

Supplemental Appendix for Online Publication

Procurement Auctions vs. Regulated Competition in Selection Markets

José Ignacio Cuesta and Pietro Tebaldi

S.1 Details on Data Construction

S.1.1 Provider Networks

To construct measures of provider networks offered by each plan, we leverage detailed data on the universe of providers that each plan offers. Covered California mandates every qualified health plan to provide monthly information on all PCPs, specialists, hospitals, and other individual and facility providers covered by them. For each physician and facility, the data include the name, the national provider identification number (NPI), specialty, zipcode, and the plans it has contracted with. For our analysis, we use the second quarter of each year. There is minimal variation in provider networks reported across quarters within a year. In addition, the quality of the data for 2014 is low, and hence, we impute the data from 2015 to 2014. In addition, we focus on the networks of PCPs and hospitals. We leverage the fact that we observe the location of both households and providers at the zipcode level to construct measures of provider networks offered by each plan in each zipcode. It is possible in the data for the same provider to have multiple practice locations within a zip code. In such cases, we treat these locations as one for the same provider within a zip code. In contrast, if the same provider offers multiple locations in different zip codes within a catchment area, we count them as separate providers.

S.1.2 Measuring the Number of Potential Buyers

The enrollment data describe choices among plans in Covered California, but do not provide information on the outside option for this population. Hence, we develop a procedure to measure the set of potential buyers and hence the outside share. This procedure is similar to those adopted in recent work by Tebaldi et al. (2023) and Dickstein et al. (2024).

We start by estimating the likelihood of a person being a potential buyer using data for California from the American Community Survey (ACS). We define potential buyers in Covered California as individuals who are either individually insured or uninsured. We then estimate the

linear regression

$$\text{Potential buyer}_i = f(X_i) + \varepsilon_i,$$

where $f(X_i)$ is a flexible function of individual characteristics. In particular, we first define 8 coarse age groups given by [20-26), [26-31), [31-36), [36-41), [41-46), [46-51), [51-56), [56-61) and [61-64]; and 7 coarse FPL groups given by [100-140), [140-150), [150-200), [200-250), [250-300), [300-350) and [350-400]. The function $f(X_i)$ includes the interactions between indicators for these two sets of variables with continuous age and income as a share of FPL; interactions for indicators for regions and gender with continuous age and income as a share of FPL; interactions between indicators for regions and gender; and interactions between indicators for age and income bins and gender. This regression is weighted using ACS person weights. Let the fitted likelihood of individual i being a potential buyer be τ_i , which takes on the same value for any individual with the same age, FPL, sex, region, and year.

The next step is to use the estimates $\hat{\tau}_i$ to obtain the number of potential buyers among consumers of a particular type. Let d index types and d_i encode the type of an individual as defined by gender and the bins of age and income as a share of FPL defined above. Define $\mathcal{D}_{drt} = \{i : d_i = d, r_i = r, t_i = t\}$ as the set of individuals of type d . We compute the number of potential buyers of type d in region r and year t as

$$NPB_{drt} = \sum_{i \in \mathcal{D}_{drt}} \hat{\tau}_i \cdot w_i,$$

where w_i are person weights in the ACS.

The final step in the procedure is to move from the number of potential buyers of a given type at the region level to the zipcode level. This is not feasible to implement using ACS data only, using population counts are not available from such data at the zipcode level. We complement that data source with population counts at the zipcode level from www.greatdata.com. Using these data, we calculate the zipcode population weights w_z , such that $\sum_{z \in r} w_z = 1 \quad \forall r$, and then calculate the number of potential buyers of a given type in a zipcode as

$$NPB_{dzt} = NPB_{drt} \cdot w_z,$$

which requires the assumption that the distribution of population across zipcodes within a region is unrelated to individual types.

S.1.3 Augmenting the Enrollment Data to Account for the Outside Share

We estimate demand using the individual-level data from the marketplace and a maximum likelihood estimator. In the context of that strategy, we incorporate our measures of market size by augmenting the data to match the outside shares implied by our estimates of the number of potential buyers. We start by calculating the number of potential buyers of type θ_i in a given region and year who choose the outside option by taking the difference between $NPB_{\theta_{zt}}$ and the number of enrollees in Covered California in the enrollment data. Let $N_{\theta_{zt}}^0$ denote this number. For each type θ in a zipcode and year, we then randomly draw $N_{\theta_{zt}}^0$ individuals from the enrollment data, and set their choice to be the outside option. We then append these synthetic enrollees to our dataset and proceed with estimation as described in Section 5.1.