Online Appendix

Vertical Integration and Plan Design in Health Insurance Markets

José Ignacio Cuesta, Carlos Noton, and Benjamín Vatter

A Data and Setting Appendix

This appendix provides additional details about our data, preliminary steps taken to estimate risk and extract negotiated prices, and form choice sets for the plan demand analysis. In addition, we provide details about the Chilean healthcare market and regulatory environment to complement those in the main text.

A.1 Data Construction

A.1.1 Enrollment and Claims Data. Our analysis relies on two main administrative datasets. First, an enrollment dataset covering the universe of private plan policyholders, including plan choices, household composition, and sociodemographic characteristics. Second, a claims dataset documenting each claim processed by private insurers, with details on hospitals, services rendered, charged prices, and reimbursements. We focus on data from 2013–2016. This section details the processing steps transforming the raw datasets into the analysis sample.

The raw enrollment dataset comprises 1,484,897 policyholders across 61,450 distinct plan identifiers from five insurers. Many plans exhibit low enrollment, with a median of 12 policyholders per plan. Two factors drive this pattern. First, guaranteed renewability allows policyholders to retain plans no longer actively marketed; indeed, most plans never appear on the spot market. While our analysis primarily examines spot-market plans, legacy plans are incorporated into demand estimation (see Section 6). Second, insurers frequently offer nearly identical plans under distinct codes to reduce regulatory scrutiny (see Section A.2).

Given these characteristics, we consolidate the original 61,450 plan identifiers by insurer, inpatient and outpatient coverage, primary preferential provider, and deciles of base premiums. This grouping captures essential financial dimensions of plan design and insurer competition relevant to VI. Similar approaches have been adopted in prior research on the Chilean market (Atal, 2019; Dias, 2022). We use claims data to determine preferential providers, leveraging indications of preferential provider status and consumer cost-sharing to infer coverage tiers. We observe claims for nearly all enrolled plans, ensuring minimal information loss from this procedure.

We apply two additional filters to the enrollment data. First, if a policyholder switches plans within a given year, we retain only the plan held longest. Second, we restrict the dataset to policyholders aged 2564. The resulting sample consists of 1,247,125 policyholders and 4,110 plan-years, with 1,431 plans actively offered on the spot market.

The raw claims dataset contains 18,957,306 inpatient service claims processed by the sampled insurers at Santiago hospitals. Claims are aggregated into distinct admission events as identified in the data, each associated with ICD-10 diagnosis codes. Ten percent of admissions (eight percent of total revenue) lack sufficient diagnostic detail. Admissions involving multiple providers are assigned to the hospital delivering the majority of services. Admissions at the markets 11 primary hospitals, comprising 74 percent of events, are used for the main analysis; admissions at smaller private or public hospitals form the outside option. Prices and reimbursements are aggregated within each admission event and linked to the corresponding policyholder and plan.

We further refine the claims dataset by removing admissions lacking matching policyholder records in the enrollment data or adequate diagnostic information. Additionally, we restrict the dataset to 16 major diagnosis categories: infections and parasites, neoplasms, blood diseases, endocrine disorders, nervous system diseases, ocular conditions, ear diseases, circulatory diseases, respiratory diseases, digestive diseases, skin disorders, musculoskeletal diseases, genitourinary diseases, pregnancy, perinatal treatments, and congenital malformations. Finally, admissions involving patients aged 65 or older are excluded. The final claims sample includes 773,264 admission events. .

A.1.2 Diagnosis Risk. Throughout our analysis, we aggregate diagnoses at the ICD-10 chapter level. Using claims data, we calculate each beneficiary's average annual risk of experiencing a medical event by diagnosis group. Specifically, we determine the average number of events per enrollee-diagnosis-year within age-gender categories. Ages between 5 and 65 are grouped into 5-year intervals, with separate bins created for those under age 2 and those between ages 2 and 5.

A.1.3 Negotiated Prices and Resource Intensity Weights. We compute negotiated prices and resource intensity weights based on observed total payments to providers per diagnosis following recent work in the literature (e.g., Gowrisankaran *et al.* 2015; Cooper *et al.* 2018). We regress log total hospital payments per admission on insurer-hospital-year and gender-diagnosis-age fixed effects:

$$\log p_{ihjt}^{\text{Total}} = \log p_{m(j)ht} + \log \omega_{\kappa(i)d(i)} + \varepsilon_{ihjt}$$

where $\log p_{m(j)ht}$ are insurer-hospital-year fixed effects capturing log negotiated prices for an insurer-hospital pair $m - j$ in year t, and $\log \omega_{\kappa(i)d(e)}$ are the log resource intensity weights associated with the type of the patient in admission i, $\kappa(i)$, and the diagnosis $d(i)$ associated with admission i. We normalize weights relative to delivery for a woman aged 25–40. Finally, we recenter the estimated negotiated prices such that the mean predicted total payment equals the observed mean. In estimation, we drop the top 99*th* and bottom 1*st* percentiles of spending to reduce the impact of outliers.

A.1.4 Insurance Plan Choice Sets and Publicly Insured Consumers. To form plan choice sets, we select the top 70 percent most popular insurance plans sold by each insurer in each market segment and year. On average, this captures 27 percent of the plans but 95.2 percent of the enrollment. Each household's choice set includes these popular plans and their previous year's plan, regardless of whether it is currently offered in the spot market, to conform to Chile's guaranteed renewability regulation. For each plan, we compute the expected network surplus for each household member as indicated in the main text. Given the number of diagnoses and available options, this is a lengthy process. To reduce computational costs, we use a random 30 percent sample of households when estimating plan demand.

One of the insurers in our sample operates using two brands. We account for this feature by allowing for heterogeneity in preferences over the services of these brands when estimating demand, but otherwise treat them as a single firm.

To provide a comprehensive depiction of the insurance market, we also integrate data on public insurance enrollees from the 2013, 2015, and 2017 waves of the CASEN survey (CASEN, 2015), as described in the main text. Representative public enrollees are created and matched to corresponding private market segments based on the modal neighborhoods of similar private enrollees, including assigning representative characteristics to dependents. Subsequently, we calculate their expected utility from each private insurance plan. Sampling weights reflecting the number of public enrollees represented by each household are applied in all subsequent estimations.

A.2 Details about Health Insurance in Chile

A.2.1 Public and Private Sector Interaction. The public and private health systems operate in a notably isolated manner: Private enrollees account for 97 percent of private hospital revenue and only 3 percent of public hospital revenue (Galetovic and Sanhueza, 2013). Research on sorting across sectors highlights differences in premiums as the key driver of enrollment decisions (Pardo and Schott, 2012).

During our period of study, private and public insurance enrollment have remained steady at around 18 and 76 percent, respectively. While we incorporate substitution between private and public insurance, substitution across the markets is low (Duarte, 2011). The evidence suggests that the public sector is a safety net for private enrollees, who utilize it primarily when unemployed.

A.2.2 The Regulatory Environment. In May 2005, the government introduced Law 20,015 known as *Ley Larga de Isapres*, imposing several regulations on the private insurance sector. This law limited risk pricing by enforcing risk-rating functions on insurers and capping annual premium increases. It also prohibited direct VI by banning insurers from providing healthcare services.

However, Chile's law code has enabled industry players to find legal ways to risk-price consumers and vertically integrate with hospitals. To bypass the pricing step-function rules, insurers often duplicate their plans, selling different versions to different consumers, thus segmenting the market without significant regulatory constraints. To capture these strategies, we aggregate plans based on coverage, premium, and preferential providers rather than identifiers reported by insurers. Similarly, insurers have circumvented the VI ban by forming holding companies that own both insurers and hospitals. This ability to bypass regulations has been the focus of much political debate over the last decade.

In September 2004, the government enacted Law 19,966, mandating coverage for a list of medical conditions. Effective June 2005, public and private insurers must provide adequate treatment and insurance for these conditions, known as AUGE-GES guarantees. The law ensures access to adequate treatment and requires hospitals to certify their care quality. It also mandates that private plans cover 80 percent and public insurance 100 percent for these conditions. However, private enrollees must file forms and request approval to use these benefits, leading to significantly lower utilization among them (Alvear-Vega and Acuña, 2022). In our analysis, we consider this regulation as an overall shifter of the value of public insurance relative to private insurance and as a determinant of the resource intensity weight multiplier identified in the price regressions.

Regulation also dictates that private plans must offer coverage at least as generous as the public insurance system (Article 190 of DFL #1, 2005). However, the enforcement of this requirement is incomplete, and some plans fall below the 50 percent effective coverage of the public option. Enforcement, however, does occur. For example, in 2008, one of the insurers in our sample was found to have undercut the legal floor and was forced to re-liquidate charges and reimburse nearly a thousand beneficiaries (Superintendencia de Salud de Chile, 2008). Similarly, there have been public complaints by lawmakers due to a lack of enforcement in other coverage and access guarantees (Mostrador, 2013). The usual recourse used by policymakers, regulators, and consumers when regulation is violated is to sue insurers (Oliveira *et al.*, 2021).

Finally, guaranteed renewability regulations mandate indefinite plan durations, requiring insurers to clearly outline and notify consumers of coverage changes within 60 days, with a reciprocal 30-day consumer notice required for exiting contracts. Although renewability theoretically enables permanent plan retention, insurers effectively displace undesirable enrollees by raising premiums.

Due to this imperfect enforcement of regulatory constraints, we model the regulatory environment in plan design as a function to be estimated rather than imposing strict constraints that would not align with the data. The use of litigation as a means to enforce compliance with access regulation justifies our use of disagreement penalties in the bargaining model.

A.2.3 Risk Pricing and Selection. The public insurer primarily differentiates consumers based on income and household size, offering no variation in plans based on other characteristics. In contrast, private insurers provide diverse plans targeting specific demographic groups, such as women without dependents, men without dependents, and families. These targeted groups often explicitly appear in plan names and are imperfectly reported to the regulator, as observed in our data. Although regulations prevent insurers from charging different premiums for identical plans or rejecting consumers based on gender, insurers effectively segment the market by designing plans with gender-specific coverages, notably pregnancy-related services. Thus, insurers circumvent regulatory pricing constraints by offering specialized plans to different age and gender groups. Our analysis captures this strategy by treating each gender and age cohort as distinct market segments.

Regulations permit private insurers to deny coverage based on pre-existing conditions. However, our analysis does not consider this selection margin, as insurers rely on consumer disclosure, information absent from our data. Furthermore, it is unclear whether consumers are incentivized to disclose pre-existing conditions or how effectively insurers enforce compliance. Recent developments post-dating our sample period include a major insurer significantly reducing consideration of pre-existing conditions to attract enrollees, though this initiative had minimal impact on enrollment (La Tercera, 2020). This observation suggests that selection based on pre-existing conditions may not be a significant market concern.

The extensive variety and heterogeneity of private plans indicate that insurers strategically use plan design for risk selection. Our primary analysis highlights this phenomenon, quantifying how network-based selection contributes to adverse selection in the market. Additionally, it implicitly influences equilibrium plan design in counterfactual scenarios.

Finally, this market features minimal risk adjustment mechanisms. Private insurers manage a shared fund dedicated to pooling and redistributing expenses related to AUGE-GES conditions. As mentioned earlier, AUGE-GES conditions constitute a small proportion of inpatient care costs, rendering this adjustment mechanism negligible for our analysis. Nonetheless, because AUGE-GES significantly impacts outpatient expenditures, we model non-inpatient costs (referred to as administrative costs) as plan-specific rather than consumer-specific.

A.2.4 Coverage and Networks. Hospitals in Chile operate on a fee-for-service basis, submitting charges to insurers who then allocate costs to patients based on their plan's cost-sharing terms. Plans include deductibles, coinsurance rates, and coverage caps. Deductibles require consumers to pay full costs up to a defined threshold; however, these thresholds are relatively small compared to typical inpatient charges, thus playing a minor role in our analysis. Coverage caps set annual limits on insurer payments, serving as the Chilean counterpart to U.S. maximum out-of-pocket limits, though such limits are absent in Chile. These caps are generally high, and government-provided catastrophic insurance often covers excess charges. Consequently, coinsurance rates primarily determine patient cost-sharing, a factor explicitly modeled in our analysis.

Regulations mandate three types of plan networks. First, unrestricted network plans provide uniform coverage across all hospitals, similar to U.S. PPO plans. Second, preferential provider plans offer tiered coverage across selected hospitals, analogous to U.S. HMO plans. Third, restricted provider networks limit coverage exclusively to designated hospitals. Our analysis focuses on unrestricted and preferential provider plans, as restricted networks represent less than eight percent of plan codes and under four percent of enrollment. Coverage terms presented in our analysis aggregate benefits at the preferential and non-preferential levels. In practice, coverage within tiers can vary—for instance, plans might differ in their full or partial coverage of specific services such as imaging or surgery fees. However, within a coverage tier, treatments are uniformly covered. We utilize average effective coverage, closely reflecting consumer-facing plan descriptions. This simplification prevents us from analyzing selection driven by variations in coverage for specific medical treatments. Hospitals must provide care to insured patients and emergency care regardless of insurance status. Therefore, all consumers can access all hospitals, though insurance coverage levels may vary.

A.2.5 Solvency and Liquidity Regulation. Private insurers are required to maintain a solvency margin, with net equity equal to at least 30 percent of total liabilities. Insurers must also sustain a liquidity ratio of at least 0.8, meaning that liquid assets have to cover at least 80 percent of short-run obligations. They are also mandated to maintain a statutory guaranteed deposit covering obligations to beneficiaries and providers. For further details, see Superintendencia de Salud de Chile (2014). For our analysis, these regulations imply that insurers that offer more generous coverage face higher costs beyond their medical bills, as regulation requires them to have higher capital reserves, more liquidity, and larger guaranteed deposits.

A.3 Additional Descriptive Statistics

A.3.1 Insurance Market. Table A.2-A summarizes the variation in paid premiums and market shares across insurers. We document substantial variation in premiums across insurers, with the difference between the highest and the lowest insurer average premium being 36 percent of the market average. We observe significant variation in premiums within an insurer, likely driven by a combination of preference heterogeneity leading to compositional differences and the corresponding risk-rating and cost differences.

A.3.2 Hospital Market. Table A.2-B documents substantial price and market share variation across hospitals. The industry is moderately concentrated: no hospital has a market share higher than 13 percent, five hospitals have market shares between 8 and 13 percent, while the rest have market shares below 5 percent. The overall HHI is 1,387. The outside option—smaller private hospitals and all public ones—has a market share of 26 percent. Dispersion in admission prices across hospitals is substantial. For example, hospitals h_1 and h_6 charge average prices more than double those charged by h_4 and h_{11}. Differences in location, infrastructure, and real and perceived quality explain this price dispersion.

A.3.3 Plan Tiering. Table A.3-A describes the structure of plan preferential tiers by documenting the share of plans that have each hospital in its preferential tier. In most cases, VI insurers are more likely to place their integrated hospitals in the preferential tier of their plans relative to rival hospitals. For instance, m_a places its integrated hospitals in the preferential tier of between 37 and 69 percent of its plans, and m_b does so between 49 and 96 percent. In contrast, averaging across insurer-hospital combinations shows hospitals are preferential in only 20 percent of plans. Notably, regardless of vertical linkages, most hospitals feature in the preferential tier of plans offered by all insurers.

A.3.4 Admission Flows. Table A.3-B displays the breakdown of hospital admissions per insurer and shows that this pattern holds when looking at individual firms. For each integrated hospital, its integrated insurer is the dominant source of admissions, accounting for between 43 and 71 percent. Nevertheless, all integrated hospitals receive a substantial share of patients from integrated and non-integrated rival insurers, and all non-integrated hospitals receive patients from integrated insurers.

A.3.5 Comparison Between VI and non-VI patients. One explanation for outcome differences across admissions covered by insurers that are VI and non-VI with a hospital is that patients may differ in observables. To study this possibility, we estimate different versions of equation (2) in the main text, using patient observables as the dependent variable. Finding differences in observables depending on VI would suggest that the patient groups we are comparing are not balanced. Table A.4 shows the results.

The unconditional comparison in column (2) shows that patients from VI insurers at their integrated hospital are almost three years older, 6 percentage points less likely to be female, equally likely to be employed, and have 16 percent lower income than other patients. However, once we control for the set of fixed effects in equation (2), those differences get much closer to zero and become less statistically significant. Moreover, once we include an insurer-hospital fixed effect, VI and non-VI insurer patients only differ in age. The former are around one year younger on average and are otherwise balanced. This suggests that our comparison is based on groups of patients with similar characteristics. Regardless, we control for these observables in our analysis in Section 4.[1]

A.4 Additional Evidence on VI and Admission Outcomes

We complement the analysis of Section 4 by exploring whether VI affects the provision of services for which physicians enjoy some discretion. We focus on C-sections and ultrasounds during pregnancy, hemogram tests in digestive diagnoses, and chest X-rays and cross-section imaging in respiratory diagnoses. C-sections provide a particularly

[1]Note that the regression R^2 increases substantially once location, hospital, plan, and diagnosis controls are included in column (3). This suggests observables explain patient sorting across hospitals, which suggests accounting for observable heterogeneity along these dimensions in demand. Our demand estimates in Section 6 show substantial observable heterogeneity in preferences over plans and hospitals.

convenient setting for this test as, upon delivery, a physician must either implement a C-section or proceed with natural birth—there is no extensive margin decision. Therefore, we know the exact alternatives for physicians in these cases, which is not as evident for the other services we study. Table A.5-A displays the results, which provide mixed evidence on the association between VI and the provision of these services. On the one hand, we find that admissions from VI insurers tend to provide fewer C-sections, consistent with evidence from the U.S. (Cutler *et al.*, 2000; Johnson and Rehavi, 2016). On the other hand, we find the opposite effect on imaging and no significant effect on ultrasounds, hemograms, and chest X-rays. Moreover, once we focus on hospital-insurer variation only by including hospital-insurer fixed effects, the coefficients become less statistically significant, as shown by Table A.5-B. Taken together, these results do not suggest a strong relationship between VI and hospital provision of services. Hence, we do not model these dimensions of hospital behavior explicitly. However, in our counterfactual analysis in Section 7, we consider the role of cost efficiencies associated with VI.

B Model Appendix

This appendix presents additional details about how we formulate and implement our model. It also includes the proofs for Proposition 1 from the main text.

B.1 Connection with the Regulatory Environment

Relative to the regulatory environment described in Appendix A.2.2, the primary distortion arises from insurers' capacity to duplicate plans. Legally, insurers must assign a base premium to each plan and define up to two price schedules governing premium variations by age and gender. However, by duplicating plans under different codes, insurers are able to price across demographic segments differentially. In fact, many plans explicitly target demographic groups through branding. Hence, rather than directly following regulatory guidelines, our model reflects insurers' actual practices, informed by examining insurers' websites and marketing materials to approximate demographic segmentation. Similarly, we ignore the regulatory limits on premium increases as enforcement depends on consumer-submitted complaints within specific periods (Chile Atiende, 2024). Due to enforcement barriers and documented lax oversight (CIPER, 2013), these caps did not significantly constrain insurers during our study period.

To account for incomplete enforcement of minimum coverage requirements, discussed in Appendix A.2.2, we model a generalized plan-design cost component. Specifically, the regulatory cost aims to capture the probability and associated expense of regulatory inspection tied to plan design.

We incorporate guaranteed renewability in two simple ways. First, we allow consumers to re-enroll in their previously chosen plan when estimating demand. In counterfactuals, however, we treat all consumers as new potential enrollees. Second, we embed legal penalties representing disputes between insurers and hospitals over renewability and coverage, consistent with the evidence of enforcement via litigation discussed in Appendix A.2.2. We ignore the dynamic effects of guaranteed renewability both because they seem secondary to the study of VI and because the enforcement of these guarantees is incomplete, as noted in Appendix A.2.2.

We abstract from insurers' ability to reject applicants based on pre-existing conditions. As discussed in Appendix A.2, the impact of these practices is unclear. Insurers rely on consumers to voluntarily disclose their conditions, and a recent waiver of this requirement by an insurer led to no effects on enrollment. In addition, there is public supplementary insurance for pre-existing conditions, which also acts to counteract this selection.

Finally, the market features a minimal risk pool compensating insurers for spending variations on specific conditions. Given that few of these conditions are treated in inpatient settings, our analysis omits explicit risk adjustment modeling. Nonetheless, these conditions notably influence outpatient expenditures, justifying our approach of modeling non-inpatient (administrative) costs as plan-specific rather than consumer-specific.

B.2 The Pricing Subgame

We formulate the pricing subgame through two nested fixed-point conditions to improve computational efficiency and respect our equilibrium refinement. The outer layer corresponds to the stacked conditions associated with the optimality of negotiated and VI-optimal prices, while the inner loop operationalizes the premium-setting condition. In what follows, we ignore time indices when possible to reduce the notational burden.

B.2.1 Premiums. We use the approach of Morrow and Skerlos (2011) to form a fixed point formulation of the optimality condition associated with premium setting. This results in:

$$\Phi \quad = \Lambda(\Phi)^{-1}[\Gamma(\Phi)\Phi + C(\Phi)] \tag{1}$$

where Φ is the stacked vector of all plan premiums, Λ is a diagonal matrix with j-th element equal to $\sum_{i\in\mathcal{I}} |\mathcal{F}_i|^2 \alpha_i^M D_{ij}^M(\Phi)$, and Γ is a sparse matrix with (j, j')-th entry equal to $\sum_{i\in\mathcal{I}} |\mathcal{F}_i|^2 \alpha_i^M D_{ij}^M(\Phi) D_{ij'}^M(\Phi)$ if plan j and j' belong to the same insurer and zero otherwise.

Finally, $C(\cdot)$ is a vector capturing all costs and the effect of premiums on VI hospitals:

$$C(\Phi)[j'] = \sum_{i \in \mathcal{I}} |F(i)| D^M_{ij'}(\Phi) \left[\alpha^M_i (ec_{ij'} + \eta_j - \sum_{j \in \mathcal{J}_m} D^M_{ij}(\Phi)(ec_{ij} + \eta_j)) - 1 - \alpha^M_i \left(\pi^{H|m(j')}_{ij'} - \sum_{j \in \mathcal{J}} D^M_{ij}(\Phi) \pi^{H|m(j')}_{ij} \right) \right]$$

where ec_{ij} is consumer i's expected inpatient cost as described in equation (7) and $\pi^{H|m(j')}_{ij}$ captures insurer $m(j')$'s weighted profits at integrated hospitals from claims paid by plan j if individual i enrolls in it, namely:

$$\pi^{H|m(j')}_{ij} = \sum_{h \in \mathcal{H}} \theta^{m(j)}_h \sum_{i' \in \mathcal{F}_i} \sum_{d \in \mathcal{D}} r_{i'd} D^H_{i'hd|j} \omega_{i'd} (p_{m(j)h} - k_{hm(j)}).$$

As noted by Morrow and Skerlos (2011), while there are no theoretical guarantees that equation 1 is a contraction, it appears to be so in practice. We obtain consistent, rapid convergence to the same stable set of premiums, regardless of initial conditions.

B.2.2 Prices. For any hospital system $s \subset \mathcal{H}$, we follow the logic of Gowrisankaran *et al.* (2015) to express the optimality conditions associated with negotiated prices and VI-optimal prices (if any) as:

$$\boldsymbol{p}_s = \boldsymbol{k}_s - (\nabla_{p_s} D^{H\prime}_s \cdot diag(\boldsymbol{\theta}_s) + diag(\chi_s) \cdot \Lambda_s)^{-1} (diag(\boldsymbol{\theta}_s) \cdot D^H_s + \nabla_{p_s} \pi^M_{m(s)} + diag(\chi_s) \cdot \Gamma^{VI}_s)$$

where $\boldsymbol{p}_s$ and $\boldsymbol{k}_s$ are the vectors of prices and hospital costs across the system's hospitals and all insurers, D^H_s is the vector of total resource-weighted expected hospital demand from each insurer, $diag(\cdot)$ is the diagonalization operator, and χ_s is an indicator vector that equals one when the hospital and insurer in each row are not VI. Finally, Λ_s and Γ_s are defined as:

$$\begin{aligned}
\Lambda_s[i,j] &= \left(\frac{\tau_h}{1-\tau_h} \frac{\frac{\partial \pi^M_m}{\partial p_{h,m}} + \sum_{h' \in \mathcal{H}} \theta^m_{h'} \frac{\partial \pi^H_{h'}}{\partial p_{h,m}}}{\Delta_{m,s} V^M_m} \right) [i] \times \theta^m_{h'}[j] \Delta_{m[i],s} \tilde{D}^H_{m'h'}[j] \\
\Gamma^{VI}_s[i] &= \left(\frac{\tau_h}{1-\tau_h} \frac{\frac{\partial \pi^M_m}{\partial p_{hm}} + \sum_{h' \in \mathcal{H}} \theta^m_{h'} \frac{\partial \pi^H_{h'}}{\partial p_{h,m}}}{\Delta_{m,s} V^M_m} \Delta \pi^M_{m(s)} \right) [i] \\
\Delta_{m,s} V^M_m[i] &= \Delta_{m[i],s} \pi^M_{m[i]} + \sum_{h' \in \mathcal{H}} \theta^{m[i]}_{h'} \Delta_{m[i],s} \pi^H_{h'} + l_{m[i]h[i]} \Delta_{m[i],s} WTP_{m[i]s}
\end{aligned}$$

where $\Delta_{ms} WTP_{ms}$ is the total expected loss in network surplus from removing hospital system s's members from all of insurer m's networks. When computing this value, we assume the courts would compute consumers' expected network surplus using the average prices negotiated by rival insurers with the removed hospitals. We also assume they will use the realized insurance demand (under disagreement) to compute total losses. These assumptions mostly affect the identified distribution of multiplier L. Our estimates remain largely unchanged if we assume the courts use the full-agreement insurance demand or previous hospital prices as the basis for computing network surplus losses.

Unlike the premium fixed point formulation, we find that the expression above appears to be a contraction mapping only locally around the equilibrium. To address this, we use a globally convergent Anderson Acceleration routine (Zhang *et al.*, 2020). This procedure results in an efficient pricing subgame equilibrium solver, with the outer loop solving the more challenging price fixed-point problem and the inner loop providing equilibrium premiums conditional on the current guess of hospital prices.

B.3 Solving the Plan Design Problem

We begin by proving Proposition 1 from the main text, which is the central result for solving the plan design problem. We restate the proposition below for completeness.

Proposition 1. *A continuous extension of K_{mt} always exists and is bounded whenever K_{mt} is bounded. Moreover, let $\mathfrak{c}^*$ be the set of solutions to the design problem and $\tilde{\mathfrak{c}}(\lambda, G)$ be the set of solutions of its relaxation, then $\mathfrak{c}^* = \lim_{\lambda\to\infty} \tilde{\mathfrak{c}}(\lambda, G)$.*

Proof of Proposition 1. In what follows, we drop time indices to reduce notational burden. We begin by proving that a continuous extension of K_m exists. First note that $\mathcal{C} = \{(\underline{c}, \bar{c}, w) \in [0,1]^{2+|\mathcal{H}|} \mid \underline{c} \leq \bar{c} \ \wedge w_h(1-w_h) = 0 \ \forall h\}$, is the intersection of $\mathcal{C}^*$ and the sets $\bar{\mathcal{C}}_h = \{(\underline{c}, \bar{c}, w) \in [0,1]^{2+|\mathcal{H}|} \mid w_h(1-w_h) = 0\}$. Note that $\mathcal{C}^*$ is the preimage of the closed set $[0,\infty)$ of the continuous function $f : (\underline{c}, \bar{c}, w) \mapsto \bar{c} - \underline{c}$. Hence, $\mathcal{C}^*$ is closed. Analogously, $\bar{\mathcal{C}}_h$ is the preimage of the closed set $\{0\}$ of the continuous $f : (\underline{c}, \bar{c}, w) \mapsto w_h(1-w_h)$. Hence each $\bar{\mathcal{C}}_h$ is closed. Thus, $\mathcal{C}$ is closed and, by the Tietze extension theorem, $\hat{K}_m$ exists and coincides with K_m on $\mathcal{C}$. Boundedness of the extension follows from the same theorem.

We can now restate the original problem. Denote as $\mathcal{J}_m$ the set of plans to be optimized and the firm's objective $V(C) = \tilde{\pi}_m(C, \boldsymbol{\phi}^*(C), \boldsymbol{p}^*(C)) - \sum_{j\in\mathcal{J}_m} N_{s(j)}\hat{K}_m(C)$, where $C \in [0,1]^{|\mathcal{J}_m|\times|\mathcal{H}|}$ is the matrix of coverages from each plan at each hospital. Note that we have defined the problem using the continuous extension of K, $\hat{K}$, as it exists and coincides with K at any tiered coverage. Denote the set of tiered coverages as $\mathcal{C}_m = \{C \in [0,1]^{|\mathcal{J}_m|\times|\mathcal{H}|} | C_{(j,\cdot)} \in \mathcal{C}\}$ where $C_{(j,\cdot)}$ is the row associated with plan j and $\mathcal{C}$ is defined as in equation (11) in the main text. Thus, the original combinatorial optimization problem established in equation (11) can be expressed as $\mathcal{P}^* : \max_{C\in\mathcal{C}_m} V(C)$.

Let $\mathcal{P}^0$ denote the following continuous-control non-convex optimization problem:

$$\max_{\underline{c},\bar{c}\in[0,1]^{|\mathcal{J}_m|},W\in[0,1]^{|\mathcal{J}_m|\times|\mathcal{H}|}} V(diag(\underline{c})(1-W)+diag(\bar{c})W) \tag{2}$$

$$s.t \quad W_{jh}(1-W_{jh})=0 \quad \forall j\in\mathcal{J}_m, h\in\mathcal{H} \tag{3}$$

Denote $C_m^0 = \{C \in [0,1]^{|\mathcal{J}_m|\times|\mathcal{H}|} \mid \exists \underline{c},\bar{c} \in [0,1]^{|\mathcal{J}_m|}, W \in [0,1]^{|\mathcal{J}_m|\times|\mathcal{H}|}$, s.t. $C = diag(\underline{c})(1-W)+diag(\bar{c})W\}$. Note that $C_m \subset C_m^0$, since any element of C_m corresponds to an element of C_m^0 where the weight matrix W is an extreme point of its domain. Moreover, note that any element of C_m has a representation $(\underline{c},\bar{c},W)$ that satisfies the tiering constraint $W_{jh}(1-W_{jh})=0$ and any element of C_m^0 that satisfies the constraint is an element of C_m.

As $V(\cdot)$ is continuous and the set C_m^0 is compact, the Weierstrass extreme value theorem guarantees that a solution to $\mathcal{P}^0$ exists. Therefore, $\arg\max\mathcal{P}^* = \arg\max\mathcal{P}^0$ necessarily and neither is empty. By contradiction, suppose that $\exists\tilde{C}\in\arg\max\mathcal{P}^*$ not in $\arg\max\mathcal{P}^0$. Then there exists $\hat{C}\in\arg\max\mathcal{P}^*$ such that $V(\hat{C}) > V(\tilde{C})$ and moreover $\hat{C}$ satisfies the tiering constraint, hence it is an element of C_m. Analogously, if there exists $C^0\in\arg\max\mathcal{P}$ not in $\arg\max\mathcal{P}^*$, then there exists $C'\in C_m$ such that $V(C') > V(C^0)$ that satisfies the tiering constraint. Therefore, problem $\mathcal{P}^*$ and $\mathcal{P}^0$ are equivalent.

Let $\lambda > 0$ and $G(\cdot)$ be a positive, continuous, strictly monotonic function. Without loss, normalize $G(0)=0$. Define $\mathcal{P}^1(\lambda)$ the original statement of the proposition:

$$\max_{\underline{c},\bar{c}\in[0,1]^{|\mathcal{J}_m|},W\in[0,1]^{|\mathcal{J}_m|\times|\mathcal{H}|}} V(diag(\underline{c})(1-W)+diag(\bar{c})W)-\lambda\sum_{j\in\mathcal{J}_m}\sum_{h\in\mathcal{H}} G(W_{jh}(1-W_{jh})) \tag{4}$$

Again, this is a continuous optimization problem over a compact domain. Hence, a solution exists. Moreover, by Berge's maximum theorem, $\arg\max\mathcal{P}^1(\lambda)$ is continuous in λ. Note that any solution to $\mathcal{P}^0$ attains the minimum tiering penalty $\sum_{j\in\mathcal{J}_m}\sum_{h\in\mathcal{H}} G(W_{jh}(1-W_{jh}))=0$ and that for any convergent sequence of solution to $\mathcal{P}^1$, $C^1(\lambda)$, the tiering penalty at $C^1(\lambda)$ must be weakly decreasing in λ. Therefore, by the upper hemicontinuity of the solution to $\mathcal{P}^1(\lambda)$ established by the maximum theorem, $\lim_{\lambda\to\infty}\arg\max\mathcal{P}^1(\lambda) = \arg\max\mathcal{P}^0 = \arg\max\mathcal{P}^*$. □

The proof above contains a simple idea: Whether insurers face a strict requirement to offer tiered networks or a penalty for offering plans with untiered networks is equivalent. The proposition establishes a sequence of design problems under increasing penalties, which allows us to trade off exploration versus optimality when solving the problem. The rate at which these problems converge as the penalty λ increases is inherently tied to the value firms have from offering intermediate tiers of coverage. In our setting, the value of doing so is low because insurers use tiers to steer consumers to specific hospitals. To do so effectively, they must introduce a large wedge between the out-of-pocket price at the hospitals they want consumers to visit and those they do not. Designing plans with coverages between the preferential and base tiers counteracts the efforts to steer consumers. In Appendix C, we discuss how we implement the convergent sequence of design problems and find intersections of insurers' best responses.

C Methodological Appendix

This appendix provides additional details about the identification, estimation, and counterfactual simulation methodology.

C.1 Demand Instruments

C.1.1 Hospital Demand Instruments. The main endogeneity problem in identifying preferences over hospitals is that prices might be negotiated with knowledge of unobserved preferences for care. We instrument prices using supplementary data on claims paid by *closed* insurers—large employers who have formed insurance companies exclusively to cover their employees and who do not sell insurance on the open private market we study. For each medical event of a closed-insurer enrollee, we regress the total bill amount (i.e., the total amount paid to the hospital) on diagnosis-year, hospital fixed effects, age, and gender fixed effects. We use the estimated parameters to predict total closed-insurer payments for each option in our hospital demand panel. We multiply the predicted price by the enrollee's coinsurance rate to account for differential coverage. We use this predicted out-of-pocket price as our instrument.

Table A.6-A shows the first stage of the instrument on private enrollees' out-of-pocket hospital prices within the hospital demand panel. To match the estimating equation, the regression controls for distance and year-diagnosis-insurer-hospital fixed effects. As shown, the instrument has a strong and significant positive relationship with the endogenous price variable. While Table 2 in the main text shows the effect with the instruments included, Table A.6-B shows the impact of ignoring them. As usual, failing to instrument for endogenous prices results in a lower predicted price elasticity.

C.1.2 Plan Demand Instruments. The main endogenous variable in our plan demand estimation corresponds to plan premiums. As explained in the main text, we instrument this variable using the public hospital system's list prices to compute each household's expected spending. Adjusting spending by each plan's coverage, we construct a predicted cost metric for each insurer, which we call the actuarially fair premium. We instrument each plan's premium with the average of its rivals' actuarially fair premiums. Table A.6-A shows the first stage regression of the instrument, while Table A.6-B shows the effect of ignoring the instruments on estimated preferences. As with hospital demand, ignoring the endogeneity of premiums results in a lower predicted elasticity.

C.2 Identification of Price- and Premium-setting Parameters

This section provides the proof for Propositions 2 and 3 in the main text. We begin by developing some general properties of our model and argue that we can discuss identification formally in a simplified framework. We then prove that the additional identification challenges imposed by the introduction of VI are overcome once hospital costs are known. Thus, we focus on the separate identification of bargaining weights from hospital costs. We prove a general non-identification result, a somewhat technical necessary and sufficient condition result, and an easy-to-implement sufficient condition result for identification. Finally, we extend the model to include statistical uncertainty on the insurers' side and show conditions for identification of the extended model.

Since the proof uses cross-sectional variation only, we omit time subscripts. Before proceeding with the proof, we introduce some additional notation. First, we denote the total weighted demand to hospital h from household i conditional on enrolling in plan j as $D^H_{ih|j} = \sum_{i'\in\mathcal{F}_i}\sum_{d\in\mathcal{D}} r_{i'd}\omega_{i'd}D^H_{i'hd|j}$ and, similarly, the total weighted demand to hospital h from insurer m as $D^H_{hm} = \sum_{j\in\mathcal{J}_m}\sum_{i\in\mathcal{I}} D^M_{ij}D^H_{ih|j}$. Moreover, we denote the vector of total weighted demand from each insurer to hospital h as $\boldsymbol{D}^H_h$. Similarly, we denote by $\boldsymbol{D}^M_m$ the vector $(D^M_j)_{j\in\mathcal{J}_m}$ of demand for each plan of insurer m where $D^M_j = \sum_i D^M_{ij}$. Denote the difference in demand for hospital h from insurer m' under full agreement and under disagreement between the pair (m,h) as $\Delta_{mh}D^H_{hm'}$. Denote the matrix of demand for hospital h from each insurer (columns) under every potential disagreement (rows) as $\Delta_{h|M}\boldsymbol{D}^H_h$. Furthermore, denote the vector of premiums of insurer m as $\boldsymbol{\phi}_m$, the vector of hospital prices for h from all insurers by $\boldsymbol{p}_h$, and the vector of prices between insurer m and its integrated hospitals (if any) as $\boldsymbol{p}^{VI}_m$. We define $VI : \mathcal{H} \to \mathcal{M} \cup \{-1\}$ as a function that returns the integrated insurer of any hospital, or -1 if the hospital is not VI. Throughout, we retain the convention that bold letters denote vectors, and denote $diag(\cdot)$ the diagonalization operator, which takes a vector and returns a matrix with a diagonal that matches the input vector. As in the main text, we denote the distribution of observables as generated by the DGP, $F_{\boldsymbol{X},\boldsymbol{Y}}$.

We rely on a fundamental result from the study of additive random utility models, which establishes convexity properties of surplus functions and their connection with choice probabilities. We state and prove this result here for completeness.

Theorem C.1 (The Williams-Daly-Zachary Theorem)**.** *Let* $V(\boldsymbol{v}) = \int (\max_i v_i + \epsilon_i) f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon}$ *for* $v \in \mathbb{R}^n$ *and let* f *be absolutely continuous. Then* V *is continuous, convex, and satisfies* $\frac{\partial V}{\partial v_i} = \mathbb{P}_{\boldsymbol{\epsilon}}[i = i^* | \boldsymbol{v}]$ *where* i^* *denotes the optimal choice. Moreover, for a given* v_0*, define* $\bar{V} : \mathbb{R}^{n-1} \to \mathbb{R}$ *as* $\bar{V}(\bar{\boldsymbol{v}}) = V(0, \bar{\boldsymbol{v}})$*, then* $\bar{V}$ *is continuous, strictly convex, and satisfies* $\frac{\partial \bar{V}}{\partial \bar{v}_i} = \mathbb{P}_{\boldsymbol{\epsilon}}[i = i^* | \bar{\boldsymbol{v}}]$.

Proof. Continuity of V follows directly from that of f and the fact that the set $\{v_i\}_{i=1}^n$ is finite. Convexity follows from noting that V is the composition of the linear expectation operator and the convex maximum. As V is convex, it is differentiable Lebesgue almost everywhere. Let ι_j be the j-th basis vector in $\mathbb{R}^n$ and let $\alpha > 0$. Note that:

$$\begin{aligned}
V(\boldsymbol{v} + \alpha\iota_j) - V(\boldsymbol{v}) &= \int \Big[(\max_i v_i + \alpha \mathbb{1}\{i = j\} + \epsilon_i) - (\max_i v_i + \epsilon_i)\Big] f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \\
&= \int \mathbb{1}\{v_j + \alpha + \epsilon_j \geq \max_{i \neq j} v_i + \epsilon_i \geq v_j + \epsilon_j\} * \\
&\quad \Big[(\max_i v_i + \alpha \mathbb{1}\{i = j\} + \epsilon_i) - (\max_i v_i + \epsilon_i)\Big] f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon} \\
&\quad + \int \mathbb{1}\{v_j + \alpha + \epsilon_j < \max_{i \neq j} v_i + \epsilon_i\} 0 \, d\boldsymbol{\epsilon} \\
&\quad + \int \mathbb{1}\{v_j + \epsilon_j \geq \max_{i \neq j} v_i + \epsilon_i\} \alpha \, d\boldsymbol{\epsilon}
\end{aligned}$$

where the first equality is by definition and the second follows from decomposing the integral into three parts: (i) the region in which $v_j + \alpha + \epsilon_j$ is the maximum value action but $v_j + \epsilon_j$ is not; (ii) the region in which $v_j + \alpha + \epsilon_j$ is not the maximum value (iii) the region in which $v_j + \epsilon_j$ is the maximum value. Note that that the first term of the second equality is bounded above by $\alpha(\mathbb{P}_{\boldsymbol{\epsilon}}[j = i^* | \boldsymbol{v} + \iota_j \alpha] - \mathbb{P}_{\boldsymbol{\epsilon}}[j = i^* | \boldsymbol{v}])$ and below by zero. Note also that the last term is exactly the choice probability of action j multiplied by α. Hence, we have:

$$\alpha \mathbb{P}_{\boldsymbol{\epsilon}}[j = i^* | \boldsymbol{v}] \leq V(\boldsymbol{v} + \alpha\iota_j) - V(\boldsymbol{v}) \leq \alpha(\mathbb{P}_{\boldsymbol{\epsilon}}[j = i^* | \boldsymbol{v} + \iota_j \alpha] - \mathbb{P}_{\boldsymbol{\epsilon}}[j = i^* | \boldsymbol{v}]) + \alpha \mathbb{P}_{\boldsymbol{\epsilon}}[j = i^* | \boldsymbol{v}]$$

Dividing by α and taking the limit to zero yields the derivative in one direction. The proof for $\alpha < 0$ is symmetric. Moving to $\bar{V}$, continuity, convexity, and the connection between the gradient and choice probabilities follows directly from the preceding proof. To prove strict convexity, it suffices to show that the Hessian of $\bar{V}$ is positive definite. By the Gershgorin circle theorem, it suffices to show that the Hessian, $H\bar{V}$, has a dominant diagonal. Note that:

$$\frac{\partial \bar{V}}{\partial v_j} = \mathbb{P}_\epsilon[j = i^*|\boldsymbol{v}] = \int F(v_j + \epsilon_j, v_j - v_1 + \epsilon_j, \ldots, \epsilon_j, v_j - v_{j+1} + \epsilon_j, \ldots) d\epsilon_j$$

where F is the CDF associated with f. It is easy to see that:

$$\frac{\partial^2 \bar{V}}{\partial v_j^2} = \sum_{i=0, i\neq j}^{n} \left| \frac{\partial^2 \bar{V}}{\partial v_j \partial v_i} \right| > \sum_{i=1, i\neq j}^{n} \left| \frac{\partial^2 \bar{V}}{\partial v_j \partial v_i} \right|$$

Hence HV is positive definite and thus $\bar{V}$ is strictly convex. □

This proof approach combines elements from Lindberg (2012) and Hotz and Miller (1993). For the purpose of this section, we are interested primarily in the implied convexity properties. From now on, we refer to Theorem C.1 as the WDZ theorem. The theorem has immediate corollaries for the study of aggregate consumer surplus (and thus aggregate demand) functions:

Corollary 1 (Social Surplus)**.** *For every $i \in \mathcal{I}$ and $j \in \mathcal{J}$, denote δ_{ij} the ex-ante indirect utility for i of choosing j. If $\boldsymbol{x} \in \mathbb{R}^{|\mathcal{J}|}$ denotes an attribute such that $\delta_{ij} = \bar{\delta}_{ij} + v_i(x_j)$ for some v_i (weakly) convex and the social surplus function is given by $S(\boldsymbol{x}, \bar{\delta}; \boldsymbol{\omega}) = \sum_{i\in\mathcal{I}} \omega_i \int \left(\max_{j\in\mathcal{J}} \bar{\delta}_{ij} + v_i(x_j) + \epsilon_{ij} \right) dF(\boldsymbol{\epsilon}_i)$ for some positive welfare weights $\boldsymbol{\omega}$, then S is convex in $\boldsymbol{x}$ and $\nabla_x S$ equals a weighted sum of choice probabilities. Moreover, let $\delta_{i0} = 0$ for all i and denote $\bar{S}$ the normalized function. Then: (i) $\bar{S}$ is strictly convex in $\boldsymbol{x}$ and, (ii) if v_i is affine and v_i' has a single sign $s \in \{-1, 1\}$ for all i and $\omega_i = 1/|v_i'|$, then $\partial \bar{S}/\partial x_j = sP_j$, where P_j is the choice probability of j.*

The convexity properties follow directly from the fact that the positive-weighted sum of convex functions is convex and that the composition of a weakly convex function with a strictly convex increasing function is strictly convex. The derivative properties follow directly from social surplus being additive in individual surplus functions, satisfying the WDZ theorem. A second corollary is a law of demand for insurance:

Corollary 2 (Law of Demand for Insurance)**.** *Let $\alpha_i^M < 0$ and denote $V_i = \int \max_j (\delta_{ij}^M + \epsilon_{ij}^M) dF(\boldsymbol{\epsilon})$ and $S^M = \sum_{i\in I} \frac{1}{|\alpha_i^M||\mathcal{F}_i|} V_i$, denote individual and aggregate consumer surplus from insurance market. Then, $D_j^M = -\partial S^M / \partial \phi_j$ and $\nabla_{\phi_m} \boldsymbol{D}_m^M$ is negative definite.*

This result follows directly from noting that δ_{ij}^M is affine decreasing in ϕ_j with derivative $\alpha_i^M |\mathcal{F}_i|$. Hence, $\nabla_{\boldsymbol{x}} \bar{S}$ is positive definite and $\nabla_{\phi_m} \boldsymbol{D}_m^M$ is negative definite. A weaker result also applies to hospital aggregate demand.

Corollary 3 (WDZ for Hospitals). *Let $\alpha_i^H < 0$, $c_{jh} \in (0,1)$ for all j, h, and $\beta^M > 0$. For all i, j, d define $V_{ijd}^H = \int(\max_h \delta_{ihd|j}^H + \epsilon_{ihd}^H)dF(\boldsymbol{\epsilon}_{id}^H)$, $WTP_{ij}^H = \sum_{d\in\mathcal{D}} \frac{r_{id}}{|\alpha_i^H|} V_{ijd}^H$, and $V_i^H = \int\Big(\max_j \bar{\delta}_{ij} + \beta^M \sum_{i'\in\mathcal{F}_i} WTP_{i'j}^H + \epsilon_{ij}^M\Big)dF(\boldsymbol{\epsilon}_i^M)$ and $S^H = \sum_{i\in\mathcal{I}} \frac{1}{\beta^M} V_i^H$. Denote $o_{jh} = (1-c_{jh})p_{m(j)h}$, then $D_{h|m}^H = -\nabla_{\boldsymbol{o}_{hm}} S^H \mathbf{1}$.*

We note from these results that, in general, the presence of multiple medical conditions, multiple members per household, and even price preference heterogeneity is largely superfluous for the analysis of supply-side identification. The proofs below rely on the strong convexity properties of surplus and optimality of firm behavior, which are independent of these dimensions of heterogeneity. Put simply, the presence of multiple policyholders with heterogeneous additive product preferences is enough to rule out the use of any specialized tools for analyzing homogeneous demand. Thus, under the following assumption, we can restrict attention to a simplified model, noting that extending the proof to the heterogeneous case only adds cumbersome notation.

Assumption 1. *$\alpha_i^M < 0, \alpha_i^H < 0$ for all i, $\beta^M > 0$. For every pair (h, m) denote $\Delta_{hm}^H \tilde{\pi}_h \equiv \tilde{\pi}_h - \tilde{\pi}_{h\setminus m}$ and $\Delta_{hm}^M \tilde{\pi}_m = \tilde{\pi}_m - \tilde{\pi}_{m\setminus h}$. Further, assume that $\Delta_{hm}^H \tilde{\pi}_h > 0$ and $\Delta_{hm}^M \tilde{\pi}_m > 0$ and $(\partial\tilde{\pi}_m/\partial p_{hm})(\partial\tilde{\pi}_h/\partial p_{hm}) < 0$.*

We can now state the generic vertical market problem we will use to study supply-side identification. As in the main model, we denote $\mathcal{I}$ the set of consumers, $\mathcal{J}$ the set of plans, $\mathcal{M}$ the set of insurers, and $\mathcal{H}$ the set of hospitals. Denote $\mathcal{J}_m \subset \mathcal{J}$ for $m \in \mathcal{M}$ the set of plans owned by insurer m. Let f be an absolutely continuous density with full support and, for a vector $\boldsymbol{\epsilon} \in R^n$, denote $F(\boldsymbol{\epsilon}) = \prod_{l=1}^n F(\epsilon_l)$, where F is the CDF of f. Define the network surplus of plan $j \in \mathcal{J}$ for consumer $i \in \mathcal{I}$ to be

$$W_{ij} = \int(\max_{h\in\mathcal{H}\cup\{0\}} \delta_{ijh}^H - (1-c_{jh})p_{m(j)h} + \epsilon_{ijh})dF(\boldsymbol{\epsilon}_i)$$

with the outside option $h = 0$ having a normalized indirect utility of ϵ_{ij0}. By the WDZ theorem, W_{ij} is strictly convex in the vector $\boldsymbol{v}_{ij}^H = (v_{ijh}^H)_{h\in\mathcal{H}}$ where $v_{ijh}^H = (\delta_{ijh}^H + (1-c_{jh})p_{m(j)h})$.

Similarly, let $V_i = \int(\max_{j\in\mathcal{J}\cup\{0\}} \delta_{ij}^M - \phi_j + W_{ij} + \epsilon_{ij})dF(\boldsymbol{\epsilon}_i)$ denote consumer i's surplus. ϕ_j denotes the premium of plan j. The outside option $j = 0$ satisfies $\delta_{i0}^M = \phi_0 = W_{i0} = 0$. Hence, again, V_i is strictly convex with respect to $\boldsymbol{v}_i^M = (v_{ij}^M)_{j\in J}$ where $v_{ij}^M = \delta_{ij}^M - \phi_j + W_{ij}$.

We now present a result about the logit demand system that extends Corollary 3. While an equivalent result can be derived for the general absolutely continuous, fully supported error case, the conditions that lead to it and the respective proof are less transparent. Hence, we present it for the specific use case in which we apply our model. In what follows, we adopt the notation $\partial_x F$ to denote the Jacobian of F with respect to x; we use it to avoid confusion with the demand functions (D) and the set of plans ($\mathcal{J}$). We retain the convention that for a scalar-valued function, the gradient (∇) is a column and that the Jacobian is row-ordered.

Lemma 1 (Law of Demand for Hospitals)**.** *Let $\epsilon^H_{ihdt|j}$ be distributed iid T1EV across i, h, d, t, and j. For any $h \in \mathcal{H}$, if $c_{jh} \in (0,1)$ for at least one $j \in \mathcal{J}_m$ and every $m \in \mathcal{M}$, then $-\partial_{p_h} D^H_h$ is an M-matrix (positive dominant diagonal, non-positive off-diagonal). Moreover, if $c \in (\underline{c}, \bar{c}) \subset (0,1)$ and for every $i \in \mathcal{I}$, $D^M_{i0} > \frac{\bar{c}-\underline{c}}{2-(\bar{c}+\underline{c})}$, then the symmetric part of $\partial_{p_h} D^H_h$ is negative definite. In addition, $\Delta_{h|M} D^H_h$ is also an M-matrix and has a positive definite symmetric part.*

Proof. Define $D^H_{ihm} = \sum_{j \in \mathcal{J}_m} D^M_{ij} D^H_{ih|j}$ and $\tilde{D}^H_{ihm} = \sum_{j \in \mathcal{J}_m} D^M_{ij} D^H_{ih|j}(1 - c_{jh})$. It is easy to verify that:

$$\partial D^H_{hm} / \partial p_{hm'} = \sum_i \tilde{D}^H_{ihm} D^H_{ihm} - \mathbb{1}\{m = m'\} \tilde{D}^H_{ihm}$$

Hence:

$$\partial_{p_h} D^H_h = \sum_i \boldsymbol{D}^H_{ihM} \tilde{\boldsymbol{D}}^{H\prime}_{ihM} - diag(\tilde{\boldsymbol{D}}^H_{ihM})$$

where $\boldsymbol{D}^H_{ihM}$ and $\tilde{\boldsymbol{D}}^H_{ihM}$ are the natural vector extensions of the elements defined above. Thus, note that $\partial D^H_{hm} / \partial p_{hm'}$ has a negative diagonal and strictly positive off-diagonal entries. Moreover, note that the matrix has a dominant diagonal if the following inequality holds:

$$\begin{aligned} \tilde{D}^H_{ihm'}(1 - D^H_{ihm'}) \quad &> \quad \sum_{m \neq m'} \tilde{D}^H_{ihm'} D^H_{ihm} \\ \iff (1 - \sum_{j \in \mathcal{J}_{m'}} D^H_{ih|j} D^M_{ij}) \quad &> \quad \sum_{j \in \mathcal{J} \setminus \mathcal{J}_{m'}} D^H_{ih|j} D^M_{ij} \\ \iff 1 \quad &> \quad \sum_{j \in \mathcal{J}_m} D^H_{ih|j} D^M_{ij} \end{aligned}$$

We note that this holds always as $D^H_{ih|j} < 1$ and $\sum_{j \in \mathcal{J}_m} D^M_{ij} < 1$ due to the presence of the outside option. Hence, the matrix has a dominant diagonal. Given that the set of matrices with dominant positive diagonal and negative off-diagonal is closed under addition, $\partial_{p_h} D^H_h$ is an M-matrix.

Now, to get the symmetric definite part, note that by the Gershgorin circle theorem, a sufficient condition for this is that the diagonal of the symmetric part is dominant, or:

$$\begin{aligned}
\tilde{D}^H_{ihm'}(1 - D^H_{ihm'}) &> \frac{1}{2}\Big(\tilde{D}^H_{ihm'} \sum_{m \neq m'} D^H_{ihm} + D^H_{ihm'} \sum_{m \neq m'} \tilde{D}^M_{ihm}\Big) \\
\iff \tilde{D}_{ihm'}\Big(1 - \sum_{m \in \mathcal{M}} D^H_{ihm}\Big) + \tilde{D}^H_{ihm'} - D^H_{ihm'} \sum_{m \in \mathcal{M}} \tilde{D}^H_{ihm} &> 0
\end{aligned}$$

Using this inequality, we have that:

$$\begin{aligned}
\tilde{D}_{ihm'}\Big(1 - \sum_{m \in \mathcal{M}} D^H_{ihm}\Big) + \tilde{D}^H_{ihm'} - D^H_{ihm'} \sum_{m \in \mathcal{M}} \tilde{D}^H_{ihm} &> (1-\bar{c})D^H_{ihm'}D^M_{i0} + (1-\bar{c})D^H_{ihm'} - D^H_{ihm'}(1-\underline{c}) \sum_{m \in \mathcal{M}} D^H_{ihm} \\
&> D^H_{ihm'}\Big((1-\bar{c})(1+D^M_{i0}) - (1-\underline{c})(1-D^M_{i0})\Big)
\end{aligned}$$

It is easy to verify that the term in parentheses is positive whenever $D^M_{i0} > \frac{\bar{c}-\underline{c}}{2-(\bar{c}+\underline{c})}$.

Moving to the properties of $\Delta_{h|M} D^H_h$, note that for $m \neq m'$,

$$\begin{aligned}
\Delta_{hm'} D^H_{hm} &= \sum_{j \in \mathcal{J}_m} D^H_{ih|j}(D^M_{ij} - D^{M-m'\setminus h}_{ij}) \\
&= \sum_{j \in \mathcal{J}_m} D^H_{ih|j}\left(\frac{e^{\delta^M_{ij}}}{1 + \sum_{j' \in \mathcal{J}_{-m'}} e^{\delta^M_{ij'}} + \sum_{j' \in \mathcal{J}_{m'}} e^{\delta^M_{ij'}}} - \frac{e^{\delta^M_{ij}}}{1 + \sum_{j' \in \mathcal{J}_{-m'}} e^{\delta^M_{ij'}} + \sum_{j' \in \mathcal{J}_{m'}} e^{\delta^{M-m'\setminus h}_{ij'}}}\right) \\
&= \sum_{j \in \mathcal{J}_m} D^H_{ih|j} D^{M-m'\setminus h}_{ij} \cdot \sum_{j \in \mathcal{J}_{m'}} (D^M_{ij}(\sigma=0) - D^M_{ij}(\sigma=1))
\end{aligned}$$

where have used the superscript $-m' \setminus h$ to denote the variables whose values depend on the network of m' excluding h and have parameterized the presence of h in the network of m' by a variable $\sigma \in [0,1]$.[2] Hence, $D^M_{ij}(\sigma = 1) = D^M_{ij}$ and $D^M_{ij}(\sigma = 0) = e^{\delta^{M-m'\setminus h}_{ij}}/(1+\sum_{j' \in \mathcal{J}} e^{\delta^M_{ij'}})$. By the WDZ theorem, $D^M_{ij}(\sigma)$ is convex in σ, hence we have that $\frac{\partial D^M_{ij}(1)}{\partial \sigma} > D^M_{ij}(1) - D^M_{ij}(0) = D^M_{ij} - D^{M-m'\setminus h}_{ij}$ and that $\frac{\partial D^M_{ij}(1)}{\partial \sigma} = D^M_{ij} D^H_{ihj}$. Now note that $\Delta_{hm'} D^H_{hm}$ has a dominant diagonal if:

$$\sum_{j \in \mathcal{J}_{m'}} D^H_{ih|j} D^M_{ij} > \sum_{m \neq m'} |\Delta_{hm'} D^H_{hm}| = \sum_{j \in \mathcal{J} \setminus \mathcal{J}_{m'}} D^H_{ih|j} D^{M-m'\setminus h}_{ij} \cdot \sum_{j \in \mathcal{J}_{m'}} (D^M_{ij}(\sigma=1) - D^M_{ij}(\sigma=0))$$

Substituting in the convex inequality, we have that this holds trivially due to the presence of the insurance outside option. The proof for the positive definiteness of the symmetric part of $\Delta_{h|M} D^H_h$ follows the exact logic as before: write the diagonal dominance condition, use the convex inequality twice to show it holds, and appeal to the Gershgorin circle

[2] That is, multiply the contribution of h to the surplus of network j by σ.

theorem. $\square$

The result in Lemma 1 is useful in two ways. First, $-\partial_{p_h} D_h^H$ being an M-matrix implies it is invertible. Moreover, its sum with another M-matrix is also invertible, which will prove useful as $-\partial_{p_h} D_h^H$ and $\Delta_{h|M} D_h^H$ will show up additively in one of the propositions below. The negative definiteness of its symmetric part implies that $x'\partial_{p_h} D_h^H x > 0$ for any non-zero vector x. Moreover, the sufficient condition implied in the statement of Proposition 1 holds trivially in our data. In our setting, $\bar{c} \approx 0.8$ and $\underline{c} \approx 0.5$, which implies a bound of $D_{i0}^M > 3/7$, well below the market share of the public insurance option.

We highlight that the conditions stated in Proposition 1 above are significantly stronger than necessary. The M-matrix property is a natural consequence of the law of demand; the convex inequality used is a consequence of the WDZ theorem; and the negative definitions of the symmetric parts will hold in most cases. In particular, it holds whenever coverage to hospital h is symmetric. It is in no way special to the logit demand system—albeit proving it in the logit case is easier.

We can now state the first part of Proposition 2 from the main text. To recall, we have that the supply side objectives are characterized by profit functions:

$$\begin{aligned}
\pi_m^M &= \sum_{j \in J_m} \sum_i D_{ij}^M \left(\phi_j - \sum_{h \in H} D_{ih|j} c_{jh} p_{mh} - \eta_j \right) \\
\pi_h^H &= \sum_{j \in J} \sum_i D_{ij} D_{ih|j} (p_{m(j)h} - k_{mh})
\end{aligned}$$

And for any VI hospital or insurer, their objective is:

$$\tilde{\pi}_m = \pi_m^M + \sum_h \theta_h \pi_h^H$$

where we have normalized the integration weights ($\boldsymbol{\theta}$) relative to the insurer weight (θ_m).

Proposition 2 (Identification of $\boldsymbol{\eta}, \boldsymbol{\theta}$)**.** *Let Assumption 1 hold. Then $\boldsymbol{\eta}_m$ is identified from $F_{\boldsymbol{X},\boldsymbol{Y}}$ for all non-VI m. Let $\boldsymbol{k}$ be identified and $c_{jh} \in (0,1)$ for all j, h such that $m(j)$ and h are VI. Then $(\boldsymbol{\eta}, \boldsymbol{\theta})$ are identified from $F_{\boldsymbol{X},\boldsymbol{Y}}$ up to θ_m.*

Proof. The first order conditions for the optimal premium of a (potentially VI) insurer and optimal within-VI pricing are:

$$\begin{aligned}
\nabla_{\boldsymbol{\phi}_m} \pi_m^J - \partial_{\boldsymbol{\phi}_m} \boldsymbol{D}_m^{M\prime} \boldsymbol{\eta}_m + \partial_{\boldsymbol{\phi}_m} \boldsymbol{\pi}^H(\boldsymbol{k_m})' \boldsymbol{\theta}_m = 0 \\
\nabla_{\boldsymbol{p}_m^{VI}} \pi_m^J - \partial_{\boldsymbol{p}_m^{VI}} \boldsymbol{D}_m^{M\prime} \boldsymbol{\eta}_m + \partial_{\boldsymbol{p}_m^{VI}} \boldsymbol{\pi}^H(\boldsymbol{k_m})' \boldsymbol{\theta}_m = 0
\end{aligned}$$

where $\pi^J_m = \sum_{j\in J_m}\sum_i D^M_{ij}(\phi_j - \sum_{h\in H} D^H_{ih|j} c_{jh} p_{mh})$, $\boldsymbol{D}^M_m$ is the vector of $(\sum_i D^M_{ij})_{j\in J_m}$, and $\boldsymbol{\pi}^H$ is the vector of hospitals profits. By corollary 2, $\partial_{\phi_m} D^M_m$ is negative definite, so we can express this as:

$$\begin{aligned} \eta_m &= (\partial_{\phi_m} D^M_m)^{-1}(\nabla_{\phi_m}\pi^J_m + \partial_{\phi_m}\pi^H\theta_m) \\ \nabla_{p_m}\pi^J_m - (\partial_{p^{VI}_m} D^M_m)(\partial_{\phi_m} D^M_m)^{-1}(\nabla_{\phi_m}\pi^J_m) + A\theta_m &= 0 \\ A &\equiv \partial_{p^{VI}_m}\pi^{H\prime} - (\partial_{p^{VI}_m} D^M_m)(\partial_{\phi_m} D^M_m)^{-1}\partial_{\phi_m}\pi^{H\prime} \end{aligned}$$

From the first equation, it is direct that $\boldsymbol{\eta}$ is identified for all non-VI insurers, for which $\theta_m = 0$. For VI insurers, it suffices to show that A is invertible, in which case the system can be solved uniquely. Note that A has $|VI^{-1}(m)|$ columns, each with form:

$$\begin{aligned} A_h &= \partial_{p^{VI}_m}\pi_h - (\partial_{p^{VI}_m} D^M_m)(\partial_{\phi_m} D^M_m)^{-1}\partial_{\phi_m}\pi_h \\ &= \partial_{p^{VI}_m}\pi_h - (\partial_{p^{VI}_m} D^M_m)(\partial_{\phi_m} D^M_m)^{-1}(\partial_{\phi_m} D^M_m D^H_{h|m}(p_{mh} - k_{mh}) - \sum_{m'\neq m}\partial_{\phi_m} D^M_{m'} D^H_{h|m'}(p_{m'h} - k_{m'h})) \\ &= \partial_{p^{VI}_m}\pi_h - \partial_{p^{VI}_m} D^M_m D^H_{h|m}(p_{mh} - k_{mh}) - \sum_{m'\neq m}(\partial_{p^{VI}_m} D^M_m)(\partial_{\phi_m} D^M_m)^{-1}\partial_{\phi_m} D^M_{m'} D^H_{h|m'}(p_{m'h} - k_{m'h})) \\ &= (\partial_{p^{VI}_m} D^H_{h|m}) D^M_m (p_{mh} - k_{mh}) + D^{M'}_m D^H_{h|m}\iota_h \sum_{m'\neq m}(\partial_{p^{VI}_m} D^M_{m'} - (\partial_{p^{VI}_m} D^M_m)(\partial_{\phi_m} D^M_m)^{-1}\partial_{\phi_m} D^M_{m'}) D^H_{h|m'}(p_{m'h} - k_{m'h})) \end{aligned} \tag{5}$$

Now we note that $\partial_{p^{VI}_m} D^M_{m'} = (\partial_{p^{VI}_m} D^M_m)(\partial_{\phi_m} D^M_m)^{-1}\partial_{\phi_m} D^M_{m'}$. To show this, we express the condition in terms of surplus V. Without loss we can write the surplus as $V(u,v)$ where $(u,v) = (-\phi_m + W_m(p^{VI}_m), -\phi_{-m} + W_{-m})$. We can then write:

$$\begin{aligned} \partial_{p^{VI}_m} D^M_{-m} &= \partial_{p^{VI}_m}(-\partial_{\phi_{-m}} V) \\ &= \partial_{p^{VI}_m}(\partial_v V) \\ &= \partial_{p^{VI}_m} W_m \partial_{uv} V \end{aligned}$$

and:

$$\begin{aligned} (\partial_{p^{VI}_m} D^M_m)(\partial_{\phi_m} D^M_m)^{-1}(\partial_{\phi_m} D^M_{-m}) &= (\partial_{p^{VI}_m}\partial_u V)(\partial_{\phi_m}\partial_u V)^{-1}(\partial_{\phi_m}\partial_v V) \\ &= (\partial_{p^{VI}_m} W \partial_{uu} V)(-\partial_{uu} V)^{-1}(-\partial_{uv} V) \\ &= \partial_{p^{VI}_m} W \partial_{uv} V \end{aligned}$$

Hence, the last term in equation (5) is zero. This condition is essentially an envelope condition on the problem. So we have that:

$$A_h = (\partial_{p_m^{VI}} D_{h|m}^H) D_m^M (p_{mh} - k_{mh}) + D_m^{M\prime} D_{h|m}^H \iota_h$$

We can apply the same approach to simplify the original first-order condition as:

$$\begin{aligned}\nabla_{p_m^{VI}} \pi_m^J - (\partial_{p_m^{VI}} D_m^M)(\partial_{\phi_m} D_m^M)^{-1}(\nabla_{\phi_m} \pi_m^J) &= \partial_{p_m^{VI}}(D_m^M(\phi_m - ec_m)) - (\partial_{p_m^{VI}} D_m^M)(\partial_{\phi_m} D_m^M)^{-1} \partial_{\phi_m} D_m^M (\phi_m - ec_m) \\ &= -(\partial_{p_m^V I} ec_m) D_m^M\end{aligned}$$

Hence, the first-order condition for prices within VI firms is:

$$-\sum_{j \in J_m} \Big(\sum_h \nabla_{p_m^{VI}} D_{h|j}^H D_j^M c_{jh} p_{mh} + \sum_{h \in VI^{-1}(m)} D_{h|j}^H D_j^M c_{jh} \Big) + \sum_{h \in VI^{-1}(m)} ((\partial_{p_m^{VI}} D_{h|m}^H) D_m^M (p_{mh} - k_{mh}) + D_m^{M\prime} D_{h|m}^H \iota_h) \theta_h = 0$$

which is exactly the first-order condition of optimality for prices set by the VI firm in a gate-kept economy in which enrollment is fully inertial. From this, we can see that for any finite vector $\boldsymbol{\theta}$, VI prices must lie between those optimal for the hospital system and those for the insurer. From this, we can see that A is necessarily invertible: Suppose by contradiction A is not invertible, then there exists a non-zero $\boldsymbol{\theta}$ such that $A\boldsymbol{\theta} = 0$. This implies that the equilibrium prices coincide with the hospital's set prices in a gate-kept economy. As we noted above, this cannot happen at any finite θ. Hence, A must be invertible. □

A consequence of Proposition 2 is that we can ignore VI from the rest of the identification problem—conditional on identifying hospital costs, the degree of integration θ is identified. Thus, any additional complexity imposed by VI can be taken as given. We can thus state the remainder of Propositions 2 and 3 from the main text, assuming a non-VI market structure, with the extension to the VI setting being direct. However, we first prove a preliminary lemma that speaks to the general identification of the model:

Lemma 2. *For any hospital h, denote the identified set (from $F_{\boldsymbol{X},\boldsymbol{Y}}$) of cost vectors as K_h and the set of identified bargaining weight vectors as T_h. Then there exists a diffeomorphism $g_h : K_h \to T_h$.*

Proof. Fix non-VI insurer-hospital pair (h, m). Denote $\tilde{\tau}_{hm} = (1 - \tau_{hm})/\tau_{hm}$ and note that the Nash-Bargaining protocol implies a first-order optimality condition for hospital prices:

$$\frac{|\partial \pi_m^M / \partial p_{hm}|}{\Delta_{mh} \pi_m^M} = \tilde{\tau}_{hm} \frac{\partial \pi_h^H / \partial p_{hm}}{\Delta_{mh} \pi_h^H}$$

where, by Proposition 2, the left-hand side is identified, and hence we can treat it as a known component $\hat{\psi}_{hm} \equiv \frac{|\partial \pi^M_m / \partial p_{hm}|}{\Delta_{mh} \pi^M_m}$. Therefore, we can rewrite the equation above as:

$$\hat{\psi}_{hm} = \tilde{\tau}_{hm} \frac{\sum_{m' \in \mathcal{M}} \partial D^H_{hm'} / \partial p_{hm} (p_{hm'} - k_{hm'}) + D^H_{hm}}{\sum_{m \in \mathcal{M}} \Delta_{mh} D^H_{hm'} (p_{hm'} - k_{hm'})}$$

Denote $f_{hm}(\boldsymbol{k})$ the term multiplying τ_{hm} such that stacking the expression across insurers for a single h is:

$$\boldsymbol{\psi}_h = diag(\boldsymbol{\tau}_h) \boldsymbol{f}_h(\boldsymbol{k}_h) \tag{6}$$

In general, the expression above can be reorganized as:

$$\begin{aligned} \hat{\psi}_{hm} \sum_{m' \in \mathcal{M}} \Delta_{mh} D^H_{hm'} (p_{hm'} - k_{hm'}) &= \tilde{\tau}_{hm} (\sum_{m' \in \mathcal{M}} \partial D^H_{hm'} / \partial p_{hm} (p_{hm'} - k_{hm'}) + D^H_{hm}) \\ -\tilde{\tau}_{hm} D^H_{hm} + \sum_{m' \in \mathcal{M}} (\hat{\psi}_{hm} \Delta_{mh} D^H_{hm'} - \tilde{\tau}_{hm} \partial D^H_{hm'} / \partial p_{hm}) p_{hm'} &= \sum_{m' \in \mathcal{M}} (\hat{\psi}_{hm} \Delta_{mh} D^H_{hm'} - \tilde{\tau}_{hm} \partial D^H_{hm'} / \partial p_{hm}) k_{hm'} \end{aligned}$$

which we stack over all negotiations of the hospital h to obtain:

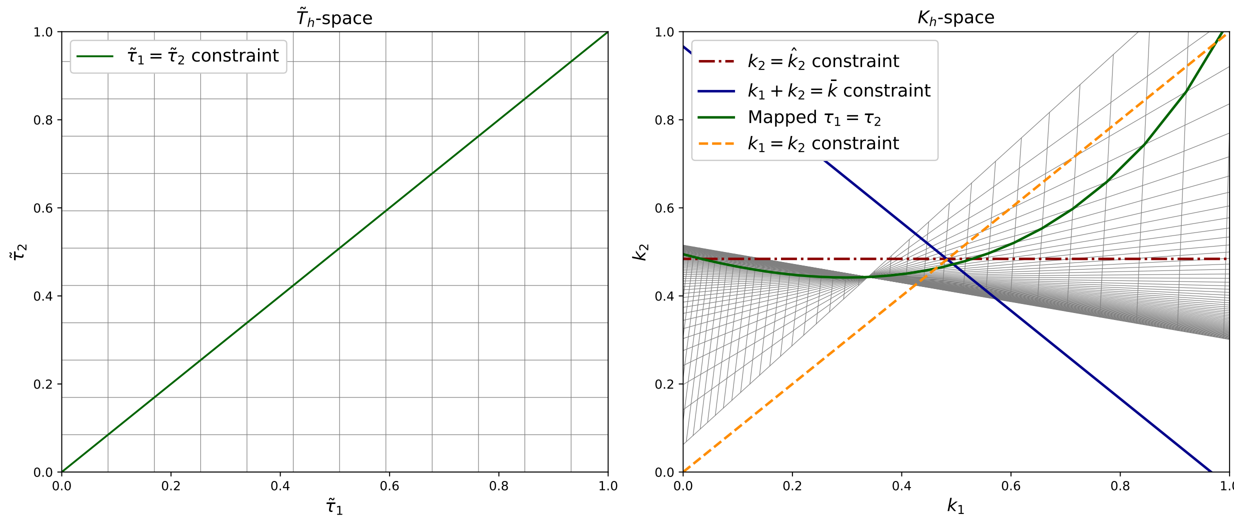
$$\begin{aligned} -diag(\tilde{\boldsymbol{\tau}}_{hM}) \boldsymbol{D}^H_h + (diag(\hat{\boldsymbol{\psi}}_h) \Delta_{h|M} D^H_h - diag(\tilde{\boldsymbol{\tau}}_{hM}) \partial_{\boldsymbol{p}_h} D^{H'}_h)(\boldsymbol{p}_h - \boldsymbol{k}_h) &= 0 \\ -(diag(\hat{\boldsymbol{\psi}}_h) \Delta_{h|M} D^H_h - diag(\tilde{\boldsymbol{\tau}}_{hM}) \partial_{\boldsymbol{p}_h} D^{H'}_h)^{-1} diag(\tilde{\boldsymbol{\tau}}_{hM}) \boldsymbol{D}^H_h + \boldsymbol{p}_h &= \boldsymbol{k}_h \end{aligned} \tag{7}$$

To see that the inverse $(diag(\hat{\boldsymbol{\psi}}_h) \Delta_{h|M} D^H_h - diag(\tilde{\boldsymbol{\tau}}_{hM}) \partial_{\boldsymbol{p}_h} D^{H'}_h)^{-1}$ exists and is unique, note that by Lemma 1, the two non-diagonal matrices in the expression are M-matrices. The class of M-matrices is closed under pre-multiplication by positive diagonals and addition, hence the term in parentheses is an M-matrix. Therefore, its inverse exists and is positive. Note that this is the case even if $c_{jh} = 1$ for all j, as then $\partial_{p_h} D^H_h = 0$.

Therefore, the data establishes a bijection $g_h : K_h \to T_h$, from the set of admissible costs K_h to the set of admissible bargaining weights T_h. $g_h(\boldsymbol{\tau})$ is defined by equation (7) and its inverse as $g_h^{-1}(\boldsymbol{k}_h) = diag(\boldsymbol{f}_h(\boldsymbol{k}_h))^{-1} \boldsymbol{\psi}_h$. Moreover, the bijection and its inverse are differentiable, making it a diffeomorphism. □

This diffeomorphism is a generalization of the cost inversion formula of Gowrisankaran *et al.* (2015) and Ho and Lee (2017), to a setting that includes both cost-sharing and downstream recapture. To illustrate the implications of this lemma, Figure A.1 illustrates the diffeomorphism for a single hospital with two insurers. On the left, we have the space of bargaining weights and a potential model constraint of homogenous bargaining weights

Figure A.1: Illustration of the model diffeomorphism



Notes: This figure illustrates the class of diffeomorphisms implied by the bargaining model in two dimensions. For ease of illustration, the inverse diffeomorphism is plotted, and the space of bargaining weights is plotted in its adjusted form $\tilde{\tau} = (1-\tau)/\tau$. On the left, the figure shows the space of bargaining weights for a given hospital and the constraint of homogeneous bargaining weights across insurers. On the right, the hospital costs space is shown. The deformed grid corresponds to the same grid of $\tilde{T}$ as mapped through the diffeomorphism. The homogeneity constraint on bargaining weights is also shown in this space, along with three potential constraints on the space of costs.

across insurers $\tilde{\tau}_1 = \tilde{\tau}_2$. On the right, the space of bargaining weights is transformed into the identified space of costs, K_h. The figure illustrates that the model contains some identifying power over costs, as only the gridded region is part of the identified space. However, the figure also illustrates clearly the lack of identification result noted in the first part of the second item in Proposition 2 in the main text. In general, we can see the issue in equation (6): there are $|\mathcal{H}|$ such conditions, each with $|\mathcal{M}|$ rows, but there are $2 \times |\mathcal{H}| \times |\mathcal{M}|$ parameters to identify. In particular, let $\boldsymbol{k}_h^1$ be any vector of hospital costs such that $\boldsymbol{f}_h(\boldsymbol{k}_h^1) > 0$ element-wise, and $\boldsymbol{\tau}_h^1$ its associated bargaining vector, such that the equality holds. Note that $\boldsymbol{f}_h$ is continuous, hence at there exists a neighborhood of $\boldsymbol{k}_h^1$ such that for all $\boldsymbol{k}_h^2$ in the neighborhood, $\boldsymbol{f}_h(\boldsymbol{k}_h^2) > 0$. Define $\boldsymbol{\tau}_h^2 = \boldsymbol{\tau}_h^1 diag(\boldsymbol{f}_h(\boldsymbol{k}_h^1)) diag(\boldsymbol{f}_h(\boldsymbol{k}_h^2))^{-1}$, then $(\boldsymbol{k}_h^2, \boldsymbol{\tau}_h^2)$ is admissible and observationally equivalent to $(\boldsymbol{k}_h^1, \boldsymbol{\tau}_h^1)$. Therefore, if there is an admissible tuple of cost and bargaining weights vector given observed prices, then there is a continuum of such tuples.

We are clearly missing $|\mathcal{H}| \times |\mathcal{M}|$ additional restrictions. The diffeomorphism, g_h, allows us to apply constraints across the two spaces, in a property we summarize in the following lemma, which matches item 2 of Proposition 2 of the main text.

Lemma 3. *Let Assumption 1 hold and denote g_h the identifying diffeomorphism from Lemma 2. Then, $(\boldsymbol{k}_h, \boldsymbol{\tau}_h)$ are identified if and only if there exists $\mu^K : \mathbb{R}^{|\mathcal{M}|} \to \mathbb{R}^{m_k}$ and $\mu^T : \mathbb{R}^{|\mathcal{M}|} \to \mathbb{R}^{m_T}$ differentiable constraints with full row-rank Jacobians, such that $\boldsymbol{k}_h$ satisfies the constraints if*

$\mu^K(\boldsymbol{k}) = 0$ and $\boldsymbol{\tau}_h$ satisfies the same if $\mu^T(\boldsymbol{\tau}_h) = 0$ such that the set $\{\boldsymbol{k} \in K_h \mid \mu^K(\boldsymbol{k}) = 0 \wedge \mu^T(g(\boldsymbol{k})) = 0\}$ is singleton.

Proof. If $\{\boldsymbol{k} \in K_h \mid \mu^M(\boldsymbol{k}) = 0 \wedge \mu^H(g(\boldsymbol{k})) = 0\}$ is singleton, then there is a unique $\boldsymbol{k}_h$ that satisfies all model constraint and complies with the optimality condition of bargaining. Thus the identified tuple is $(\boldsymbol{k}_h, g_h(\boldsymbol{k}_h))$. For the converse, we can consider μ^K to be the constraint that restricts K_h to a strictly monotonic one-dimensional line crossing $\boldsymbol{k}_h^*$ and μ^T such that its inverse map through g_h is a strictly monotonic (with an opposite sign to μ^K) one dimensional line. □

A natural consequence is that any broad set of additional dimension-reducing restrictions can help identify the model. These include homogeneity constraints on the bargaining weights, observational equivalence on costs, or even more complex conditions, such as variance restrictions. However, these constraints must satisfy the conditions of the lemma, which are non-trivial. To illustrate this, Figure A.1 also shows how the diffeomorphism transforms a constraint of homogeneity on bargaining weights within a hospital across insurers and how this constraint can pair successfully or unsuccessfully with three different constraints on the space of costs. First, consider the case in which the researcher has knowledge of the hospital's cost for one insurer and imposes $k_2 = \hat{k}_2$ for some observed $\hat{k}_2$. The dash-dotted horizontal line shows such a case. As can be seen in the figure, this constraint intersects twice with the mapped bargaining weights constraints, hence identification is not achieved. A second example is a constraint of equal costs across insurers within a hospital. In that case, the constraint equals the upward-sloping yellow dashed line. Again, we see that this constraint intersects twice with the bargaining constraint. Hence, identification is not achieved. Finally, suppose the researcher knows the total cost of a hospital. This case is illustrated in the downward-sloping blue line in the figure on the right. Clearly, this constraint intersects exactly once with the homogeneity constraint. Hence, identification is achieved. Note that all three constraints intersected with the bargaining constraint reduce the dimensionality of the identified set to zero, yet only one of them achieves point identification. Thus, the figure illustrates the transversality requirement on constraints implied by the lemma, showing that dimension counting (e.g., counting that we have more restrictions than unknowns) is insufficient. Intuitively, the transversality condition is the non-linear equivalent to a rank condition for a linear model.

The following proposition, which matches the third point in the main text Proposition 2, is one such case:

Proposition 3 (Conditional identification of $(\boldsymbol{\tau}, \boldsymbol{k})$)**.** *Let Assumption 1 hold. Then $\boldsymbol{k}$ and $\boldsymbol{\tau}$*

*are **not** separately identified from $F_{\boldsymbol{X},\boldsymbol{Y}}$. Let $\bar{\boldsymbol{k}}$ denote the mean cost of each hospital and let $\bar{\boldsymbol{X}}$ denote the extended set of observables that includes observation of $\bar{\boldsymbol{k}}$, then $\boldsymbol{k}$ and $\boldsymbol{\tau}$ are identified from $F_{\bar{\boldsymbol{X}},\boldsymbol{Y}}$ up to heterogeneity within hospitals in bargaining weights.*

Proof. The case in the proposition is an example of a scalar-reducing set of constraints. In particular $\mu^H(\boldsymbol{\tau}) = \sum_{i=1}^{|\mathcal{M}|-1} |\tau_i - \tau_{i+1}|$, hence $m_{\mu^H} = 1$ and the gradient is full row rank. $\mu^M(\boldsymbol{k}) = \frac{\sum_{m\in\mathcal{M}} D_{hm}k_m}{\sum_{m\in\mathcal{M}} D_{hm}} - \bar{k}_h$ hence $m_{\mu^M} = 1$ and the Jacobian is a diagonal matrix with strictly positive entries, and hence row rank. Given the dimensionality reduction implied by the constraints, the intersection is necessarily one-dimensional. For the intersection to be unique, we must show that the hyperplane implied by μ^M intersects exactly once with the one-dimensional line implied by the inverse of μ^H. To see that this is the case, note that we can take equation (7) and premultiply it by D_h^H to get:

$$D_h^{H'}(\tau_h^{-1} diag(\psi_h)\Delta_{h|m} D_h^H - \partial_{p_h} D_h^H)^{-1} D_h^H = \sum_m D_{hm}^H (p_{hm} - k_{hm})$$

Note that now the right-hand side is known, since $D_h^{H'}\boldsymbol{k}_h = \bar{k}_h(D_h^{H'}\mathbf{1})$. The left-hand side is a function of the unknown τ_h. By Lemma 1, $\Delta_{h|M} D_h^H$ has a definite symmetric part, which implies the function on the left is monotonic in τ_h. Therefore, there is a unique value of τ_h that satisfies the equality. Note that if $c_{jh} = 0$ for all $j \in \mathcal{J}$, the expression above is identical except that the term $\partial_{p_h} D_h^H$ is omitted, which does not change the result. □

Lemma 3 also speaks to the common practice of using orthogonality restrictions (i.e., instruments) to identify the Nash-in-Nash bargaining model. In particular, without further restrictions on bargaining weights, identification of $\boldsymbol{k}_h$, and thus $\boldsymbol{\tau}_h$, requires exactly $|\mathcal{M}|$ linearly independent instruments. These instruments need not only be orthogonal to costs, but also differentially so across insurers within a hospital, in a non-trivial way, as they must form a complete basis for $\mathbb{R}^{|\mathcal{M}|}$. Evidently, finding $|\mathcal{M}|$ such instruments is often difficult. This might explain why it is also common practice to impose homogeneity restrictions on bargaining weights (e.g., common weights within a hospital across insurers). In that case, however, Lemma 3 imposes further restrictions: the instruments must satisfy a transversality condition relative to the homogeneity constraints. Intuitively, it must be that the instruments on $\boldsymbol{k}_h$ do not imply some of the degrees of homogeneity on τ_h, or vice versa. This is why in the proof of Proposition 3 it does not suffice to argue that once one assumes homogeneity of bargaining weights within a hospital, then a single cost restriction is sufficient to identify the model. The proof, instead, must show that the cost restriction mapped through the model diffeomorphism g_h implies a feasible set for the weights that has a unique intersection with the feasible space implied by the homogeneity constraints. In particular, replacing the average-cost constraint with any other constraints of the type $\boldsymbol{T}\boldsymbol{k}_h = \boldsymbol{k}_o$ for a known map $\boldsymbol{T}$ and outcome $\boldsymbol{k}_o$ implies (through equation (7)) a non-quadratic—and thus, often non-monotonic—relationship between costs and weights. Such constraints might result in multiple solutions and thus partial identification.

Having established our key results for the general framework, we now move to the slightly modified version specialized to the Chilean setting. In particular, we introduce the disagreement penalty ζ_{hm}, such that insurer m's gains from trade with hospital h are now given by $\Delta^M_{hm}\tilde{\pi}_m = \tilde{\pi}_m - \tilde{\pi}_{m\backslash h} + \zeta_{hm}$. As in Assumption 1, we retain the assumption that $\Delta^M_{hm}\tilde{\pi}_m > 0$ for every non-VI pair (h, m). We assume that $\zeta_{hm} = z_{hm} l_{hm}$ where z_{hm} is a known scalar and l_{hm} is distributed iid L across insurers and hospital pairs. In addition, we adopt a parametric form on hospital costs to reduce its dimensionality. To make the assumptions of the extended model explicit, we summarize them in the following statement, together with the remaining assumptions required for our final result.

Assumption 2. *The following conditions hold: (i) There exists $\boldsymbol{w}_{hm} \in \mathbb{R}^d$ observed such that $k_{hm} = \bar{k}_h + \boldsymbol{w}_{hm}\boldsymbol{\beta}$ with $d < |\mathcal{H}|(|\mathcal{M}| - 1)$ and stacked matrix $\boldsymbol{W}_h$ column rank for each h. (ii) Bargaining weights are homogeneous within hospital, $\tau_{hm} = \tau_h$. (iii) The distribution of disagreement penalties, L, is absolutely continuous and has full support, and $\mathbb{E}[l_{hm}|\bar{\boldsymbol{X}}, \boldsymbol{W}] = 0$. (iv) Either $c_{jh} = 1$ for every $h \in \mathcal{H}$ and $j \in \mathcal{J}$; or $c_{jh} \in (\underline{c}, \bar{c}) \subset (0, 1)$, and for every $i \in \mathcal{I}$, $D^M_{i0} > \frac{\bar{c}-\underline{c}}{2-(\bar{c}+\underline{c})}$.*

Part (i) of Assumption 2 allows for parametric dimension reduction of costs in flexible ways as $\boldsymbol{w}_{hm}$ can include either continuous variables or discrete indicators. Part (ii) is a direct consequence of Proposition 3, which establishes that only bargaining weights up to hospital heterogeneity are identified from our data. Part (iii) states our assumption on the disagreement penalty, taking the usual exogeneity and distributional assumptions imposed on shocks in the non-parametric identification literature. Finally, part (iv) implies that either the derivative of hospital demand with respect to prices conditional on enrollment is zero or that the outside option is strong enough to establish a law of demand. As we noted above, this condition holds trivially in our data and is stronger than necessary in most cases.

The results developed before can be easily extended to this new specification. All the results treat $\Delta^M_{hm}\tilde{\pi}_m$ as an identified quantity; thus, they can all be stated as including the assumption that ζ_{hm} is also identified. The following proposition leverages these prior results and establishes our final identification result, which corresponds to Proposition 3 from the main text:

Proposition 4. *Let Assumptions 1 and 2 hold. Then $(\boldsymbol{k}, \boldsymbol{\beta}, \boldsymbol{\tau}, L)$ are identified from $F_{\bar{\boldsymbol{X}},\boldsymbol{Y}}$.*

Proof. From Proposition 3, we can treat $\bar{k}_h$ and τ_h as identified from observing of $\bar{\boldsymbol{k}}$. We can write the bargaining optimality conditions as:

$$\frac{\Delta_{mh}\pi_m^M}{z_{hm}} = \frac{|\partial\pi_m^M/\partial p_{hm}|}{z_{hm}\tilde{\tau}_h}\frac{\Delta_{mh}\pi_h^H(\boldsymbol{k}_h)}{\partial\pi_h^H/\partial p_{hm}(\boldsymbol{k}_h)} + l_{hm}$$

Note that this is a non-linear regression model. We will now show that the first term on the right is a monotonic function of $\boldsymbol{k}_h$, and thus, taking expectation over l_{hm}, can be inverted to recover $\boldsymbol{\beta}$. From there, non-parametric identification of L follows from standard arguments (Matzkin, 2007). We show the proof for the case of $c_{jh} < 1$ for all pairs (j, h). The proof for the case of full insurance ($c_{jh} = 1$) follows directly as below, but removing any derivative of demand with respect to price.

Denote the function on the right of the equality as f_{hm} and consider its stacked version over m, $\boldsymbol{f}_h : \mathbb{R}^{|\mathcal{M}|} \to \mathbb{R}^{|\mathcal{M}|}$. Note that the Jacobian of $\boldsymbol{f}_h$ has an (i, j) entry given by:

$$[\partial_{p_h}\boldsymbol{f}_h]_{ij} = \left(\frac{|\partial\pi_m^M/\partial p_{hm}|}{z_{hm}\tilde{\tau}_h}\right)_i \frac{\Delta_{hm_i}D_{hm_j}(\sum_{m'}\partial_{p_{hm_i}}D_{hm'}^H(p_{hm'} - k_{hm'})) - \partial_{p_{hm_i}}D_{hm_j}^H(\sum_{m'}\Delta_{hm'}D_{hm'}^H(p_{hm'} - k_{hm'}))}{(\sum_{m'}\partial_{p_{hm_i}}D_{hm'}^H(p_{hm'} - k_{hm'}) + D_{hm}^H)^2}$$

We now want to show that this gradient has a positive definite symmetric part and appeal to the Gale-Nikaido Theorem (Gale and Nikaido, 1965), which yields injectivity of $\boldsymbol{f}_h$. To see that, we examine the diagonal dominance of the Jacobian. Note that diagonal dominance is not affected by premultiplication by a positive diagonal matrix. Hence, the property depends exclusively on the term:

$$\Delta_{hm_i}D_{hm_j}(D_{hm_i}^H + \sum_{m'}\partial_{p_{hm_i}}D_{hm'}^H(p_{hm'} - k_{hm'})) - \partial_{p_{hm_i}}D_{hm_j}^H(\sum_{m'}\Delta_{hm_i}D_{hm'}^H(p_{hm'} - k_{hm'}))$$

Note that $(D_{hm}^H + \sum_{m'}\partial_{p_{hm_i}}D_{hm'}^H(p_{hm'} - k_{hm'})) = \frac{\partial\pi_h^H}{\partial p_{hm}}$, which is positive by Assumption 1. Similarly, $\sum_{m'}\Delta_{hm_i}D_{hm'}^H(p_{hm'} - k_{hm'})$ are the gains from trade of hospital h when negotiating with insurer m', which are also positive by assumption. Hence, by Lemma 1, the expression is the positive sum of two M-matrices with positive definite symmetric parts: $\Delta_{hM}D_h$ and $-\partial_{p_h}D_h$. Thus, the gradient of $\boldsymbol{f}_h$ is injective, concluding the proof. □

We see this final result as the most practical one from this section. Conditional on simplifying the bargaining weight structure—a consequence of Proposition 3—it provides boundaries on our ability to shift identifying variation from residual cost heterogeneity for the hospital to residual disagreement cost heterogeneity for the insurer. If $\boldsymbol{w}_{hm}$ is taken to be a full set of fixed effects for each hospital and $|\mathcal{M}| - 1$ insurers, then we recover the full cost heterogeneity, and L will be point identified as a mass point at zero. However, it also allows us to identify specifications in which cost heterogeneity is coarser, in which case L becomes non-degenerate. It is important to note that these different models are observationally equivalent; hence, where to assign the heterogeneity should depend on the context (i.e., it is a modeling decision).

This result also implies that counterfactuals that depend on the distinction between heterogeneity in bargaining weights and hospital costs are also, crucially, not identified. For example, it states that this framework cannot be used (without further model restrictions) to determine whether price increases following a merger stem from increased costs versus higher bargaining weights. In contrast, our counterfactual of banning VI can hold bargaining weights and cost heterogeneity fixed across negotiating parties. As noted by our descriptive evidence, even if we tried to identify cost changes stemming from VI, we would likely fail to estimate anything meaningful, as there does not seem to be variation in service provision or service intensity. This lack of identification also justifies our bounding approach to cost effects, in which we simulate counterfactual cost efficiencies.

Finally, it is worth noting two practical benefits of having the structure in Proposition 4. First, the introduction of L allows us to specify a well-defined data likelihood for all negotiations. Importantly, it enables us to distribute statistical uncertainty between hospitals' gains from trade (through $\boldsymbol{k}_h$) and insurers' gains from trade (through ζ_{hm}). Without this uncertainty, the model is rejected by the data, as a few insurers have minimally negative gains from trade with some high-priced non-VI hospitals. Second, removing within-hospital heterogeneity in bargaining weights implies their identification from average hospital costs, and allows us to use the nonlinear regression inversion developed in the constructive proof of identification in Proposition 4. This approach allows us to recover a point estimate of l_{hm} for every hospital-insurer pair. Hence, in counterfactuals, we can hold a firm's legal cost fixed, avoiding the need to simulate multiple draws to integrate over potential equilibria of the pricing subgame.

C.3 Price and Premium Setting Parameter Estimator

We estimate our model's price and premium setting parameters using an iterative two-step procedure. The outer loop takes a guess of hospital costs and applies a linear inversion to recover hospital plan administrative cost (η_{jt}) and VI weights (θ_{hmt}) from the optimality conditions associated with plan premiums and VI prices. Conditional on plan administrative cost and VI weights, the inner loop solves a constrained maximum likelihood problem to recover hospital costs, bargaining weights, and the distribution of the disagreement penalty multiplier. The estimator can be described as follows:

$$
\begin{aligned}
\max_{\boldsymbol{\tau}\in[0,1]^{|H|},\boldsymbol{k}^H,\tilde{\boldsymbol{k}}^H,\mu_l,\sigma_l} \; & \sum_{h,m,t\in\mathcal{B}} \ln \mathcal{L}(p^*_{hmt}|\boldsymbol{\beta}^{\lambda},\mu^L,\sigma^L,\boldsymbol{\beta}^c) \\
\text{s.t.} \quad & k^H_h + \tilde{k}^H_{mt} \geq 0 \quad \forall h,m,t && (C1) \\
& \Delta_{mh}\tilde{\pi}_{ht} \geq 0 \quad \forall h,m,t && (C2_H) \\
& \Delta_{mh}\tilde{\pi}_{mt} + l_{hmt}\Delta_{mh}WTP_{mht} \geq 0 \quad \forall h,m,t && (C2_M) \\
& \frac{\partial \Delta_{mh}\tilde{\pi}_{ht}}{\partial p_{hmt}} \geq 0 \quad \forall h,m,t && (C3) \\
& k^o_{ht} = \sum_{m'} D^H_{h|m}(k^H_h + \tilde{k}^H_{mt}) \quad \forall h,t && (MATCH) \\
& \nabla_{[\phi_m,\boldsymbol{p}^{VI}_m]}\tilde{\pi}_{mt} = 0 \quad \forall \text{ VI-}m && (INV)
\end{aligned}
$$

where the likelihood of hospital prices is evaluated over all hospital-insurer price negotiations. The first four constraints match the requirements of our identification results: $C1$ imposes that hospital costs are positive, $C2_H$ and $C2_M$ that negotiations are individually rational, and $C3$ that hospital gains from trade are increasing in price at the observed prices. These constraints impose bounds on hospital costs, as neither bargaining parameters nor the penalty distribution enters them. The $MATCH$ condition incorporates our additional hospital cost data, where $D^H_{h|m}$ denotes the demand for hospital h from insurer m, weighted by resource intensity ω_{id}. Finally, the INV constraint captures the linear inversion procedure for recovering plan administrative costs and VI weights from VI firms' optimality conditions.

Following the Nash bargaining model formulated in Section 5, the likelihood of observing p^*_{hmt} for the negotiation between hospital h and insurer m in year t is the likelihood with which observed prices satisfy the first-order optimality condition:

$$
\begin{aligned}
\mathcal{L}(p^*_{hmt}|\boldsymbol{\beta}^{\lambda},\mu^L,\sigma^L,\boldsymbol{\beta}^c) &= \mathbb{P}\left(\lambda_h \frac{\partial\tilde{\pi}_{mt}/\partial p_{hmt}}{\Delta_{mh}\tilde{\pi}_{mt} + l_{mht}\Delta_{mh}WTP_{mt}} + \frac{\partial\tilde{\pi}_{ht}/\partial p_{hmt}}{\Delta_{mh}\tilde{\pi}_{ht}} = 0\right) \\
&= \mathbb{P}\left(l_{hmt} = -\frac{\lambda_h}{\Delta_{mh}WTP_{mt}}\frac{\partial\tilde{\pi}_{mt}}{\partial p_{hmt}}\frac{\Delta_{mh}\tilde{\pi}_{ht}}{\partial\tilde{\pi}_{ht}/\partial p_{hmt}} - \frac{\Delta_{mh}\tilde{\pi}_{mt}}{\Delta_{mh}WTP_{mt}}\right)
\end{aligned}
$$

where $\lambda_h = \frac{\tau_h}{1-\tau_h}$. The analytic form for this expression is easily derived from the assumption that l is normally distributed.

A relevant advantage of splitting the estimation into two nested loops is that the resulting constraints on the maximum likelihood problem are linear in parameters. Moreover, only cost parameters ($\boldsymbol{k}^H, \tilde{\boldsymbol{k}}^H$) feature in them, appearing only in hospital profits, which are linear in costs.[3] Hence, all constraints are linear, and the feasible set of cost solutions is a regular convex set. It is easy to see that because $\tilde{\pi}_{mt}$ is linear in administrative costs (η_{jt}) and VI weights (θ_{hmt}), these are uniquely determined from the optimality condition of VI prices and premiums, conditional on a guess of hospital costs (see the proof of Proposition 2 in the main text). Hence, this estimator is simple and efficient when implemented in a two-step procedure, and the feasible region of parameters is easy to explore. None of the first four constraints is binding, and the other two can be absorbed within the likelihood objective. Therefore, the asymptotic properties of the estimator are well described by those of standard maximum likelihood. Figure A.6 shows the convergence rate of the solution in our estimation. It shows that within two iterations, we attain a stable solution for supply-side parameters, which takes only a few minutes to solve.

C.4 Plan Design Cost Estimator

The cost of designing plan coverage is estimated in two steps. First, we estimate the regulatory cost component, which is a function of the continuous coverage level of each plan. We specify the cost as $K^r_m(\boldsymbol{c}_{jt}) = \exp(c^K(\boldsymbol{c}_{jt})) + \underline{c}_{jt}\underline{\zeta}_{jt} + \bar{c}_{jt}\bar{\zeta}_{jt}$, where $c^K(\cdot)$ is a flexible polynomial of coverage and $(\bar{\zeta}, \underline{\zeta})$ are mean zero iid normal shocks of unknown variance. We estimate this component of the model by using the optimality condition associated with coverage levels of each plan $(\underline{c}_{jt}, \bar{c}_{jt})$:

$$\begin{aligned}
\frac{\partial \tilde{\pi}_{mt}(\boldsymbol{c}, \boldsymbol{\phi}^*(\boldsymbol{c}), \boldsymbol{p}^*(\boldsymbol{c}))}{\partial \underline{c}_{jt}} - M_{jt}\frac{\partial K^r_m(\boldsymbol{c}_j)}{\partial \underline{c}_{jt}} &= 0 \\
\frac{\partial \tilde{\pi}_{mt}(\boldsymbol{c}, \boldsymbol{\phi}^*(\boldsymbol{c}), \boldsymbol{p}^*(\boldsymbol{c}))}{\partial \bar{c}_{jt}} - M_{jt}\frac{\partial K^r_m(\boldsymbol{c}_j)}{\partial \bar{c}_{jt}} &= 0
\end{aligned}$$

Given the pricing subgame, the primary challenge in this estimation procedure is computing the derivative of equilibrium profits. With the average insurer offering 81 plans with two tiers each over four years of data, this necessitates calculating 648 derivatives of the subgame equilibrium for each insurer. This is made feasible by recent advances in GPU-accelerated linear algebra and automatic differentiation(Bradbury *et al.*, 2018).

Having estimated the continuous regulatory cost, we turn to estimating the fixed cost of tiering. We define $\tilde{V}_{mt} = \tilde{\pi}_{mt} - \sum_{j \in \mathcal{J}_{mt}} N_{jt} K^r_m(\boldsymbol{c}_j)$, and $\Delta_{jht}\tilde{V}_{mt}$ denotes the change in $\tilde{V}_{mt}$ when the tiering position of hospital h in plan j is inverted. The model implies the

[3]In practice, we do not impose the *MATCH* constraint directly but add it as a large quadratic cost to the solver. This is to avoid issues with empty feasible sets on single hospitals for some intermediate solver iterations. On convergence, the solution satisfies the constraint.

inequalities:

$$[(\Delta_{jht}\tilde{V}_{mt} - M_{jt}\vartheta_{hmt})(1 - w_{hjt}) - \bar{\varsigma} w_{hjt}]Z^K_{hjt} \leq 0 \tag{8}$$

$$[(\Delta_{jht}\tilde{V}_{mt} + M_{jt}\vartheta_{hmt})w_{hjt} - \bar{\varsigma}(1 - w_{hjt})]Z^K_{hjt} \leq 0 \tag{9}$$

where $\Delta_{jht}\tilde{V}_{mt}$ is the difference in insurance profit minus regulatory costs when the tiering decision of hospital h in plan j is inverted. Equation (8) implies a lower bound on tiering costs ϑ_{hmt}, stating that if insurer m decided to leave hospital h in the base tier of plan j, then tiering costs must have been sufficiently large. Equation (9) states that if plan j chose to make hospital h preferential, tiering costs must have been sufficiently small, placing an upper bound on costs.

We implement equations (8) and (9) following the approach of Canay *et al.* (2023). We form a grid of nearly two thousand potential tiering costs, ranging from -3 to 3 million dollars per hundred thousand enrollees. We use the test of Chernozhukov *et al.* (2019) to admit points into the identified set. To evaluate the associated confidence interval, we identify the minimum and maximum points that satisfy the inequality and refine the cost grid to pinpoint the exact values where the inequality binds. We apply the bootstrap approach described in Chernozhukov *et al.* (2019), taking 300 random samples of our data with replacement.

Evaluating the impact of tiering decisions on equilibrium conditions is computationally intensive. With 11 hospitals to tier, we need to evaluate approximately 17,820 different subgame equilibria to form the estimator. Again, this is facilitated by optimizing the subgame equilibrium fixed-point solver and employing GPU acceleration.

C.5 Solving Counterfactual Equilibria

This section discusses our approach to solving the counterfactual equilibrium. In Appendix C.2 we show that Proposition 3 of the main text allows us to write the bargaining first-order conditions as a non-linear least squares regression, which allows us to recover a point estimate of l_{hmt} for every observation. Hence, in counterfactuals, we can hold a firm's legal penalty fixed, avoiding the need to integrate over potential equilibria of the pricing subgame. We seek to simulate the "average" outcome of banning VI, and hence simulate the counterfactual with the tiering-cost shock set to zero. Hence, simulating counterfactual equilibria involves two primary challenges: solving the combinatorial plan design problem and identifying a set of equilibrium strategies for all firms.

We tackle the first problem by relying on the result of Proposition 1 and XLA/JAX.

We define a grid of tiering penalties $\lambda \in \{0, 1, 10, 100\}$ and a penalty function $G(x) = x^2$. Given that our costs are measured in millions, these choices imply a maximum penalty of 68.75 million per plan for violating the tiering constraint. Fixing rival designs, we solve each firm's design problem in parallel, determining the solution $\tilde{c}^*(\lambda)$ along the grid. The initial condition at $\lambda = 0$ and the first iteration of the solver are always set to the status quo, with subsequent iterations updating from the previous solution. If at any $\lambda < 100$ the single-firm optimum has a tiering violation ($\sum_{j \in \mathcal{J}_{mt}} \sum_{h \in \mathcal{H}} G(w_{jh}(1 - w_{jh}))$) below 10^{-4}, we accept it as optimal for the best response iteration.

To illustrate the performance of this approach, we compare it to solving the problem by brute force using grid search. We select two firms at random and choose a single plan for each. We consider a grid of potential coverage levels, taking values in [0.65, 0.7, 0.75, 0.8, 0.85], and evaluate the profitability of every potential tier design while keeping the design of all other plans in the market fixed. In total, each firm evaluates 20,480 configurations. As the effect of a single plan on equilibrium prices and premiums is small, we shut down this channel for this exploration, which vastly reduces the computational cost of grid search. To illustrate its stability, we run our regular convergent solver from 30 random starting points. Figures A.7a and A.7b show the results. On average, our algorithm attains between 95 and 101 percent of the maximum value of the grid search.[4] For insurer 1, the objective curve is slightly smoother, likely contributing to improved performance. Our worst solution attains 98.7 percent of the maximum grid value. For insurer 2, the objective is slightly more jagged, likely contributing to a reduced worst-case performance of 91 percent.[5] Figure A.7c shows the difference in computing times for both approaches. While grid search is about as fast as our approach for a single plan, it is about two orders of magnitude slower for two plans and nearly seven for three plans. When evaluating these results, it is worth remembering that this class of problems is NP-hard; insurers are likely relying on commercial mixed-integer packages to solve these design problems, which are not guaranteed to obtain the optimum. As noted by Murray and Ng (2010), this approach is often as good or better than commercial solvers.

The second challenge involves finding the intersection of firms' best responses. We employ a Gauss-Seidel approach: starting with status quo designs, we identify each firm's best response to the current state and update the state accordingly. We iterate until the change in plan coverages is less than 10^{-5} in the Euclidean norm.

[4]Our algorithm can improve upon the result of the grid because it is not constrained to the same coarse coverage options.

[5]The figures also show that the grid optimum is isolated from other designs. This suggests that firms' best responses might have a unique solution.

To illustrate the performance of the approach, we return to the two single-product firms described above. We solve the equilibrium between the two firms by grid search and through our regularized approach, starting from 30 random points as before. Denoting the equilibrium profit of firm j under the brute-force grid search as π_j^B, and under the regularized approach as π_j^R, we define the relative fit of our approach as $\varepsilon_j = (\pi_j^R - \pi_j^B)/\pi_j^B$. We describe equilibrium fit based on $\varepsilon^* = (\sum_{j=1}^{2} \varepsilon_j^2)^{1/2}$. Figure A.7d shows the distribution of ε_j and ε^* across the different starting points. The relative difference in insurer profits (ε_j) is minuscule, with an absolute value below 0.3 percent for insurer 1 and below 0.1 percent for insurer 2. The equilibrium fit (ε^*) displays an interquartile range below 0.25 percent. Most of the mismatch between the approaches is due to improved best responses found by the regularized approach rather than from meaningful differences in plan design.

Finally, below is an illustration of our full process for one counterfactual equilibrium exercise, which converged in four iterations of best response intersections:

```
Outer [Relaxed][0] Time: 839.7 seconds
 BR: [764.778, 501.524, 210.011, 874.270, 4.082]
 NI: [506.341, 294.806, 89.630, 632.665, −264.353]
 Penalties: [5.329∗10^−14, 3.908∗10^−14, 5.684∗10^−14, 1.279∗10^−13, 1.778∗10^−8]
 Delta prices 10.0
 Delta premiums 2.032∗10^−1

Outer [Relaxed][1] Time: 793.041 seconds
 BR: [748.513, 491.870, 210.517, 862.170, 1.267]
 NI: [748.512, 489.813, 210.513, 862.134, 0.017]
 Penalties: [5.329∗10^−14, 5.329∗10^−14, 6.040∗10^−14, 1.386∗10^−13, 4.718∗10^−4]
 Δ prices 1.815∗10^−4
 Δ premiums 3.250∗10^−4

Outer [Relaxed][2] Time: 670.903 seconds
 BR: [748.501, 492.114, 210.460, 862.138, 1.242]
 NI: [748.501, 492.113, 210.453, 862.138, 1.241]
 Penalties: [4.619∗10^−14, 5.329∗10^−14, 5.329∗10^−14, 1.386∗10^−13, 4.718∗10^−04]
 Δ prices 2.295∗10^−6
 Δ premiums 1.270∗10^−7

Outer [Strict][3] Time: 352.612 seconds
 BR: [747.797, 491.862, 208.375, 853.319, 1.176]
 NI: [747.795, 491.861, 208.375, 853.319, 1.177]
```

Penalties: [$8.171 * 10^{-14}$, $4.263 * 10^{-14}$, $1.741 * 10^{-13}$, $5.933 * 10^{-13}$, $5.149 * 10^{-13}$]

These iterations display several key properties of our approach and the problem. First, they illustrate how convergence occurs through various margins. At each iteration, we present the best response profits (BR) versus the profits implied by all the best responses (NI, for Nikaido-Isoda) for each firm. Formally, for an initial condition vector $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_5)$ and best responses $(\boldsymbol{c}_1^*(\boldsymbol{c}_{-1}), \ldots, \boldsymbol{c}_5^*(\boldsymbol{c}_{-5}))$, BR profits correspond to $\tilde{V}_m(\boldsymbol{c}_m^*(\boldsymbol{c}), \boldsymbol{c}_{-m})$ and NI profits are $\tilde{V}_m(\boldsymbol{c}_m^*(\boldsymbol{c}), \boldsymbol{c}_{-m}^*(\boldsymbol{c}))$, where $\tilde{V}_m$ is the total profit of insurer m net of underwriting cost. A vector of coverages $\boldsymbol{c}^*$ is a Nash equilibrium if and only if BR and NI profits match. The iterations reveal the rate at which this convergence is attained.

Additionally, the iterations show how insurer tiering penalties converge. They indicate that throughout our iterations, all but one firm satisfy the tiering constraint almost exactly. Our algorithm identifies that at the third iteration (marked [Relaxed][2]), the coverage vector is stable, but tiering constraints are not all satisfied. This triggers a new cycle (marked [Strict]) in which any violating coverage vector is adjusted to its nearest tiered configuration, serving as a starting point, and the convergence tolerance on firms' best response solver is substantially increased. As shown, the subsequent iteration results in an equilibrium with insignificant tiering violations. Intuitively, this improvement occurs because the profit objective of the last insurer is nearly flat on its coverage decision, necessitating a shift in its initial condition and convergence criteria to attain the optimum.

Finally, the iterations illustrate how changes to the coverage structure affect the subgame price equilibrium at the end of every iteration. Changes are shown in the Euclidean norm, indicating that as coverage converges, equilibrium prices and premiums also converge. Importantly, this reflects the local stability of the equilibrium: small changes in coverage do not induce large changes in prices or premiums.

It is worth noting that the counterfactual presented above was chosen for its few iterations to demonstrate full convergence. Our main VI ban counterfactual takes considerably longer to solve (approximately 8 hours) but shares the same properties.

D Additional Results

D.1 Moral Hazard and Adverse Selection

Based on our model estimates, we quantify two relevant frictions that challenge market efficiency: moral hazard and adverse selection. Figure A.8a shows the effect of allocative moral hazard on total spending. Holding plan choices fixed, allocative efficiency is achieved when consumers choose hospitals based on cost and quality. Status quo allocations are distorted by price negotiations, coverage choices, and the VI hospital demand shifter. This increases total spending by 47 percent, 8.7 percent of which is attributable to the VI demand shifter. This distortion is roughly equal across VI and non-VI insurers. However, the VI demand shifter nearly doubles inefficient spending at VI hospitals and decreases it at non-VI hospitals. Coverage distortions increase spending valued below cost at all private hospitals. Despite these distortions being relatively larger at VI hospitals, non-VI hospitals are more expensive on average and contribute more to moral hazard spending. Nevertheless, VI plays an important role in spending efficiency, which shapes the welfare impacts of VI we examine in Section 7.

Figure A.8b documents adverse selection by showing the correlation between consumers' willingness to pay (WTP) and expected inpatient cost for each plan (Einav *et al.*, 2010). The correlation is 0.62 overall and 0.44 conditional on market segment and year. Within each market, consumers at the 75*th* percentile of WTP cost 33 percent more to insure than those at the 25*th* percentile, on average. Adverse selection operates through two channels: Riskier consumers benefit more from higher plan generosity and greater access to high-quality hospitals. Making plan networks homogeneous eliminates the second channel, revealing that 27 percent of adverse selection is due to selection on networks (Shepard, 2022). Hence, selection likely affects how insurers choose plan generosity and preferential tiers.

D.2 Counterfactual Analysis under Alternative Premium- and Price-sensitivity

In Section 7.4, we extend our counterfactual analysis for the impacts of banning VI to environments with different impacts of VI on costs and quality. In this section, we do an analogous analysis to extend our analysis to environments where consumers feature different premium- and price-sensitivity. We find that consumer responsiveness to prices and premiums is pivotal for the effects of VI. Theoretically, consumers must be sufficiently elastic to out-of-pocket hospital prices for plan design to steer demand among providers. Additionally, consumers must be sufficiently premium-elastic for VI insurers to attract consumers with low premiums to plans with limited access to rival hospitals. To quantify these forces, we simulate the impact of banning VI under alternative price and premium elasticities. Appendix Figure A.9 shows the results.

For moderate shifts from the baseline, VI enhances welfare when consumers are more price- than premium-elastic and reduces welfare when the opposite is true. Intuitively, when consumers are significantly more responsive to prices than premiums, steering consumers becomes costly. Consumers become very responsive to differences in hospital access, limiting VI insurers' profits from skewing coverage. In this scenario, VI firms lean on their price advantage—a product of the elimination of double marginalization—resulting in welfare improvements. In contrast, when consumers are more responsive to premiums, VI insurers can skew coverage and attract consumers with lower premiums while capitalizing on higher hospital prices. In both cases, the ability of VI firms to capture demand by increasing rivals' costs is limited by insurance competition and the ability of rival insurers to adjust premiums and coverage. Fundamentally, insurers can shift a marginal price increase to where it harms their enrollees less, lowering coverage and premiums when premiums are more important or increasing them when prices are.

When elasticities decrease by more than 50 percent from the status quo estimates, VI almost invariably enhances welfare. Intuitively, consumers must be price-sensitive for steering to be effective. Otherwise, VI firms struggle to direct demand within their networks, limiting their gains from skewing coverage. Consequently, they primarily benefit from eliminating double marginalization, which generally increases welfare.[6]

References

Alvear-Vega, S. and Acuña, M. (2022). Determinantes Sociales que Influyen en el Acceso en Chile al Plan GES, según CASEN 2017. *Revista Médica de Chile*, pp. 70 – 77.

Atal, J. P. (2019). Lock-in in Dynamic Health Insurance Contracts: Evidence from Chile, manuscript.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S. and Zhang, Q. (2018). JAX: Composable Transformations of Python+NumPy Programs.

Canay, I. A., Illanes, G. and Velez, A. (2023). A Users Guide for Inference in Models Defined by Moment Inequalities. *Journal of Econometrics*, p. 105558.

CASEN (2015). *Encuesta de Caracterización Socio-Econḿica Nacional*. Tech. rep., Ministerio de Desarrollo Social.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2019). Inference on Causal and Structural Parameters using Many Moment Inequalities. *The Review of Economic Studies*, **86** (5 (310)), pp. 1867–1900.

Chile Atiende (2024). Reclamo por alza en planes de Isapres. `https://shorturl.at/e2JSF`, accessed: 2024-07-25.

[6]For the U.S., the evidence indicates that enrollees exhibit limited elasticity to out-of-pocket prices, placing them in the region of welfare-enhancing VI. In our model, however, this positive effect of VI occurs because of a breakdown of steering mechanisms. In contrast, consumers in the U.S. are significantly responsive to network structure, all but avoiding out-of-network care. This mechanism could likely substitute for steering through coinsurance rates, acting as if consumers were price elastic in our model.

CIPER (2013). Reajuste de precios de Isapres: una historia de abuso, imposición, desigualdad y lucro. `https://shorturl.at/JySsr`, accessed: 2024-07-25.

Cooper, Z., Craig, S. V., Gaynor, M. and Van Reenen, J. (2018). The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured. *The Quarterly Journal of Economics*, **134** (1), 51–107.

Cutler, D. M., McClellan, M. and Newhouse, J. P. (2000). How Does Managed Care Do It? *The RAND Journal of Economics*, **31** (3), 526–548.

Dias, M. (2022). Selection in a Health Insurance Market with a Public Option: Evidence from Chile, manuscript.

Duarte, F. (2011). Switching Behavior in a Health System with Public Option, manuscript.

Einav, L., Finkelstein, A. and Cullen, M. R. (2010). Estimating Welfare in Insurance Markets using Variation in Prices. *The Quarterly Journal of Economics*, **125** (3), 877–921.

Gale, D. and Nikaido, H. (1965). The jacobian matrix and global univalence of mappings. *Mathematische Annalen*, **159** (2), 81–93.

Galetovic, A. and Sanhueza, R. (2013). Un Análisis Económico de la Integración Vertical entre Isapres y Prestadores, manuscript.

Gowrisankaran, G., Nevo, A. and Town, R. J. (2015). Mergers When Prices Are Negotiated: Evidence from the Hospital Industry. *American Economic Review*, **105** (1), 172–203.

Ho, K. and Lee, R. (2017). Insurer Competition in Health Care Markets. *Econometrica*, **85** (2), 379–417.

Hotz, V. J. and Miller, R. A. (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. *The Review of Economic Studies*, **60** (3), 497.

Johnson, E. M. and Rehavi, M. M. (2016). Physicians Treating Physicians: Information and Incentives in Childbirth. *American Economic Journal: Economic Policy*, **8** (1), 115–141.

La Tercera (2020). Banmédica recibirá a nuevos afiliados con preexistencias y le pone presin a la industria y al regulador. `https://shorturl.at/MWNSU`, accessed: 2024-07-25.

Lindberg, P. O. (2012). A simple derivation of the williams-daly-zachery theorem.

Matzkin, R. L. (2007). Chapter 73 Nonparametric identification. In *Handbook of Econometrics*, vol. 6, Elsevier, pp. 5307–5368.

Morrow, W. R. and Skerlos, S. J. (2011). Fixed-point Approaches to Computing Bertrand-Nash Equilibrium Prices under Mixed-logit Demand. *Operations Research*, **59** (2), 328–345.

Mostrador, E. (2013). Bancada dc oficia a superintendencia de salud para que fiscalice a isapres en cumplimiento de auge. [Online; accessed 2025-05-07].

Murray, W. and Ng, K.-M. (2010). An algorithm for nonlinear optimization problems with binary variables. *Computational Optimization and Applications*, **47** (2), 257–288.

Oliveira, S. C. d., Machado, C. V., Hein, A. R. A. and Almeida, P. F. d. (2021). Relações público-privadas no sistema de saúde do chile: regulação, financiamento e provisão de serviços. *Ciência & Saúde Coletiva*, **26** (10), 4529–4540.

Pardo, C. and Schott, W. (2012). Public versus Private: Evidence on Health Insurance Selection. *International Journal of Health Care Finance and Economics*, **12**, 39–61.

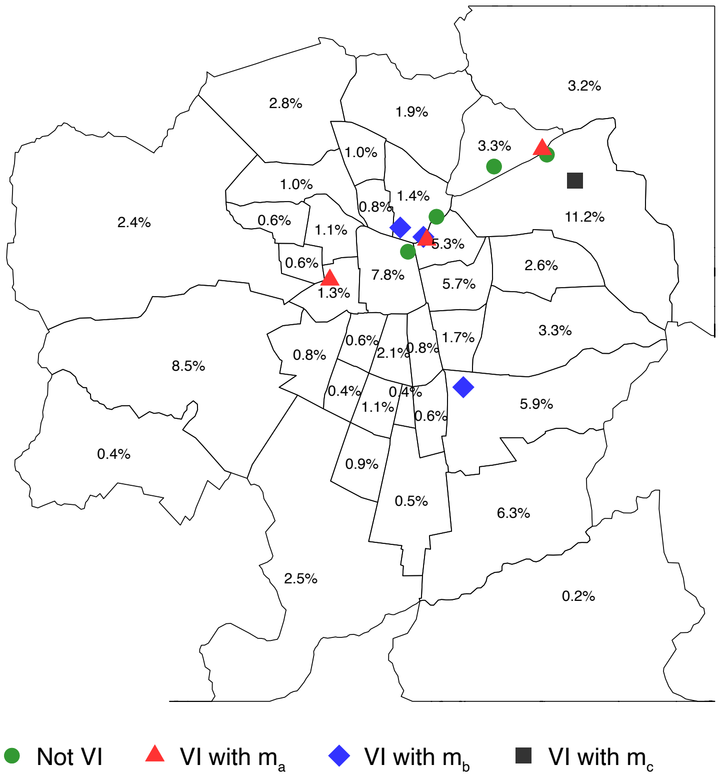
Shepard, M. (2022). Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange. *American Economic Review*, **112** (2), 578–615.

Superintendencia de Salud de Chile (2008). Resolucion exenta numero 476. Accessed: 2025-05-7.

Superintendencia de Salud de Chile (2014). Estándares legales del sistema de isapre. Accessed: 2025-05-7.

Zhang, J., O'Donoghue, B. and Boyd, S. (2020). Globally Convergent Type-I Anderson Acceleration for Nonsmooth Fixed-point Iterations. *SIAM Journal on Optimization*, **30** (4), 3170–3197.

Figure A.1: Location of hospitals and enrollees in the market



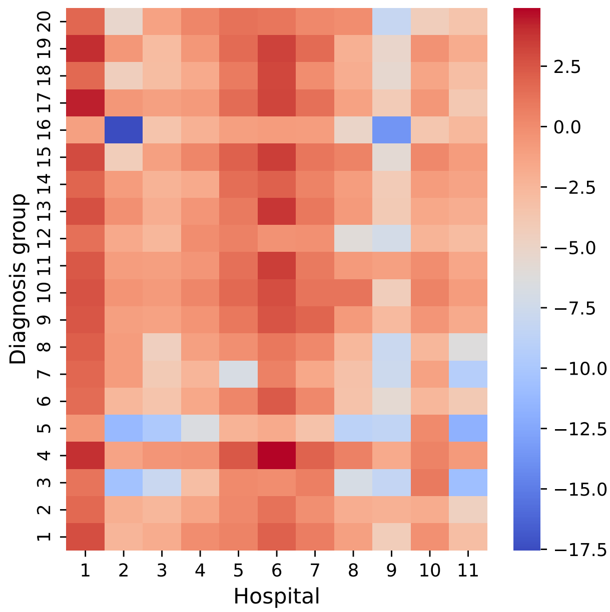
Notes: This figure shows the location of hospitals in the market. The map covers most of urban Santiago, our market of interest. Green circles indicate independent hospitals, red triangles indicate hospitals vertically integrated with insurer m_a, blue diamonds indicate those vertically integrated with insurer m_b, and black squares indicate those integrated with m_c. Finally, numbers indicate the share of enrollees from each county on the map. Some counties located further away are omitted from the plot for convenience, but are included in the analysis. Relative to the full population distribution, there is a noticeable enrollment concentration in the city's wealthier areas (from the center of the figure to the northeast). However, there is clear dispersion in the location of enrollees relative to hospitals.

Figure A.2: Vertical integration, hospital choices, and expenditure (movers subsample)

(a) Plan coverage **(b)** Any claims in VI hospital **(c)** Expenditure in VI hospital

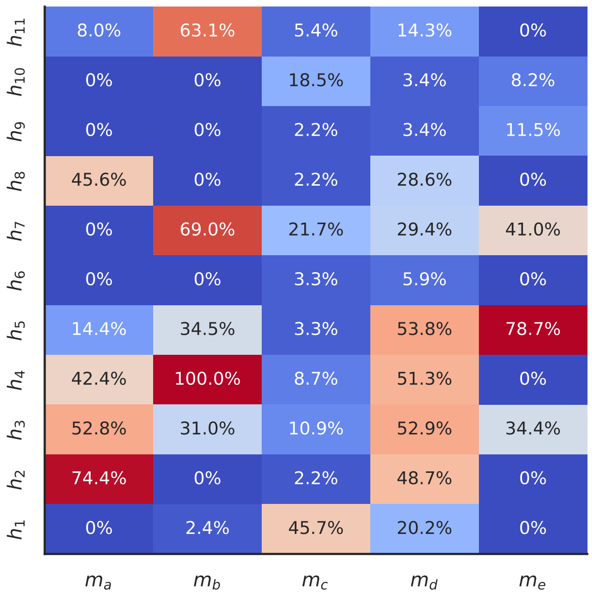
Notes: This figure displays event study estimates from equation (4) in the main text for a subsample of enrollees that move across neighborhoods. This subsample includes 18 percent of the enrollees in the main analysis. The coefficient for the year before the patient switches is set to zero. Green dots and orange squares are estimates of β_τ and γ_τ in equation (4), respectively. Dashed lines indicate 95% confidence intervals. The dependent variable in Figure A.2c is $\log(1 + y)$ to accommodate zeros, but the results are similar when using expenditure in levels.

Figure A.3: Estimates of hospital specialization

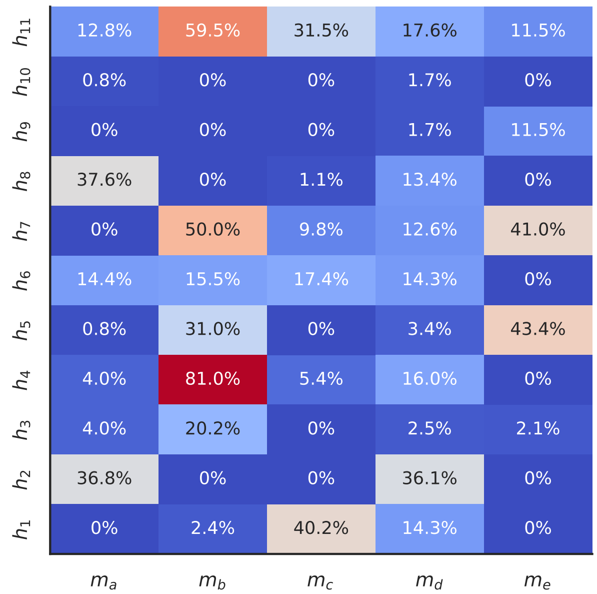


Notes: This figure presents estimates of consumers' average preference for hospital-diagnosis pairs. These correspond to χ^H_{hdt} in equation (5), averaged across years and normalized with respect to the outside option. Diagnosis groups correspond to ICD-10 diagnosis chapters, and hospitals match our anonymized identifiers. Hospitals 1 and 6 correspond to the star hospitals in our data. Hospitals 2, 3, and 8 are integrated with insurer m_a, and hospitals 4, 7, and 11 are integrated with insurer m_b. Hospital 10 is integrated with insurer m_c in the first year of the data.

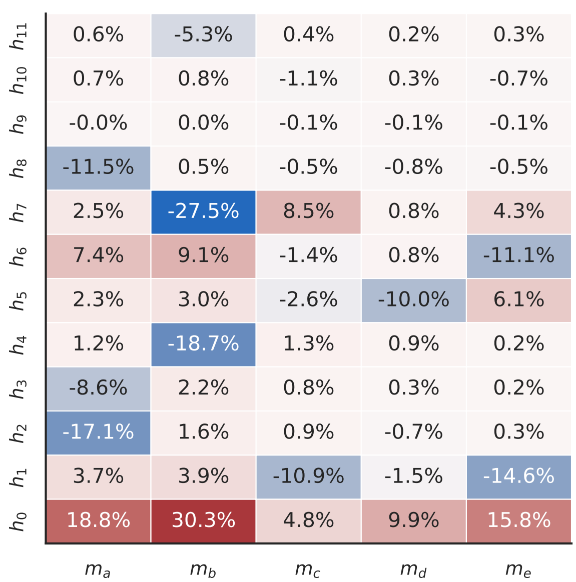
Figure A.4: Counterfactual preferential tier structure, demand, and prices



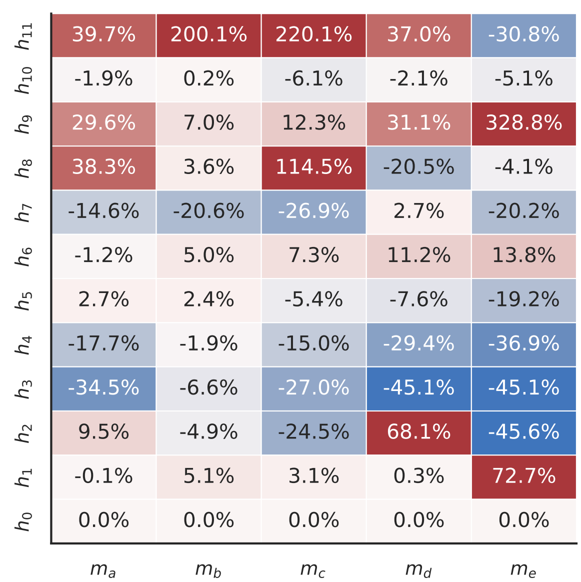
(a) Baseline plan preferential structure



(b) Counterfactual plan preferential structure



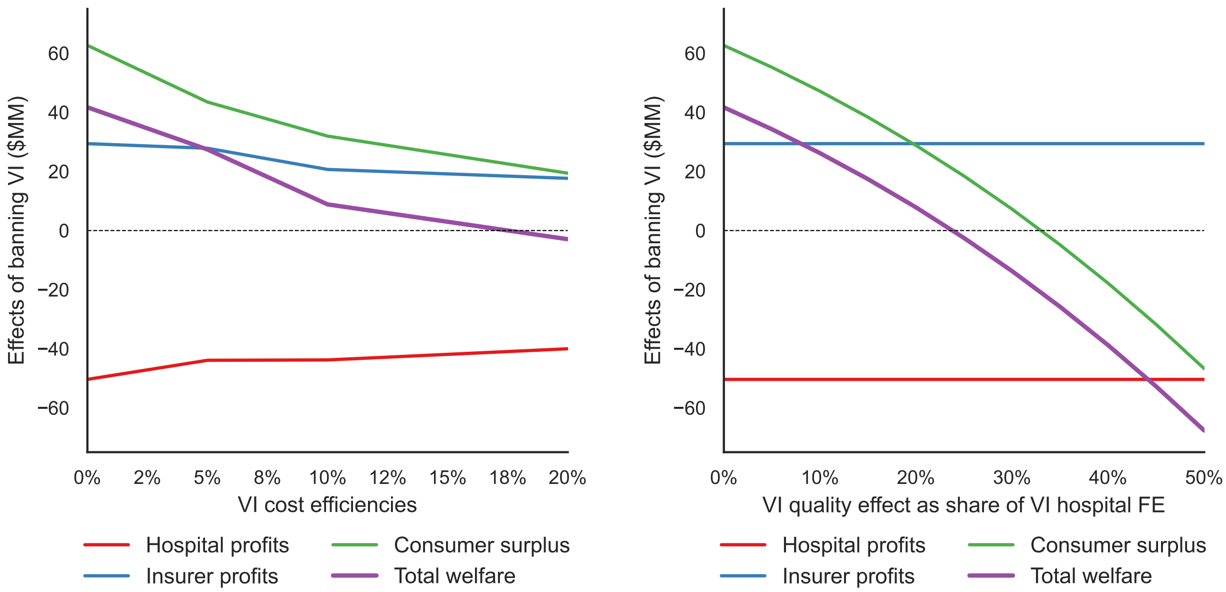
(c) Counterfactual demand change



(d) Counterfactual price change

Notes: Figures (a) and (b) illustrate the number of plans of each insurer (rows) that have each hospital (columns) as preferential. Figure (c) shows the change in hospital demand within each insurer. Changes add up to zero within each column. Figure (d) shows the percent change in negotiated prices between each hospital and insurer. Hospitals h_1 and h_6 correspond to the highest-quality and highest-priced non-VI hospitals in our data. Hospitals h_2, h_3, and h_8 are integrated with insurer m_a, and hospitals h_4, h_7, and h_{11} are integrated with insurer m_b.

Figure A.5: Effects of banning VI under cost efficiencies and quality gains

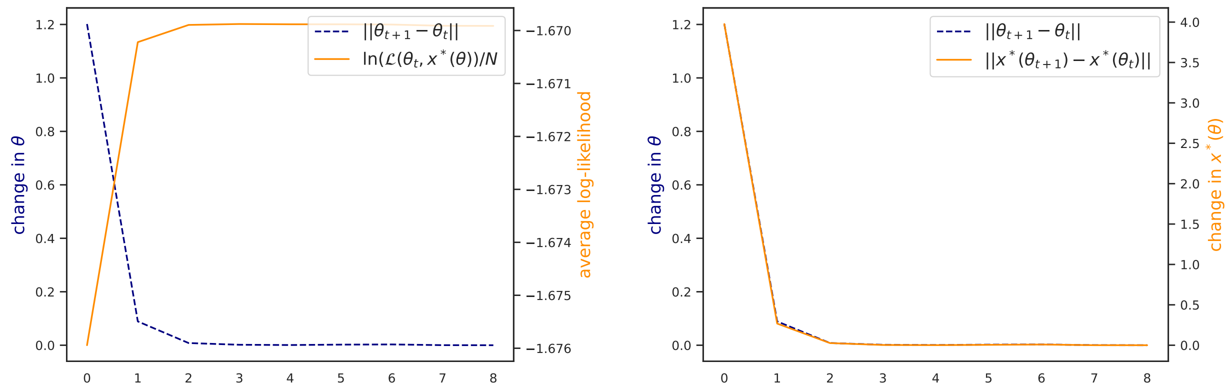


(a) Cost efficiencies

(b) Quality effect

Notes: Figure (a) shows the effect of banning VI in the presence of cost efficiencies. The x-axis represents the percent increase in hospital costs for formerly VI hospitals when serving former partners under the VI ban. Figure (b) shows the effect of banning VI under quality effects. This is computed by assuming that different shares of the VI hospital demand shifter represent true quality differentials. This only affects the baseline surplus value as both distortionary steering (e.g., referrals, marketing) and quality changes disappear in the counterfactual, leading to the same equilibrium.

Figure A.6: Convergence rate of supply-side MLE parameters

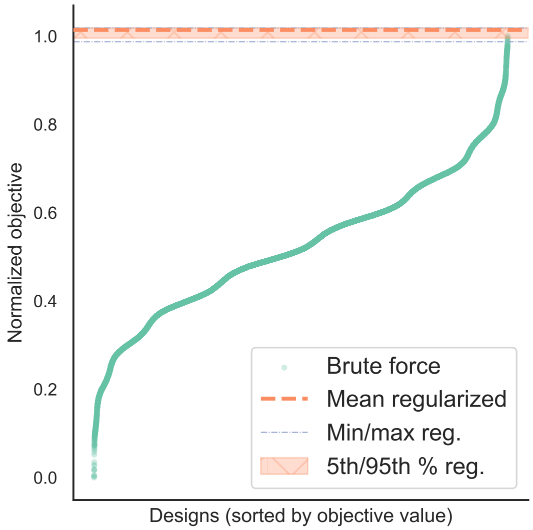


(a) Likelihood

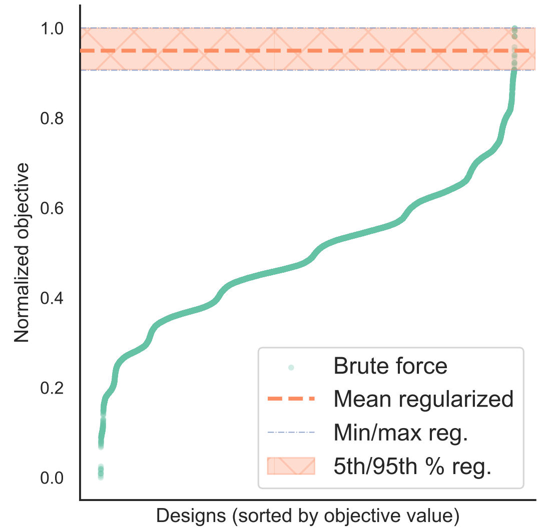
(b) Parameters

Notes: These figures show the convergence rate of model estimates governing price and premium setting. The x-axis shows the iterations of the nested algorithm. In the figures, θ_t stands for the vector of VI weights in the t-th iteration of the solver, and $x^*(\theta_t)$ is the vector of all MLE parameters conditional on the t-th iteration. Figure (a) shows the rate at which the log-likelihood and the VI weights converge. Figure (b) shows that the MLE parameters converge at a nearly identical rate to the VI weights.

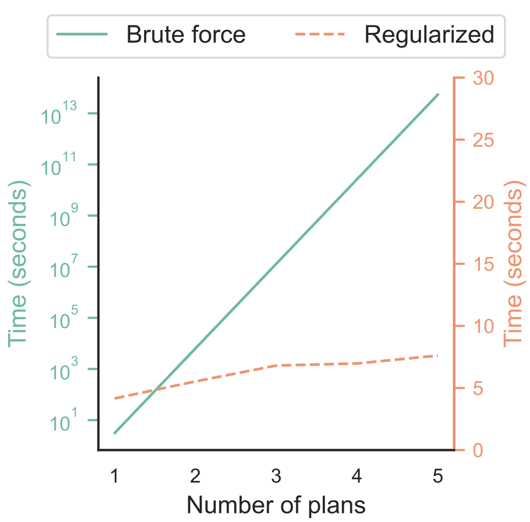
Figure A.7: Performance of plan design solver



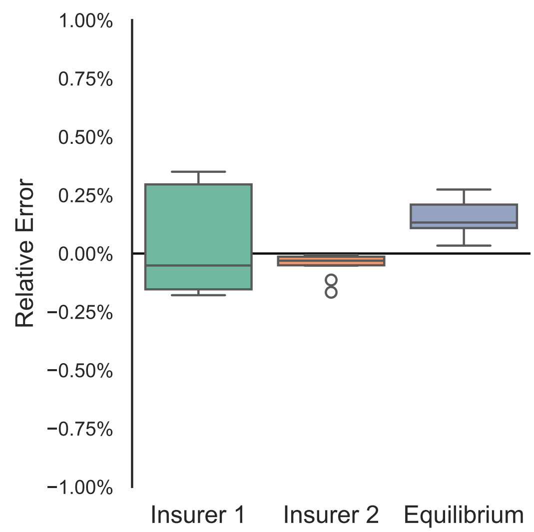
(a) Insurer 1 best response



(b) Insurer 2 best response



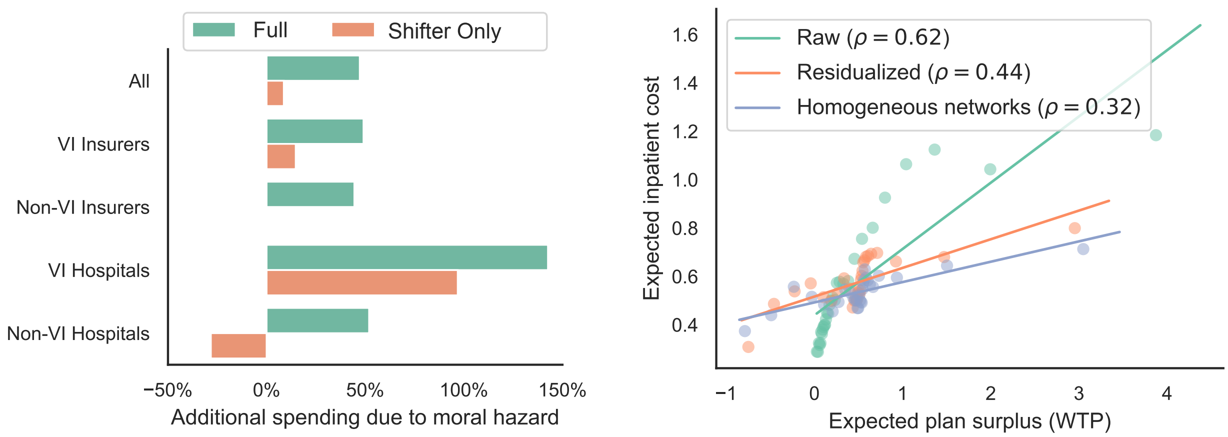
(c) Computing time



(d) Equilibrium fit

Notes: These figures show how the solution of our regularized convergent approach compares against grid-search. Panels (a) and (b) present the best responses of two firms. Green circles show the profit objective value at each of the 20,480 different designs evaluated. We run our regularized solver from 30 random starting points and report the range of results obtained. For ease of comparison, the vertical axis represents the normalized objective relative to the minimum and maximum grid-search value. Panel (c) shows the execution time associated with solving a firm's best response problem for different numbers of plans. The straight green line shows the time in seconds for the brute force grid search approach, inferred from the time it takes to compute a single objective evaluation, scaled by the number of evaluations associated with each problem. The dashed orange line is the mean computation time associated with solving the associated problem for Insurer 1 from five random starting points. All computations were done sequentially on an Nvidia A100 GPU. Panel (d) shows the fit between the equilibrium result derived from our approach relative to that found by grid search. In the first two columns, we plot the distribution of $\varepsilon_j = (\pi_j^R - \pi_j^B)/\pi_j^B$ where π_j^B is the brute-force equilibrium profit of firm j, and π_j^R is the regularized equilibrium profit of the same firm. The third column shows the distribution of $(\sum_{j=1}^{2} \varepsilon_j^2)^{1/2}$ across the 30 starting points.

Figure A.8: Adverse selection and moral hazard

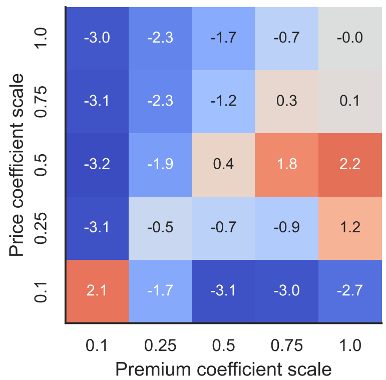


(a) Moral hazard

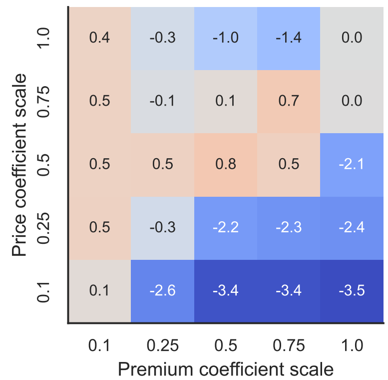
(b) Adverse selection

Notes: Figure A.8a shows the share of additional spending produced by allocative moral hazard relative to a counterfactual scenario in which consumers choose hospitals according to cost and quality, hence removing the effect of coverage, price negotiations, and the VI hospital demand shifter. *Full* bars show the additional spending in the status quo, while the *Shifter Only* bars include only the effect of VI hospital demand shifter. Figure A.8b shows the correlation between consumers' expected network surplus and their inpatient cost at each plan. Drawn circles bin the horizontal axis into 30 quantiles, plotting the mean of cost within each. Lines show the associated regression on the entire sample. *Raw* (green) is computed without controls, *Residualized* (orange) controls for a market-year fixed effect, and *Homogeneous networks* (blue) recomputes the previous by setting each plan's coverage across hospitals and each hospital's price across insurers equal to their mean. This eliminates heterogeneity across plans in access while preserving heterogeneity in coverage generosity.

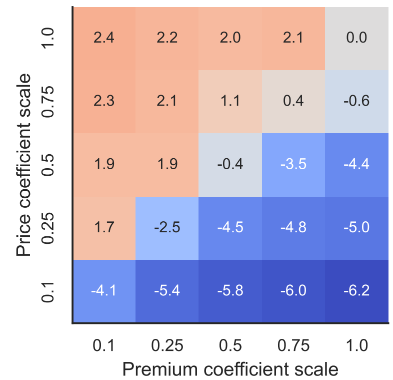
Figure A.9: Effects of banning VI under alternative elasticities



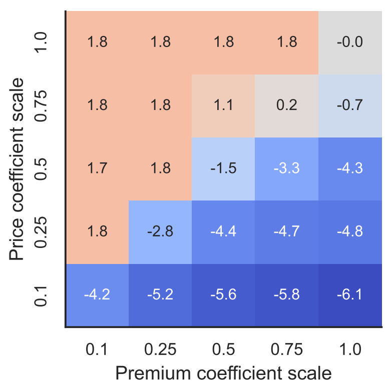
(a) Total welfare



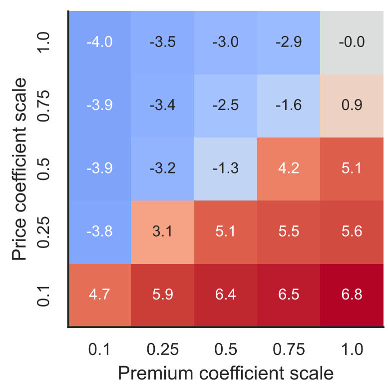
(b) Consumer surplus



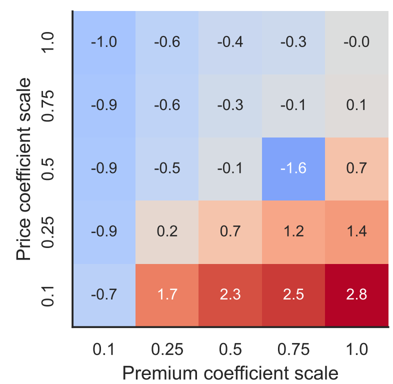
(c) Hospital profits



(d) VI hospital profits



(e) Insurer profits



(f) VI insurer profits

Notes: These figures show the change in welfare from banning VI under alternative price and premium elasticity in log scale relative to the results of the main analysis. Formally, letting (θ_x, θ_y) be the multipliers on the horizontal and vertical axes, and $\Delta TW(\alpha^H, \alpha^M)$ denote the change in total welfare produced from banning VI when consumers' price coefficient is α^H and premium coefficient is α^M, each cell consists of computing $v(\theta_x, \theta_y) = \Delta TW(\theta_x \alpha^H, \theta_y \alpha^M)/|\Delta TW(\alpha^H, \alpha^M)|$ and reporting $sign(v)log(|v|)$. Note that each cell requires computing two full (medium run) counterfactual equilibria, one for the simulated baseline and one for the counterfactual under a VI ban. In both cases, the starting point is the observed status quo to retain our equilibrium notion and consistent treatment of counterfactual simulations. This, however, implies that the VI ban counterfactual in each cell might not be the closest local equilibrium to its baseline comparison. Therefore, we adjust all values relative to our main estimates such that the welfare change in the top right corner equals the simulated baseline.

Table A.1: Descriptive statistics

	N	Mean	SD	p10	p50	p90
A. Policyholder attributes						
Age	3,946,900	39.93	10.37	27.00	38.00	56.00
Single female	3,946,900	0.24	0.42	0.00	0.00	1.00
Single male	3,946,900	0.34	0.47	0.00	0.00	1.00
Number of dependents	3,946,900	0.81	1.17	0.00	0.00	3.00
Income	3,946,900	1.63	1.02	0.00	1.58	2.93
Paid premium	3,946,900	0.17	0.11	0.07	0.15	0.31
B. Plan network structure						
Tiered network plan	1,431	0.88	0.32	0.00	1.00	1.00
Tiered network plan | VI insurer	671	0.90	0.29	1.00	1.00	1.00
Tiered network plan | Non-VI insurer	760	0.86	0.35	0.00	1.00	1.00
Has star hospital in preferential tier	1431.00	0.20	0.40	0.00	0.00	1.00
Has star hospital in preferential tier | VI insurer	671.00	0.19	0.39	0.00	0.00	1.00
Has star hospital in preferential tier | Non-VI insurer	760.00	0.22	0.41	0.00	0.00	1.00
Preferential coverage rate	1,431	77.33	12.79	60.00	80.00	90.64
Preferential coverage rate | VI insurer	671	74.37	11.96	58.32	76.70	89.11
Preferential coverage rate | Non-VI insurer	760	79.95	12.93	63.08	80.00	96.82
Base coverage rate	1,431	59.85	18.11	36.58	60.00	82.97
Base coverage rate | VI insurer	671	58.16	14.70	40.00	56.86	78.81
Base coverage rate | Non-VI insurer	760	61.34	20.55	32.69	64.82	84.74
C. Admission attributes at inside hospitals						
Full price	569,371	4.61	4.90	0.85	3.12	9.79
OOP share	569,371	0.24	0.24	0.00	0.12	0.63
Distance to hospital	569,371	7.47	6.27	0.00	5.96	15.13
Hospital is in preferential tier	569,371	0.64	0.48	0.00	1.00	1.00
Patient is enrollee of VI insurer | VI hospital	308,023	0.61	0.49	0.00	1.00	1.00
Patient is enrollee of VI insurer | Non-VI hospital	261,348	0.39	0.49	0.00	0.00	1.00
Full price at outside option	203,893	1.78	2.61	0.18	1.33	3.78

Notes: This table displays descriptive statistics for our estimating plans dataset. Panel A displays statistics across all policyholders in the sample. Panel B displays statistics for plan attributes across all plan-years in the sample offered in the spot market (i.e., excluding legacy plans held by consumers on guaranteed renewability). Panel C displays statistics across all admissions in the main hospitals included in our analysis unless otherwise noted. All monetary amounts are measured in thousands of U.S. dollars as of December 30, 2014. Distance in miles.

Table A.2: Descriptive statistics for insurers and hospitals

	A. Insurer market shares and premiums					
	Market	Paid premium				
Insurer	share	Mean	SD	p10	p50	p90
m_a	19.87	0.14	0.08	0.06	0.12	0.24
m_b	29.62	0.18	0.11	0.07	0.15	0.32
m_c	17.26	0.20	0.13	0.08	0.17	0.37
m_d	22.59	0.18	0.11	0.07	0.15	0.31
m_e	10.66	0.16	0.10	0.07	0.14	0.29
	B. Hospital market shares and prices					
	Market	Full price				
Hospital	share	Mean	SD	p10	p50	p90
h_1	13.33	6.17	5.79	0.91	4.55	12.73
h_2	4.31	2.56	2.55	0.81	1.93	4.72
h_3	4.25	2.89	3.68	0.62	2.09	5.91
h_4	11.46	3.03	3.88	0.45	2.02	5.95
h_5	10.46	3.92	3.66	0.83	3.29	7.01
h_6	7.87	7.08	6.31	1.45	5.13	14.66
h_7	13.79	4.40	4.61	0.81	3.16	9.04
h_8	2.65	4.03	4.26	0.83	2.83	8.87
h_9	1.24	3.84	4.81	0.32	2.61	7.95
h_{10}	2.26	4.63	4.03	1.05	3.62	8.56
h_{11}	2.72	2.73	2.70	0.55	2.09	5.24
Other	25.66	1.75	2.53	0.18	1.31	3.74

Notes: This table displays descriptive statistics for our estimating admissions dataset. Only admissions to the hospitals in the sample are considered for these statistics. Panel A displays statistics across all hospitals in the sample. Panel B displays statistics for market shares and full prices by hospital.

Table A.3: Preferential tiering and admission flows between hospitals and insurers

	A. Percent of plans in which hospital is in preferential tier					**B. Percent of admissions by patient insurer**				
Hospital	m_a	m_b	m_c	m_d	m_e	m_a	m_b	m_c	m_d	m_e
h_1	0.99	4.30	43.81	15.08	0.75	5.17	30.74	31.42	25.55	7.11
h_2	<u>69.17</u>	3.49	0.00	48.24	0.00	<u>57.13</u>	10.93	7.77	21.49	2.68
h_3	<u>49.21</u>	20.16	6.35	43.97	41.04	<u>63.84</u>	7.31	5.57	21.86	1.43
h_4	46.84	<u>95.70</u>	4.44	53.27	0.00	11.48	<u>70.76</u>	5.21	10.88	1.67
h_5	15.02	45.43	1.59	52.51	81.34	8.40	18.29	22.06	24.28	26.97
h_6	0.00	4.03	0.95	8.29	0.00	4.84	35.81	30.96	20.55	7.84
h_7	0.59	<u>74.73</u>	23.17	31.66	46.27	4.10	<u>56.44</u>	15.86	17.46	6.14
h_8	<u>36.76</u>	0.00	0.32	38.19	0.00	<u>42.93</u>	18.38	16.15	18.51	4.03
h_9	0.00	0.00	0.00	0.50	10.45	19.91	0.73	8.78	69.84	0.74
h_{10}	0.00	0.27	<u>17.46</u>	0.00	9.33	2.91	17.94	<u>58.87</u>	14.86	5.42
h_{11}	6.32	<u>48.92</u>	2.86	7.79	0.00	16.53	<u>66.70</u>	5.99	8.40	2.38
Other	-	-	-	-	-	25.49	13.51	18.67	30.59	11.75

Notes: Panel A displays the share of plans offered in the spot market by each insurer that has a hospital in its preferential tier, with hospitals in the rows and insurers in the columns. The sample includes the set of plans that enter demand estimation and the supply-side analysis. Panel B displays a breakdown of the admission shares. Each cell in columns labeled $m_a - m_e$ displays the share of admissions that a given hospital received from each insurer. Cells that relate to VI firms are underlined. Recall that m_c and h_{10} are only VI in the first year of our sample.

Table A.4: Relationship between patient observables and VI

	(1)		(2)	(3)	(4)
			Regression results		
Patient attribute	Mean non-VI		No controls	Controls	Within I-H
Age	31.468	$\hat{\beta}_{VI}$	2.882	1.126	-0.986
		S.E.	(0.896)	(0.318)	(0.391)
		R^2	0.006	0.290	0.291
Female	0.590	$\hat{\beta}_{VI}$	-0.059	-0.011	0.019
		S.E.	(0.022)	(0.006)	(0.015)
		R^2	0.003	0.240	0.241
Employed	0.851	$\hat{\beta}_{VI}$	0.020	-0.004	-0.012
		S.E.	(0.014)	(0.003)	(0.022)
		R^2	0.001	0.045	0.046
log(Income)	2.069	$\hat{\beta}_{VI}$	-0.170	-0.039	0.018
		S.E.	(0.046)	(0.017)	(0.120)
		R^2	0.001	0.031	0.031
Interacted FE			N	Y	Y
Plan FE			N	Y	Y
County FE			N	Y	Y
I-H FE			N	N	Y

Notes: This table shows results from estimating equation (2) using patient observables as the dependent variable, as indicated in the first column. The unit of observation is an admission. Mean non-VI indicates the mean of the dependent variable for non-VI observations, measured in levels. Each column is a different specification, as indicated in the bottom panel. For each regression, we report the coefficient associated with $\text{VI}_{m(j)ht}$. Interacted FE indicates diagnosis-hospital, diagnosis-year, and hospital-year fixed effects. Standard errors in parentheses are clustered at the insurer-hospital level.

Table A.5: Vertical integration and hospital treatment behavior

	(1)	(2)	(3)	(4)	(5)
	C-section	Ultrasound	Hemogram	Chest X-ray	Imaging
A. Main specification					
VI	-0.112	-0.001	-0.022	0.015	0.003
	(0.030)	(0.001)	(0.020)	(0.009)	(0.002)
Interacted FE	Y	Y	Y	Y	Y
Plan FE	Y	Y	Y	Y	Y
Cost proxy	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y
B. Within hospital-insurer					
VI	0.116	0.003	0.009	-0.017	0.011
	(0.081)	(0.003)	(0.042)	(0.051)	(0.006)
H-I FE	Y	Y	Y	Y	Y
Interacted FE	Y	Y	Y	Y	Y
Plan FE	Y	Y	Y	Y	Y
Cost proxy	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y
N	77,715	77,715	99,265	61,518	61,518
R^2	0.440	0.030	0.096	0.179	0.012
Mean non-VI	0.563	0.003	0.558	0.301	0.018

Notes: This table shows results from estimating equation (2). The unit of observation is an admission. The dependent variable is an indicator of whether a service that relies at least partially on physician discretion was delivered to the patient. These regressions include the following controls: diagnosis fixed effects, admission prices in the public system interacted with hospital dummies, patient age, gender, policyholder income, policyholder employment status, and county fixed effects, along with the fixed effects indicated in the table. Mean non-VI indicates the mean of the dependent variable for non-VI observations, measured in levels. Interacted FE indicates diagnosis-hospital, diagnosis-year, and hospital-year fixed effects. Standard errors in parentheses are clustered at the insurer-hospital level.

Table A.6: Instrument first stage and estimated consumer preferences without instruments

	(1)	(2)	(3)	(4)
	A. Healthcare		**B. Insurance**	
	Coef.	S.E.	Coef.	S.E.
I - Instrument first stage				
Predicted closed-insurer OOP	0.754	(0.000)		
Rival actuarially fair premiums			2.292	(0.042)
N	10,474,104		23,160	
R^2	0.605		0.856	
II - Estimates without instruments				
A: Price (α_i^H) / B: Premium (α_i^M)				
× Age ∈ $[25, 40)$	-1.410	(0.010)	-25.867	(0.051)
× Age ∈ $[40, 55)$	-1.178	(0.019)	-23.613	(0.051)
× Age ∈ $[55, 65]$	-1.139	(0.010)	-24.029	(0.052)
× Female × Single	0.243	(0.010)	10.134	(0.050)
× Has dependents	0.179	(0.009)	13.231	(0.047)
× High income	0.272	(0.005)	11.878	(0.024)
Distance (β^H)	-0.092	(0.001)		
VI demand shifter (γ^H)	2.400	(0.006)		
Network (β^M)			0.921	(0.005)
Median elasticity	-0.67		-1.96	
N	261,857		163,034,142	

Notes: Panel I presents the key first-stage estimates associated with the demand instruments. Panel A presents the regression of consumer out-of-pocket prices for each option within their hospital demand choice set on the instrument. The regression includes distance and year-provider-diagnosis-insurer fixed effects, matching the covariates that are included in the demand model. Panel B presents the regression of plan premiums on the plan's rival actuarially fair premiums. The regression also includes the network utility instruments (average rival network utilities, share of rival plans with the same preferential hospitals, and share of other plans of the same insurer in the same segment with the same preferential hospitals). It also includes insurer-age fixed effects to match our specification of plan preferences. The sample is substantially smaller as this regression is at the plan-year level, as premiums do not vary across consumers conditional on plan. Panel II shows the estimated preferences when estimated without the instruments. Panel A presents estimates of preferences for hospitals. The sample size is a 30 percent random sample of non-emergency inpatient events used to estimate preferences. The model includes insurer-hospital-diagnosis-year fixed effects, omitted from the table. Panel B presents estimates of preferences for plans. The model includes an insurer-year fixed effect, omitted from the table. The heterogeneity in price and premium preferences depends on policyholder attributes, where high income indicates those above the median income. Prices, premiums, and network surplus are measured in thousands of dollars. Network surplus is measured based on yearly risk and spending. Distance is measured in miles from neighborhood centroids to hospitals. The reported elasticities are the median own-price in Panel A and own-premium in Panel B.

Table A.7: Heterogeneity in effects of vertical integration on consumer surplus

	(1)	(2)	(3)
	A. Full effect	**B. Decomposition**	
		Endogenous prices and premiums	+Endogenous plan design
	Change	Change	Change
Policyholder neighborhood			
Center	6.94	-2.85	9.79
North	35.53	-23.74	59.27
Northeast	16.02	-34.46	50.48
Northwest	8.72	-9.90	18.62
Periphery	45.10	-27.13	72.23
South	25.88	-46.30	72.18
Southeast	39.24	-46.61	85.85
Southwest	22.29	-20.59	42.89
Policyholder household type			
Single female	1.27	-8.09	9.36
Single male	5.86	-3.03	8.90
Has dependents	32.57	-45.90	78.46
Policyholder income			
Below median	10.88	-7.47	18.35
Above median	36.23	-80.86	117.09
Policyholder age			
Below 45	9.75	-13.99	23.74
Above 45	20.79	-28.16	48.94
Policyholder insurance program			
Private insurance	28.73	-34.78	63.51
Public insurance	11.19	-17.12	28.32

Notes: This table displays average changes in consumer surplus per policyholder by population groups in dollars relative to the status quo. Panel A displays the Full effect of banning VI. Panel B displays partial changes: Column (2) accounts for changes in hospital prices and premiums keeping coverage fixed, and column (3) shows the additional impact of coverage adjustments. Their sum is the Full effect.

Table A.8: Effects of single-firm vertical integration on plan design and hospital prices

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline		A. Ban m_a's VI		B. Ban m_b's VI	
	Raw	Weighted	Raw change	Weighted change	Raw change	Weighted change
Plan coverage rate (p.p.)						
m_a insurer base coverage	58.52	58.90	-6.08	-8.54	-25.95	-29.15
m_a insurer preferential coverage	81.54	83.46	-0.71	-1.92	3.69	4.52
m_b insurer base coverage	43.63	43.07	-13.60	-13.17	-9.36	-7.73
m_b insurer preferential coverage	73.12	72.51	13.12	9.52	11.26	9.00
Non-VI insurer base coverage	58.88	58.67	-14.26	-14.75	-19.95	-18.74
Non-VI insurer preferential coverage	85.75	84.94	0.78	0.91	-1.57	-1.77
Plan preferential tiering (p.p.)						
m_a insurer self-preferencing	57.60	53.04	-31.73	-30.35	0.27	19.83
m_a insurer other-VI-preferencing	16.80	14.78	-12.53	-6.73	10.40	12.75
m_a insurer non-VI-preferencing	2.88	2.45	0.00	-0.87	8.00	1.22
m_a insurer star-hospital-preferencing	0.00	0.00	6.80	3.76	27.20	9.18
m_b insurer self-preferencing	77.38	77.82	2.78	9.39	-8.33	-15.79
m_b insurer other-VI-preferencing	10.32	11.73	-10.32	-11.73	-0.40	0.32
m_b insurer non-VI-preferencing	7.38	6.27	30.71	30.65	2.38	-1.01
m_b insurer star-hospital-preferencing	1.19	0.68	94.05	91.63	7.14	3.80
Non-VI insurer other-VI-preferencing	22.12	21.40	-8.00	-7.84	-8.62	-8.00
Non-VI insurer non-VI-preferencing	16.91	17.98	-6.36	-9.28	-6.87	-8.51
Non-VI insurer star-hospital-preferencing	13.97	18.22	2.39	-4.22	2.02	-2.47
Hospital prices ($M)						
Within m_a firm	3.43	3.43	0.27	-0.18	16.18	13.49
m_a-VI hospital to m_b insurer	4.07	4.07	-0.16	-0.25	-0.19	0.01
m_a-VI hospital to non-VI insurer	3.77	3.77	-0.29	-0.25	-0.47	-0.48
Within m_b firm	3.88	3.88	-0.79	-0.80	-0.22	-0.30
m_b-VI hospital to m_a insurer	3.59	3.59	1.02	-0.01	45.52	0.92
m_b-VI hospital to non-VI insurer	3.61	3.61	1.17	-0.15	0.49	-0.54
Non-VI hospital to m_a insurer	5.71	5.71	0.16	0.09	1.95	2.74
Non-VI hospital to m_b insurer	6.10	6.10	0.09	1.16	0.55	0.96
Non-VI hospital to non-VI insurer	5.44	5.44	1.10	0.24	1.83	0.76
All hospitals	4.32	4.32	0.45	-0.03	4.72	0.64
Plan premiums ($M)						
Insurer m_a	1.11	1.11	0.11	0.07	0.02	-0.53
Insurer m_b	1.40	1.40	0.55	0.49	0.17	0.10
Non-VI insurer	1.38	1.38	-0.02	-0.08	-0.07	-0.12
All insurers	1.31	1.31	0.12	0.09	-0.00	-0.18

Notes: Plan coverage and tiering in percentage points, prices in thousands of dollars per unit of resources, premiums in thousands per year. VI insurer self-preferencing is the likelihood that a VI hospital is preferential in a VI plan. Other-VI and non-VI are analogous to other-VI and non-VI hospitals. Odd columns display raw averages: for prices, it is across insurer-hospital; for premiums and coverages, it is across plans. Even columns display weighted averages by demand: for prices, it is by demand per unit of hospital resources; for premiums and coverage, it is by plan demand. Panels A and B display the Full effect of banning VI for m_a and m_b, respectively.

Table A.9: Effects of single-firm vertical integration ban on choices and welfare

	(1)	(2)	(3)
		A. Ban m_a's VI	**B. Ban m_b's VI**
	Baseline	Change	Change
Hospital market shares (p.p.)			
m_a-VI hospital	11.11	-5.44	3.93
m_b-VI hospital	27.56	-4.45	-12.90
Non-VI hospital	30.95	3.28	-9.85
Insurer market shares (p.p.)			
Insurer m_a	3.69	-0.84	2.04
Insurer m_b	7.15	-1.50	-2.39
Non-VI insurer	13.18	1.50	1.66
Admission shares (p.p)			
Within m_a	55.84	-43.71	5.03
m_a-VI / m_b	6.42	-1.29	3.43
m_a-VI / Non-VI	37.74	45.00	-8.45
Within m_b	78.51	-14.37	-52.81
m_b-VI / m_a	2.69	1.08	1.77
m_b-VI / Non-VI	18.80	13.29	51.05
Non-VI / m_a	9.96	0.15	0.95
Non-VI / m_b	12.99	16.31	12.44
Non-VI / Non-VI	77.05	-16.46	-13.39
Healthcare spending			
Inpatient spending | private plan ($M)	1.04	0.01	0.20
Inpatient spending ($M)	0.50	-0.00	0.06
Total household spending ($M)	1.10	0.04	-0.06
Actuarial value (%)	0.66	-0.03	-0.02
Insurance market efficiency			
Spending allocative inefficiency (%)	44.10	-2.26	-14.72
Consumer surplus			
m_a enrollees (per member, $M)	–	0.34	0.67
m_b enrollees (per member, $M)	–	-0.03	-0.23
Non-VI enrollees (per member, $M)	–	5.77	7.54
Total consumer surplus ($MM)	–	257.418	154.284
Share better off	–	0.603	0.806
Share better off relative to m_a ban	–	-	0.860
Profits ($MM)			
m_a-VI hospitals	34.90	-17.37	385.30
m_b-VI hospitals	70.14	-45.47	-39.67
Non-VI hospitals	109.57	10.94	23.42
m_a insurer	159.95	-25.29	-455.20
m_b insurer	394.73	109.74	-103.42
Non-VI insurers	713.82	58.78	79.48
Total welfare (in millions)	–	348.74	44.18

Notes: Shares in percentage points. Healthcare spending in thousands per household. Actuarial value is the share of expected payments covered by insurers. Spending allocative inefficiency is relative to the first-best inpatient spending. Consumer surplus for VI enrollees is the average surplus among VI plan enrollees, unweighted by demand. Non-VI consumer surplus is defined analogously. Profits and total consumer surplus are measured in millions of dollars per year. Panels A and B display the Full effect of banning VI for M_a and m_b, respectively.