b. Intuitively, we view such a division as a partition, although formally this is not correct, since these sets can overlap on boundary regions like $\{v \in \mathbb{R}^J : v_j - p_j = v_k - p_k\}$ where ties occurs. For the same reason, "the" MRP is not unique, since one could consider a boundary region to be in either of the sets to which it is a boundary. The boundary regions have Lebesgue measure zero in $\mathbb{R}^J$, so these caveats are unimportant given our focus on continuously distributed valuations. However, to avoid confusion, we refer to a collection of sets that would be a partition if not for regions of Lebesgue measure zero as an almost sure (a.s.) partition.

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

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

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

$$\left\{\widetilde{\mathcal{V}}_y : y \in \mathcal{Y}\right\} \quad \text{where} \quad \widetilde{\mathcal{V}}_y \equiv \bigcap_{l=1}^{L} \mathcal{V}_{y_l}(p_l) \tag{S6}$$

also constitutes an a.s. partition of $\mathbb{R}^J$.[3] Intuitively, each vector $y \equiv (y_1, \ldots, y_L)$ is a profile of L choices made under the price vectors $(p_1, \ldots, p_L)$ that comprise $\mathcal{P}$. Each set $\widetilde{\mathcal{V}}_y$ in the a.s. partition (S6) corresponds to the subset of valuations in $\mathbb{R}^J$ for which a consumer would make choices $y = (y_1, \ldots, y_L)$ when faced with prices $(p_1, \ldots, p_L)$.

The collection $\mathbb{V} \equiv \{\widetilde{\mathcal{V}}_y : y \in \mathcal{Y}\}$ is the MRP, since it satisfies Definition MRP by construction. To see this, consider any $v, v' \in \mathbb{R}^J$. If $v, v' \in \widetilde{\mathcal{V}}_y$ for some y, then by (S6), $v, v' \in \mathcal{V}_{y_l}(p_l)$ for all $l = 1, \ldots, L$, at least up to collections of v, v' that have Lebesgue measure zero. Recalling (8) and the notation of Definition MRP, this implies that $Y(v, p) = Y(v', p)$ for all $p \in \mathcal{P}$. Conversely, if $Y(v, p) = Y(v', p)$ for all $p \in \mathcal{P}$, then taking

$$y \equiv (Y(v, p_1), \ldots, Y(v, p_L)) = (Y(v', p_1), \ldots, Y(v', p_L)), \tag{S7}$$

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

However, from a practical perspective, this is an inadequate representation of the MRP, because if choices are determined by the quasilinear model (1), then many of the sets $\widetilde{\mathcal{V}}_y$ must have Lebesgue measure zero. This makes indexing the partition by $y \in \mathcal{Y}$ excessive; for computation we would prefer an indexing scheme that only includes sets that are not already known to have measure zero. For this purpose, we use an algorithm that starts with the set of prices $\mathcal{P}$ and returns the collection of choice sequences $\overline{\mathcal{Y}}$ that are not required to have Lebesgue measure zero under (1). We use this set $\overline{\mathcal{Y}}$ in our computational implementation. Note that since $\widetilde{\mathcal{V}}_y$ has Lebesgue measure zero for any $y \in \mathcal{Y} \setminus \overline{\mathcal{Y}}$, the collection $\mathbb{V} \equiv \{\widetilde{\mathcal{V}}_y : y \in \overline{\mathcal{Y}}\}$ still constitutes an a.s. partition of $\mathbb{R}^J$ and still satisfies the key property (10) of

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

Definition MRP.

The algorithm works as follows.[4] We begin by partitioning $\mathcal{P}$ into T sets (or blocks) of prices $\{\mathcal{P}_t\}_{t=1}^T$ that each contain (give or take) L_0 prices. For each t, we then construct the set of all choice sequences $\overline{\mathcal{Y}}_t \subseteq \mathcal{J}^{|\mathcal{P}_t|}$ that are compatible with the quasilinear choice model in the sense that $y^t \in \overline{\mathcal{Y}}_t$ if and only if the set

$$\left\{v \in \mathbb{R}^J : v_{y_l^t} - p_{y_l^t} \geq v_j - p_j \text{ for all } j \in \mathcal{J} \text{ and } p \in \mathcal{P}_t\right\} \tag{S8}$$

is non-empty. In practice, we do this by sequentially checking the feasibility of a linear program with (S8) as the constraint set. The sense in which we do this sequentially is that instead of checking (S8) for all $y^t \in \mathcal{J}^{|\mathcal{P}_t|}$—which could be a large set even for moderate L_0—we first check whether it is nonempty when the constraint is imposed for only 2 prices in $\mathcal{P}_t$, then 3 prices, etc. Finding that (S8) is empty when restricting attention to one of these shorter choice sequences implies that it must also be empty for all other sequences that share the short component. This observation helps speed up the algorithm substantially.

Once we have found $\overline{\mathcal{Y}}_t$ for all t, we combine blocks of prices into pairs, then repeat the process with these larger, paired blocks. For example, if we let $\mathcal{P}_{12} \equiv \mathcal{P}_1 \cup \mathcal{P}_2$—i.e. we pair the first two blocks of prices—then we know that the set of $y^{12} \in \mathcal{J}^{|\mathcal{P}_1|+|\mathcal{P}_2|}$ that satisfy (S8) must be a subset of $\{(y_1, y_2) : y_1 \in \overline{\mathcal{Y}}_1, y_2 \in \overline{\mathcal{Y}}_2\}$. We sequentially check the non-emptiness of (S8) for all y^{12} in this set, eventually obtaining a set $\overline{\mathcal{Y}}_{12}$. Once we have done this for all pairs of price blocks, we then combine pairs of blocks (e.g. $\mathcal{P}_{12} \cup \mathcal{P}_{34}$) and repeat the process. Continuing in this way, we eventually end up with the original set of price vectors, $\mathcal{P}$, as well as the set of all surviving choice sequences, $\overline{\mathcal{Y}} \subseteq \mathcal{Y}$.

The key input to this algorithm is the number of prices in the initial price blocks, which we have denoted by L_0. The optimal value of L_0 should be something larger than 2, but smaller than L. With small L_0, the sequential checking of (S8) yields less payoff, since each detection of infeasibility eliminates fewer partial choice sequences. On the other hand, large L_0 makes the strategy of combining pairs of smaller blocks of prices into larger blocks less fruitful. For the application, we use L_0 between 8 and 10, which seems to be fairly efficient, although it is likely specific to our setting.

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

S4 Proofs for Propositions 1 and 2

S4.1 Proposition 1

Suppose that $t \in \Theta^\star$. By definition, there exists an $f \in \mathcal{F}^\star$ such that $\theta(f) = t$. We will show that

$$\tilde{\phi}(\mathcal{V}|p, m, x) \equiv \int_{\mathcal{V}} f(v|p, m, x)\, dv. \tag{S9}$$

is an element of $\Phi^\star(t)$.

First, note that $\tilde{\phi} \in \Phi$, because the MRP $\mathbb{V}$ is (almost surely) a partition of $\mathbb{R}^J$, and f is a conditional probability density function on $\mathbb{R}^J$. To see that $\tilde{\phi}$ satisfies (MD'), observe that

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

where the second equality follows from (7) using the definition of $\mathbb{V}_j(p)$, and the third holds for all $f \in \mathcal{F}^\star$. Similarly, $\tilde{\phi}$ satisfies the upper bound in (IV') because

$$\begin{aligned}
\tilde{\phi}(\mathcal{V}|w, z) &\equiv \mathbb{E}\left[\tilde{\phi}(\mathcal{V}|P_i, M_i, X_i)|W_i = w, Z_i = z\right] \\
&= \mathbb{E}\left[\int_{\mathcal{V}} f(v|P_i, M_i, X_i)\, dv \middle| W_i = w, Z_i = z\right] \\
&= \int_{\mathcal{V}} f(v|w, z)\, dv \\
&\leq (1 + \kappa(z, z', w)) \int_{\mathcal{V}} f(v|w, z')\, dv = (1 + \kappa(z, z', w))\tilde{\phi}(\mathcal{V}|w, z'),
\end{aligned}$$

where the third equality follows by Tonelli's Theorem (e.g. Shorack, 2000, pg. 82), the fourth uses Assumption IV, which is satisfied by any $f \in \mathcal{F}^\star$, and the final equality reverses the steps of the first three. An analogous argument shows that $\tilde{\phi}$ also satisfies the lower bound of (IV'). That $\tilde{\phi}$ satisfies (SP') follows because $f \in \mathcal{F}^\star$ satisfies Assumption SP, so

$$\begin{aligned}
\sum_{\mathcal{V} \in \mathbb{V}^\bullet(w)} \tilde{\phi}(\mathcal{V}|w, z) &= \sum_{\mathcal{V} \in \mathbb{V}^\bullet(w)} \int_{\mathcal{V}} \mathbb{E}\left[f(v|P_i, M_i, X_i)|W_i = w, Z_i = z\right] dv \\
&= \int_{\cup\{\mathcal{V}:\mathcal{V} \in \mathbb{V}^\bullet(w)\}} f(v|w, z)\, dv \geq \int_{\mathcal{V}^\bullet(w)} f(v|w, z)\, dv = 1.
\end{aligned} \tag{S10}$$

The inequality in (S10) uses the definition of $\mathbb{V}^\bullet(w)$, which together with the fact that $\mathbb{V}$ is an a.s. partition of $\mathbb{R}^J$, implies that $\mathcal{V}^\bullet(w)$ is contained in the union of sets in $\mathbb{V}^\bullet(w)$. Inequality (S10) implies that $\tilde{\phi}$ satisfies (SP'), because

$$\sum_{\mathcal{V}\in\mathbb{V}^\bullet(w)} \tilde{\phi}(\mathcal{V}|w,z) \leq \sum_{\mathcal{V}\in\mathbb{V}} \tilde{\phi}(\mathcal{V}|w,z) = \mathbb{E}\left[\sum_{\mathcal{V}\in\mathbb{V}} \tilde{\phi}(\mathcal{V}|P_i, M_i, X_i)\middle| W_i = w, Z_i = z\right] = 1,$$

as a result of $\tilde{\phi}$ being an element of Φ. Finally, since $\tilde{\phi} \equiv \overline{\theta}(f)$ as defined in (12), we immediately have that $\overline{\theta}(\tilde{\phi}) = \theta(f) = t$ due to Condition TP. We have now established that if there is an $f \in \mathcal{F}^\star$ with $\theta(f) = t$, then $\tilde{\phi} \in \Phi^\star(t)$.

Conversely, suppose that there exists a $\phi \in \Phi^\star(t)$. Recall that W_i was assumed to be a subvector (or more generally, a function) of (P_i, M_i, X_i), and denote this function by ω, so that $W_i = \omega(P_i, M_i, X_i)$. We will show that

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

is an element of $\mathcal{F}^\star$ that satisfies $\theta(\tilde{f}) = t$. (Note that the definition of $\mathbb{V}^\bullet(w)$ ensures that the denominator of each summand is non-zero.) Intuitively, the function $\tilde{f}(\cdot|p,m,x)$ distributes total mass of $\phi(\mathcal{V}|p,m,x)$ uniformly over each set $\mathcal{V} \in \mathbb{V}^\bullet(\omega(p,m,x))$.

For all $\mathcal{V} \in \mathbb{V}$,

$$\begin{aligned}\int_{\mathcal{V}} \tilde{f}(v|p,m,x)\, dv &\equiv \sum_{\mathcal{V}'\in\mathbb{V}^\bullet(\omega(p,m,x))} \int_{\mathcal{V}} \frac{\mathbb{1}\left[v \in \mathcal{V}' \cap \mathcal{V}^\bullet(\omega(p,m,x))\right]}{\lambda\left(\mathcal{V}' \cap \mathcal{V}^\bullet(\omega(p,m,x))\right)} \phi(\mathcal{V}'|p,m,x)\, dv \\ &= \mathbb{1}[\mathcal{V} \in \mathbb{V}^\bullet(\omega(p,m,x))]\phi(\mathcal{V}|p,m,x), \end{aligned} \tag{S11}$$

since the sets in $\mathbb{V}$ and thus $\mathbb{V}^\bullet(\omega(p,m,x))$ are disjoint (almost surely). Using (S11), we have that

$$\int_{\mathbb{R}^J} \tilde{f}(v|p,m,x)\, dv = \sum_{\mathcal{V}\in\mathbb{V}} \int_{\mathcal{V}} \tilde{f}(v|p,m,x)\, dv = \sum_{\mathcal{V}\in\mathbb{V}^\bullet(\omega(p,m,x))} \phi(\mathcal{V}|p,m,x) = 1, \tag{S12}$$

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

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

and every $\phi \in \Phi$ satisfies

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

Since $\tilde{f}$ inherits non-negativity from $\phi \in \Phi^\star \subseteq \Phi$, we conclude from (S12) that $\tilde{f}$ is a conditional density, i.e. $\tilde{f} \in \mathcal{F}$.

To see that $\tilde{f}$ satisfies the upper bound of Assumption IV, notice that

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

where the inequality uses (IV'), and the final equality reverses the steps of the first three. An analogous argument can be used to show that $\tilde{f}$ also satisfies the lower bound of Assumption IV. Assumption SP is satisfied because

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

where the second equality uses Tonelli's Theorem with (S11). That $\tilde{f}$ satisfies (MD) follows from the definition of $\mathbb{V}_j(p)$, (S11), and (MD') via

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

for all $j \in \mathcal{J}$ and $(p, m, x) \in \text{supp}(P_i, M_i, X_i)$. The second-to-last equality here used the implication of (S12) that $\phi(\mathcal{V}|p, m, x) = 0$ for any $\mathcal{V} \notin \mathbb{V}^\bullet(\omega(p, m, x))$. Finally, note that in the notation of (12), (S11) says

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

This equality implies that $\overline{\phi}(\tilde{f})(\mathcal{V}|p, m, x) = \phi(\mathcal{V}|p, m, x)$ for all $\mathcal{V}$, since (S12) implies that $\phi(\mathcal{V}|p, m, x) = 0$ for $\mathcal{V} \notin \mathbb{V}^\bullet(\omega(p, m, x))$. As a consequence,

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

and therefore $t \in \Theta^\star$. *Q.E.D.*

S4.2 Proof of Proposition 2

Observe that Φ is a compact and connected subset of $\mathbb{R}^{d_\phi}$. Since (MD'), (IV'), and (SP') are linear (in)equalities, the subset of Φ that satisfies them is also compact and connected. If $\overline{\theta}$ is continuous on this subset then its image over it—which Proposition 1 established to be $\Theta^\star$—is compact and connected as well. If $d_\theta = 1$, then $\Theta^\star$ is a compact interval, so by definition its endpoints are $t^\star_{\text{lb}}$ and $t^\star_{\text{ub}}$. *Q.E.D.*

S5 Using our Methodology to Expand on a Simulation in Chesher et al. (2013)

Chesher et al. (2013, Section 4.2) consider the following discrete choice model:

$$Y_i = \underset{j \in \{0,1,2\}}{\arg\max}\; \beta_{0j} + \beta_{1j} X_i + V_{ij}, \tag{S13}$$

where V_{ij} are distributed i.i.d. type I extreme value. The coefficients for $j = 0$ are normalized to $\beta_{00} = \beta_{10} = 0$. The coefficients for the other choices are set to $\beta_{01} = \beta_{02} = 0$, $\beta_{11} = 1$, and $\beta_{12} = -0.5$. Chesher et al. (2013) compute identified sets for the counterfactual choice probabilities

$$\wp(x, j) \equiv \mathbb{P}\left[j = \underset{k\in\{0,1,2\}}{\arg\max}\, \beta_{0k} + \beta_{1k}x + V_{ik}\right] \tag{S14}$$

under the assumption that there is an instrument $Z_i \in \{-1, 1\}$ that is independent of $V_i \equiv (V_{i0}, V_{i1}, V_{i2})$. We consider the top-left panel of Figure 4 in Chesher et al. (2013), which corresponds to a case where $X_i \in \{-1, 1\}$. We generate X_i to be dependent with both V_i and Z_i exactly as in (Chesher et al., 2013, pp. 177-178), considering both a "weak" and "strong" instrument case.

Chesher et al. (2013) observe that even if the researcher correctly assumes that V_{ij} are type I extreme value, counterfactual choice probabilities $\wp(x, j)$ will still generally be partially identified because X_i and V_i are dependent. They then compute joint sharp identified sets for $(\wp(x, 1), \wp(x, 2))$ under the parametric assumption that V_{ij} are type I extreme value. We have reproduced these sets in Figure S1 using dashed lines.

To compute these parametric identified sets we used a slight modification of Proposition 2 together with the semiparametric extension proposed in Appendix S2. We created a 40^4 grid of potential values for $\beta \equiv (\beta_{01}, \beta_{02}, \beta_{11}, \beta_{12})$. For each value in this grid, we created "prices" $\tilde{P}_{ij}(\beta) \equiv -\beta_{0j} - \beta_{1j}X_i$ and the MRP implied by these prices. Then we checked whether there exists a $\phi \in \Phi^\star$ (using the characterization in Proposition 1) that also satisfies the additional parametric constraint that the mass in each set of the MRP is equal to the logistic probabilities implied by V_{ij} being type I extreme value. If this linear system of equations had a solution, then we added the values of $(\wp(x, 1), \wp(x, 2))$ implied by this value of β to the parametric identified set.[5]

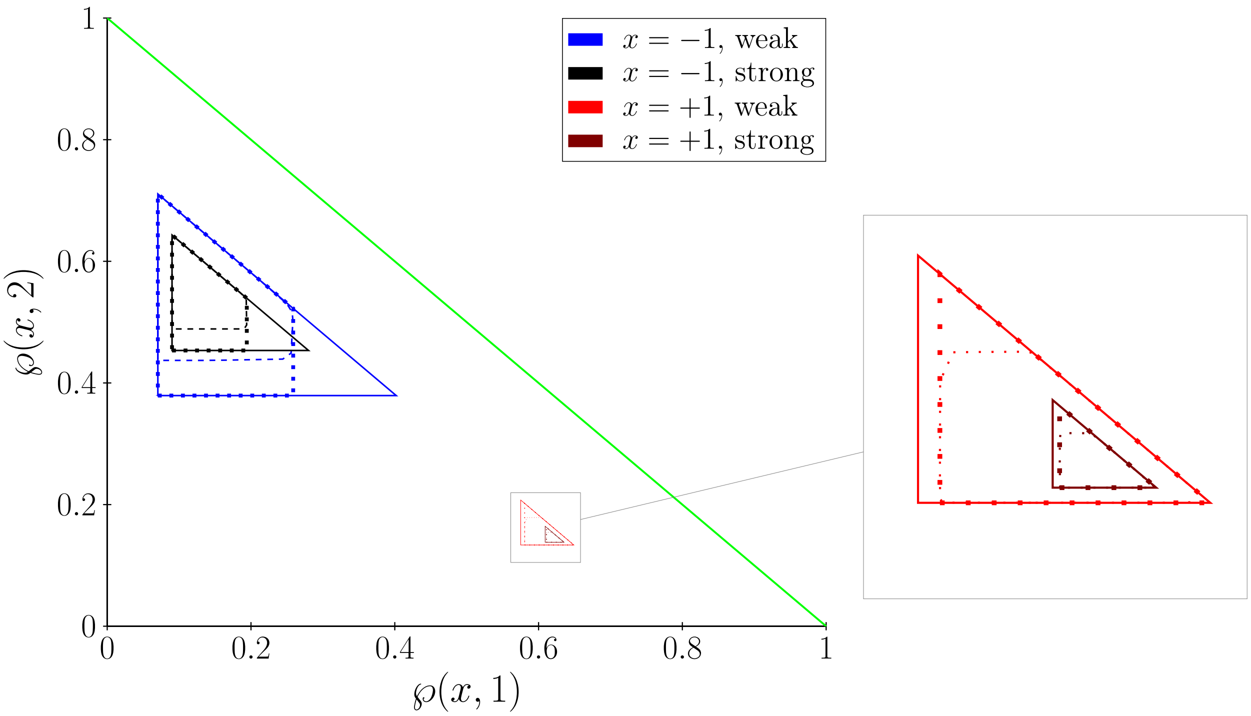
Figure S1 also reproduces the *outer* (non-sharp) nonparametric identified sets reported in Figure 4 of Chesher et al. (2013).[6] As Chesher et al. (2013, pg. 160) clearly acknowledge, these bounds do not exploit the utility maximization structure of the choice model, and thus are not sharp when assuming that choices are generated by (1).

In addition to the two types of sets reported by Chesher et al. (2013), we also compute our

[5] It is straightforward to modify the argument in Proposition 1 to show that this procedure is also sharp, just like the procedure in Chesher et al. (2013). As with any grid-based approach, sharpness in practice requires the grid used in practice to be sufficiently fine.

[6] This is the "first outer region" in equation (1.4) of Chesher et al. (2013).

Figure S1: *Replication and extension of the upper-left panel of Figure 4 in [Chesher et al. (2013)]*



Notes: *The figure shows joint identified sets for counterfactual choice probabilities* $(\wp(x,1), \wp(x,2))$ *evaluated at* $x = -1$ *and* $x = +1$ *in two designs (weak and strong). The sets in the solid lines are the nonparametric* non-sharp *identified sets reported in Chesher et al. (2013). The sets in the dashed lines are the* parametric *sharp identified sets reported in Chesher et al. (2013). The sets in the square markers are nonparametric, sharp identified sets computed using the methodology developed in this paper.*

sharp nonparametric identified sets. To compute these, we first absorb the choice–specific constants β_{0j} into V_{ij}, since the location of V_{ij} (for $j = 1, 2$) is not restricted in our approach. We also normalized (β_{11}, β_{12}) to be $(\tilde{\beta}_{11}, \tilde{\beta}_{12}) \in \mathcal{B} \equiv \{-2, -1, 0, 1, 2\}^2$, again because the scales of V_{i1} and V_{i2} are not restricted in our approach. Then we applied Proposition 2 together with the semiparametric extension proposed in Appendix S2. For each normalized choice of $(\tilde{\beta}_{11}, \tilde{\beta}_{12}) \in \mathcal{B}$, we constructed the MRP with "prices" $-\tilde{\beta}_{1j} X_i$, and then computed the joint identified set for $(\wp(x,1), \wp(x,2))$.[7] The sharp nonparametric joint identified set for $(\wp(x,1), \wp(x,2))$ is then the union of these nine sets.

Figure S1 shows our sharp nonparametric identified sets using square markers. Our sharp nonparametric identified sets are strictly contained in the non-sharp nonparametric identified sets reported by Chesher et al. (2013). Our sharp nonparametric identified sets

[7] As in Figures 5 and S2, we did this by first finding bounds for $\wp(x,1)$, then computing bounds on $\wp(x,2)$ while constraining $\wp(x,1)$ at each point in its marginal identified set.

also strictly contain the sharp parametric identified sets reported by Chesher et al. (2013). Both of these findings make sense. The non-sharp nonparametric identified sets reported by Chesher et al. (2013) are not sharp because they do not exploit the structure of the choice model, unlike our sharp nonparametric identified sets. The sharp parametric identified sets reported by Chesher et al. (2013) reflect an additional parametric assumption not maintained when computing our sharp nonparametric identified sets.

S6 Implementing Bounds on Consumer Surplus

In this section, we show how to construct the $\overline{\theta}$ function for average consumer surplus. Suppose that $\mathbb{V}$ is the MRP constructed from a set of premiums $\mathcal{P}$ that contains the two premiums, p and $p^\star$, at which average consumer surplus is to be contrasted. Let

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

denote average consumer surplus at premium $p^\star$, conditional on $(M_i, X_i) = (m, x)$ under valuation density f. Then

$$\text{CS}^{p^\star}(m, x; f) = \sum_{\mathcal{V} \in \mathbb{V}} \int_{\mathcal{V}} \left\{ \max_{j \in \mathcal{J}} v_j - p_j^\star \right\} f(v|m, x)\, dv, \tag{S15}$$

since the MRP is an (almost sure) partition of $\mathbb{R}^J$. By definition of the MRP, the optimal choice of plan is constant as a function of v within any MRP set $\mathcal{V}$. That is, using the notation in Definition MRP, $\arg\max_{j \in \mathcal{J}} v_j - p_j \equiv Y(v, p) = Y(v', p) \equiv Y(\mathcal{V}, p)$ for all $v, v' \in \mathcal{V}$ and any $p \in \mathcal{P}$. Consequently, we can write (S15) as

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

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

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

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

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

Since each MRP set $\mathcal{V}$ is polyhedral, these quantities are the optimal values of small linear programs that can be computed in an initial step. Because we do not restrict the distribution of valuations within each MRP set, a lower bound on a change in average consumer surplus is attained when this distribution concentrates all of its mass on $v_{\text{lb}}^{p\to p^\star}(\mathcal{V})$ in every $\mathcal{V} \in \mathbb{V}$. That is,

$$\begin{aligned}\Delta\text{CS}^{p\to p^\star}(m,x;f) &\geq \sum_{\mathcal{V}\in\mathbb{V}} p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)} + v_{\text{lb}}^{p\to p^\star}(\mathcal{V}) \int_{\mathcal{V}} f(v|m,x)\,dv \qquad \text{(S16)}\\ &= \sum_{\mathcal{V}\in\mathbb{V}} p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)} + v_{\text{lb}}^{p\to p^\star}(\mathcal{V}) \left[\overline{\phi}(f)(\mathcal{V}|m,x)\right] \equiv \Delta\text{CS}_{\text{lb}}^{p\to p^\star}(m,x;f).\end{aligned}$$

Similarly, an upper bound for any f is given by

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

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

$$\overline{\theta}(\phi) \equiv \sum_{\mathcal{V}\in\mathbb{V}} p_{Y(\mathcal{V},p)} - p^\star_{Y(\mathcal{V},p^\star)} + v_{\text{lb}}^{p\to p^\star}(\mathcal{V})\phi(\mathcal{V}|m,x), \qquad \text{(S17)}$$

and applying Propositions 1 or 2. The requirement that $\theta(f) = \overline{\theta}(\overline{\phi}(f))$ can be seen to be satisfied here by comparing (S16) and (S17). The upper bound is found analogously. The open interval formed by the lower and upper bounds is the sharp identified set.[8]

[8] It is an open interval instead of the closed interval in Proposition 2 because distributions that put a point mass on $v_{\text{lb}}^{p\to p^\star}(\mathcal{V})$ are not continuously distributed. It is straightforward, however, to construct continuous densities that concentrate arbitrarily closely around $v_{\text{lb}}^{p\to p^\star}(\mathcal{V})$ and $v_{\text{ub}}^{p\to p^\star}(\mathcal{V})$, for example by focusing on an $\epsilon > 0$ ball around these points.

S7 Implementing Bounds on Elasticities

In this section, we show how to estimate bounds on finite approximations to semi-elasticities of demand. These bounds can then be transformed into bounds on elasticities after normalizing by a baseline price.

The semi-elasticity of demand for good j with respect to good k in region m for buyers with characteristics x is approximately

$$\text{SElast}^{\delta}_{jk}(p, m, x; f) \equiv 100 \times \frac{1}{\delta} \left(\frac{s_j(p + \delta e_k, m, x; f) - s_j(p, m, x; f)}{s(p, m, x; f)} \right), \tag{S18}$$

where δ is a price change and e_k is a $(J + 1)$–dimensional vector with 1 in the kth place and zeros elsewhere. Condition TP is satisfied as long as the MRP contains both p and $p + \delta e_k$, in which case the corresponding $\overline{\theta}$ function is given by

$$\overline{\text{SElast}}^{\delta}_{jk}(p, m, x; \phi) \equiv 100 \times \frac{1}{\delta} \left(\frac{\sum_{\mathcal{V} \in \mathbb{V}_j(p+\delta e_k)} \phi(\mathcal{V}|p, m, x) - \sum_{\mathcal{V} \in \mathbb{V}_j(p)} \phi(\mathcal{V}|p, m, x)}{\sum_{\mathcal{V} \in \mathbb{V}_j(p)} \phi(\mathcal{V}|p, m, x)} \right). \tag{S19}$$

While $\overline{\text{SElast}}^{\delta}_{jk}(p, m, x; \phi)$ is a nonlinear function of ϕ, it is the ratio of two linear functions of ϕ. Optimization problem (14) (and the estimation counterpart (17)) thus becomes a linear-fractional program. The celebrated Charnes and Cooper (1962) transformation can be used to produce an equivalent linear program, see e.g. Boyd and Vandenberghe (2004, pg. 151) for a textbook discussion. Kamat (2020) has previously used the Charnes and Cooper (1962) transformation to bound conditional treatment effects in an instrumental variables model with discrete treatments.

In order for the linear fractional program to be well-posed, we need to ensure that $s(p, m, x; f)$ is bounded away from zero over the feasible region, so that $\text{SElast}^{\delta}_{jk}(p, m, x; f)$ remains well-defined over the feasible region. This requirement comes out of the nonparametric nature of the model, which allows for zero choice shares (vs. logit-based models), but it is quite intuitive: if a zero choice share is compatible with the data and assumptions then so too is *any* semi-elasticity of that choice.

In the application we keep the denominator bounded away from zero by changing focus in two ways. First, we group Silver, Gold, and Platinum together into a single "low-deductible" category, which helps prevents zero denominators from arising in the relatively less popular Gold and Platinum plans. Second, we consider a version of (S18) that is aggregated over demographic bins within a region:

$$\begin{aligned}&\text{SElast}^{\delta}_{jk}(m;f) \qquad\qquad \text{(S20)}\\ &\equiv 100 \times \frac{1}{\delta}\left(\frac{\sum_x \mathbb{P}[X_i = x|M_i = m]\left(s_j(\pi(m,x)+\delta e_k, m, x; f) - s_j(\pi(m,x), m, x; f)\right)}{\sum_x \mathbb{P}[X_i = x|M_i = m] s(\pi(m,x), m, x; f)}\right),\end{aligned}$$

where $\pi(m,x)$ is the premium function introduced in Section 3.2. To aggregate these region-level semi-elasticities into a single elasticity measure, we first normalize by the average premium paid for the product bundle in the region. We then report the average elasticity across regions.

S8 Statistical Inference Implementation Details

In this section we provide details on how we implement the testing procedure developed by Deb, Kitamura, Quah, and Stoye (2021, "DKQS") in our application.

The null hypothesis of the test is $H_0 : t \in \Theta^\star$, i.e. that the conjectured value t is in the sharp identified set for the target parameter. The test statistic is defined as

$$\begin{aligned}\text{TS}(t) \equiv \min_{\phi \in \Phi} \sum_{j,p,m,x} n w(p,m,x)\left(\hat{s}_j(p,m,x) - \sum_{\mathcal{V} \in \mathbb{V}_j(p)} \phi(\mathcal{V}|p,m,x)\right)^2\\ \text{subject to (IV'), (SP'), and } \overline{\theta}(\phi) = t. \qquad \text{(S21)}\end{aligned}$$

where $w(p,m,x) > 0$ is a weight, and n is the sample size. Notice that if $\hat{s}_j(p,m,x) = s_j(p,m,x)$ without error, then $\text{TS}(t) = 0$ if and only if there exists a $\phi \in \Phi^\star(t)$, i.e. if and only if $t \in \Theta^\star$. In our application, p is a deterministic function of m and x (see Section 3.1), in which case the dependence of w, $\hat{s}_j$, and ϕ on p is redundant. We take the weight $w(p,m,x) = w(m,x)$ to be proportional to the size of bin (m,x), so that larger bins receive greater weight, the same as in our estimator. We also note that (IV') and (SP') are simple equality constraints in our application, so they can be substituted out with appropriate redefinition of the parameter ϕ. After the substitution, the redefined parameter is only constrained to lie in the simplex. We directly make this substitution when applying the test, but we leave it implicit here (and throughout) for notational simplicity.

Computing a critical value involves solving a "tightened" version of (S21). Defining the tightened version requires some notation. First, since the DKQS test requires the target parameter to be linear, we abuse notation slightly and write the function $\overline{\theta}$ as a vector:

$\overline{\theta}(\phi) \equiv \phi'\overline{\theta}$. Then let

$$\theta_{\max} \equiv \max_{\phi \in \Phi} \phi'\overline{\theta} \quad \text{subject to (IV') and (SP')}, \tag{S22}$$

and define $\theta_{\min}$ to be the optimal value for the corresponding minimization problem. The set $[\theta_{\min}, \theta_{\max}]$ constitutes the range of values that the target parameter could logically take under the maintained assumptions, before confronting the data. Then define the sets of integers

$$\mathcal{I}_{\max} \equiv \left\{ i = 1, \dots, d_\phi : (\overline{\theta})_i = \theta_{\max} \right\} \quad \text{and} \quad \mathcal{I}_{\min} \equiv \left\{ i = 1, \dots, d_\phi : (\overline{\theta})_i = \theta_{\min} \right\}, \tag{S23}$$

where $(\overline{\theta})_i$ is the ith component of the vector $\overline{\theta}$, and let $\mathcal{I}_0 \equiv \{1, \dots, d_\phi\} \setminus (\mathcal{I}_{\max} \cup \mathcal{I}_{\min})$ be all the rest of the integers. The tightened version of (S21) is defined as

$$\begin{aligned} \mathrm{TS}(t;\tau) \equiv \min_{\phi \in \Phi} & \sum_{j,p,m,x} nw(p,m,x) \left(\hat{s}_j(p,m,x) - \sum_{\mathcal{V} \in \mathbb{V}_j(p)} \phi(\mathcal{V}|p,m,x) \right)^2 \\ & \text{subject to (IV'), (SP')}, \ \overline{\theta}'\phi = t, \\ & \quad \text{and } \phi_i \geq \tau \frac{(\theta_{\max} - t)}{|\mathcal{I}_{\min} \cup \mathcal{I}_0|} \text{ for all } i \in \mathcal{I}_{\min}, \\ & \quad \phi_i \geq \tau \frac{(t - \theta_{\min})}{|\mathcal{I}_{\max} \cup \mathcal{I}_0|} \text{ for all } i \in \mathcal{I}_{\max}, \\ & \quad \phi_i \geq \frac{\tau}{|\mathcal{I}_0|} \left(1 - \frac{(\theta_{\max} - t)|\mathcal{I}_{\min}|}{|\mathcal{I}_{\min} \cup \mathcal{I}_0|} - \frac{(t - \theta_{\min})|\mathcal{I}_{\max}|}{|\mathcal{I}_{\max} \cup \mathcal{I}_0|} \right) \text{ for all } i \in \mathcal{I}_0, \end{aligned} \tag{S24}$$

where $|\cdot|$ when applied to a set denotes cardinality, and $\tau \geq 0$ is a tuning parameter.

We solve the tightened problem (S24) once exactly as stated, and let $\hat{\phi}^\star$ be any optimal solution. Then we solve it again in each of B bootstrap replications. In replication b, we nonparametrically redraw choices and compute bootstrapped choice shares $\hat{s}_j^b(p,m,x) = \hat{s}_j^b(m,x)$ for each bin (m,x). Then we compute what DKQS refer to as "τ–tightened" recentered bootstrap estimators

$$\tilde{s}_j^b(m,x) = \hat{s}_j^b(m,x) - \hat{s}_j(m,x) + \sum_{\mathcal{V} \in \mathbb{V}_j(p)} \hat{\phi}^\star(\mathcal{V}|m,x). \tag{S25}$$

We solve (S24) with $\tilde{s}_j^b(m,x)$ in place of $\hat{s}_j^b(m,x)$, and let $\mathrm{TS}^b(t;\tau)$ denote the resulting optimal value. Once we have completed this B times, we find the .95 quantile of $\{\mathrm{TS}^b(t;\tau)\}_{b=1}^B$

Table S1: *Monte Carlo results*

τ	$1 - \Delta\text{Share}_0^\delta$	ΔCS^δ	ΔGS^δ
0.0025	0.930	0.850	0.845
0.005	0.985	0.970	0.990
0.01	1.000	1.000	1.000

Notes: *Proportion of 200 simulation draws in which a 95% confidence interval contained the population identified set for the specified target parameter.*

(for a level 5% test), and reject the null hypothesis $t \in \Theta^\star$ if the test statistic $\text{TS}(t;\tau)$ exceeds that quantile.

The choice of tuning parameter τ is important. Since $\text{TS}(t;\tau) \geq \text{TS}(t)$ for all τ, the likelihood of rejecting the null hypothesis decreases monotonically with τ. When $\tau = 0$, the test reduces to simply bootstrapping the test statistic, which we would not expect to control size due to the inequality constraints (see e.g. Andrews and Han, 2009).

To pick τ, we conducted a Monte Carlo simulation based on our data. We fit the simplest comparison logit model (see Section 3.4) to data from rating region 16, which covers part of Los Angeles, and is the largest region, comprising roughly 20% of potential buyers. Then we redraw data from the fitted logit model and conduct 5% tests at the endpoints of the nonparametric bounds for our three main target parameters: changes in probability of purchasing coverage, change in consumer surplus, and change in government spending, all in response to a $10 decrease in subsidies.

Table S1 reports the results, which are based on 200 draws, each with 100 bootstrap replications, the same as in the application. We find that the test produces confidence intervals with adequate coverage for $\tau = .005$, and at $\tau = .01$ the test always covers the population identified set. Since the Monte Carlo uses a smaller sample size than in our application, we decided to be extra conservative and use $\tau = .125$ in our reported results, which we found still produced acceptably short confidence intervals. We expect that our reported confidence intervals over-cover, potentially by a wide margin.

Constructing confidence intervals using the DKQS test is computationally challenging in our application. When computing bounds, we are able to leverage our empirical strategy of not using cross-region variation to separate the original program with all regions into separate programs for each region, which greatly speeds up computation and reduces memory usage. The null hypothesis constraint $\overline{\theta}'\phi = t$ in (S24) prevents us from using the same strategy for computing $\text{TS}(t;\tau)$, since the evaluation of $\overline{\theta}'\phi$ depends on all regions simultaneously.

Consequently, (S24) needs to be solved using data from all regions simultaneously. Together with bootstrapping and test inversion, this becomes a computationally demanding task.

S9 Estimation of Potential Buyers

We estimate the number of potential buyers using the California 2013 3-year subsample of the American Community Survey (ACS) public use file, downloaded from IPUMS (Ruggles et al., 2015). We use estimated potential buyers to turn the administrative data on quantities purchased into choice shares.

We define an individual i in the ACS as a potential buyer, denoted by the indicator $I_i = 1$, if they report being either uninsured or privately insured. Individuals with $I_i = 0$ include those who are covered by employer-sponsored plans, Medi-Cal (Medicaid), Medicare, or other types of public insurance. We estimate $\mathbb{P}[I_i = 1|M_i = m, X_i = x]$ and convert estimated probabilities into estimated number of potential buyers in each (m, x) pair by using the individual sampling weights provided in the ACS. To avoid excessive extrapolation, we drop 7,455 bins that are empty in the ACS.

The estimated probabilities are constructed using flexible linear regression. The main regressors are the X_i bins, that is, age in years and income in FPL (taken at the lower endpoint of the bin). We include a full set of interactions between these variables and indicators for the coarse age and income bins described in Section 3.3 (called W_i there). We also include a full set of region indicators (M_i), and interactions between these indicators and both age and income. For 62 bins, we estimate fewer potential buyers than there are actual buyers in the administrative data. We drop these bins, all of which are small.

An adjustment to this procedure is needed to account for the fact that the PUMA (public use micro area) geographic identifier in the ACS can be split across multiple counties, and so in some cases also multiple ACA rating regions. For a PUMA that is split in such a way, we allocate individuals to each rating region it overlaps using the population of the zipcodes in the PUMA as weights. This is the same adjustment factor used in the PUMA-to-county crosswalk.[9] Since the definition of a PUMA changed after 2011, we also use this adjustment scheme to convert the 2011 PUMA definitions to 2012–2013 definitions.

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

Table S2: *Region groups after clustering*

Group	Regions	Population (count)	Average Income ($)	Inpatient Days (per capita)	Hospital Spending (per capita)	Payroll Hosp. Spending (per capita)	Share in Poverty (0,1)	Share Uninsured (0,1)
A	18,19	3097907	47745	0.511	1692	682	0.138	0.168
B	4,14	832268	49900	0.845	3390	1395	0.186	0.162
C	8,9	734739	55344	0.514	1902	822	0.116	0.150
D	6,12	1530247	45758	0.623	1678	718	0.129	0.145
E	2,11	1273650	40067	0.726	1674	662	0.179	0.158
F	1,5	1211352	44651	0.402	1958	782	0.152	0.139
G	7,10	1875728	43911	0.553	2240	932	0.163	0.150
H	3,13,17	2230359	32805	0.449	1624	677	0.186	0.172
I	15,16	9889056	41791	0.658	2031	841	0.181	0.227

S10 Cross-Region Strategy

In this section, we consider an alternative strategy that uses cross-region variation to replace age variation.

The motivation for the strategy is as follows. Since the premiums are calculated from base prices following a fixed formula, insurers set base prices for a region taking into consideration its composition of potential buyers. Differences in the age composition of potential buyers mean that two individual buyers of the same age and income, but different regions, will face different post-subsidy premiums. If the different regions are otherwise comparable, then it may be reasonable to assume that these two buyers have similar preferences. This argument has been used previously in Ericson and Starc (2015); Tebaldi (2017); Orsini and Tebaldi (2017); it has the flavor of a "Waldfogel instrument" (Waldfogel, 1999).

To implement the strategy we first group the 19 Covered California rating regions in 9 separate clusters. We define the clusters based on their similarity along the vector of 7 observables *not including* the age distribution: total population, average income, hospitalizations per capita, annual hospital spending per capita, payroll hospital spending per capita, share of people in poverty, and share of under-65 who did not have health insurance before the ACA.[10] The two Los Angeles regions are grouped together, while the remaining regions are assigned to 8 different groups using single-linkage hierarchical clustering. The 9 groups are summarized in Table S2.

In the notation of Section 2, we now have W_i representing all combinations of one-year age bins, coarse FPL bin ($\{140\text{–}150, 150\text{–}200, \ldots, 350\text{–}400\}$), and region group among the

[10] The data comes from the county-level Area Health Resource Files, available at https://data.hrsa.gov/topics/health-workforce/ahrf.

Table S3: *Nonparametric bounds on changes in choice shares — cross-region strategy*

	Change in probability of choosing									
$10/month premium	Any plan		Bronze		Silver		Gold		Platinum	
increase for	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
	Panel (a): Full sample (140 - 400% FPL)									
All plans	-0.068	-0.018	-0.015	-0.004	-0.048	-0.012	-0.004	-0.001	-0.004	-0.001
Bronze	-0.015	-0.002	-0.046	-0.006	+0.001	+0.033	+0.000	+0.036	+0.000	+0.033
Silver	-0.047	-0.004	+0.000	+0.062	-0.140	-0.014	+0.000	+0.106	+0.000	+0.080
Gold	-0.003	-0.000	+0.000	+0.008	+0.000	+0.009	-0.013	-0.002	+0.000	+0.011
Platinum	-0.003	-0.000	+0.000	+0.007	+0.000	+0.006	+0.000	+0.009	-0.011	-0.001
	Panel (b): Lower income (140 - 250% FPL)									
All plans	-0.089	-0.021	-0.014	-0.003	-0.071	-0.016	-0.004	-0.001	-0.004	-0.001
Bronze	-0.013	-0.001	-0.045	-0.006	+0.001	+0.032	+0.000	+0.034	+0.000	+0.031
Silver	-0.070	-0.005	+0.000	+0.084	-0.199	-0.019	+0.000	+0.148	+0.000	+0.112
Gold	-0.003	-0.000	+0.000	+0.007	+0.000	+0.008	-0.012	-0.001	+0.000	+0.010
Platinum	-0.004	-0.000	+0.000	+0.007	+0.000	+0.006	+0.000	+0.009	-0.011	-0.001
	Panel (c): Higher income (250 - 400% FPL)									
All plans	-0.041	-0.015	-0.017	-0.005	-0.019	-0.007	-0.005	-0.001	-0.003	-0.001
Bronze	-0.016	-0.002	-0.048	-0.006	+0.000	+0.035	+0.000	+0.038	+0.000	+0.036
Silver	-0.018	-0.001	+0.000	+0.034	-0.064	-0.008	+0.000	+0.053	+0.000	+0.039
Gold	-0.004	-0.000	+0.000	+0.009	+0.000	+0.012	-0.015	-0.003	+0.000	+0.012
Platinum	-0.003	-0.000	+0.000	+0.006	+0.000	+0.006	+0.000	+0.008	-0.010	-0.001

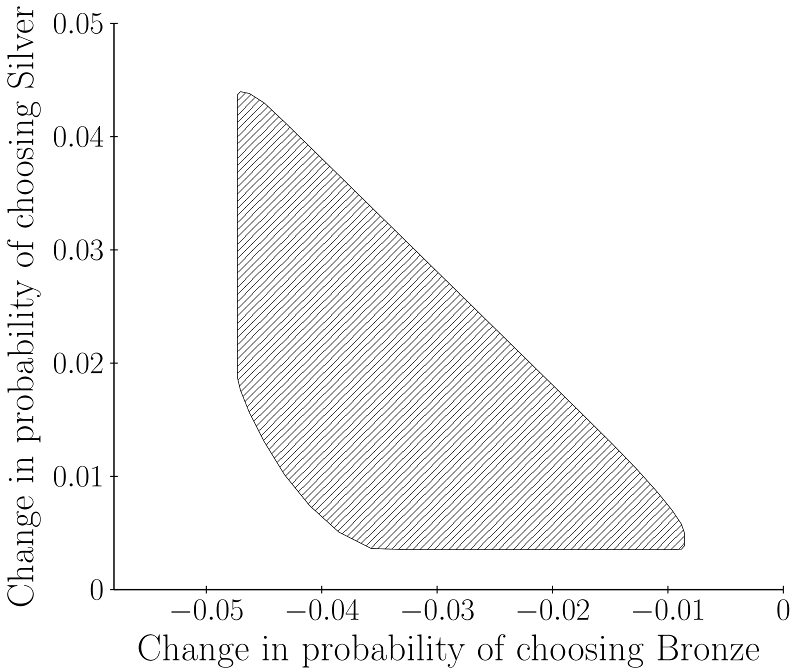
9 in Table S2. The instrument Z_i is then all bins formed by crossing a region indicator with 5% FPL bins. For example, one value of W_i corresponds to individuals who are aged 36 with incomes between 150% and 200% of the FPL and live in region group D (region 6 or 12). Within this bin, we have 20 values of Z_i, comprised of the 10 income bins crossed with the two geographic regions, and for each value we observe a different premium vector while assuming that the distribution of valuations is the same.

Table S3 reports estimated bounds on changes in choice shares, the same as Table 3 for our preferred strategy. The extensive margin responses to an increase in all premiums are nearly identical to those from our main strategy. We interpret this as corroborating our finding in Section 4.2 that our results are primarily driven by variation in income, rather than in age. Bounds on changes in consumer surplus and government expenditure (not shown) are also nearly identical to those reported in Section 5.1.

We do however see more differences in cross-tier substitution patterns. For example, using the cross-region strategy we estimate an increase in Bronze premiums by $10 would lead to an increase in the share choosing Silver of between 0.1–3.3%, versus 0.4–4.4% in our preferred strategy. As another example, the cross-region strategy tightens the upper bound on the choice share choosing Bronze when the Silver premium to 6.2%, from 12.4% in our preferred strategy.

S11 Additional Figures and Tables

Figure S2: *Effect of increasing bronze premiums by $10 on Bronze and Silver choice shares*



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

References

Andrews, D. W. K. and S. Han (2009): "Invalidity of the Bootstrap and the m out of n Bootstrap for Confidence Interval Endpoints Defined by Moment Inequalities," *Econometrics Journal*, 12, S172–S199. 18

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890. 3

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

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

Boyd, S. and L. Vandenberghe (2004): *Convex Optimization*, Cambridge university press. 15

Charnes, A. and W. W. Cooper (1962): "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, 9, 181–186. 15

Chesher, A., A. M. Rosen, and K. Smolinski (2013): "An instrumental variable model of multiple discrete choice," *Quantitative Economics*, 4, 157–196. 4, 10, 11, 12, 13

Deb, R., Y. Kitamura, J. K.-H. Quah, and J. Stoye (2021): "Revealed Price Preference: Theory and Empirical Analysis," *arXiv:1801.02702 [econ]*. 16

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. 20

Handel, B., I. Hendel, and M. D. Whinston (2015): "Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk," *Econometrica*, 83, 1261–1313. 1

Handel, B. R. (2013): "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103, 2643–82. 1

Kamat, V. (2020): "Identification of Program Access Effects with an Application to Head Start," *arXiv:1711.02048 [econ]*. 15

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

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*. 19

Shorack, G. R. (2000): *Probability for Statisticians*, Springer-Verlag. 7

Tebaldi, P. (2017): "Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA," *Working paper*. 20

Waldfogel, J. (1999): "Preference externalities: An empirical study of who benefits whom in differentiated product markets," Tech. rep., National Bureau of Economic Research. 20