Retention by Design: Long-term Contracts in Health Insurance

Valentina Reig*
Toulouse School of Economics

This version: October 2026
Please click here for the latest version

Abstract

This paper studies how regulation designed to protect consumers from rising premiums as their health worsens shapes competition and welfare when insurers have market power. In Chile, insurers commit to long-term premium schedules, while consumers remain free to switch plans. Using administrative data on plan choices and medical claims, we document that healthier individuals are more likely to switch plans. Leveraging demographic differences across groups of similar plans, we find that newly introduced plans help insurers retain switchers. To quantify how consumer switching and insurers' retention strategies jointly affect prices and welfare, we develop and estimate a supply and demand model in which insurers set prices upon introduction, anticipating future enrollment, costs, and own and rivals' plan offerings. In counterfactuals, we find that banning new plan introductions reduces insurers' costs but raises premiums by 13%, reducing the number of enrolled individuals. In a second counterfactual, in which both consumers and insurers commit to their plan for life, costs fall but consumer surplus is lower as individuals lose the flexibility to adjust coverage when their health needs change. Our findings show that policies can reduce insurers' costs without making consumers better off.

1 Introduction

Many markets feature long-term contracts between firms and consumers: phone plans fix a price for two years, energy suppliers discount fixed-term commitments, and insurers guarantee coverage terms for years. These contracts are valuable when future terms cannot be guaranteed by re-contracting: the buyer secures prices in advance, and the seller retains a stable customer (Rey and Salanie (1990)). But commitment is often one-sided: the firm cannot worsen the terms of an existing contract as it learns about each consumer, but the consumer can leave. As consumers learn about their own needs, those for whom the contract has become a bad deal leave, and this selective exit worsens the pool of consumers who stay. In perfectly competitive markets, previous work has characterized the properties of these contracts (Harris and Holmstrom (1982), Hendel

*I am grateful to Pierre Dubois for the guidance on this project. I also thank Angelique Acquatella, Léa Bignon, Matteo Bobba, Christian Bontemps, Alexandre de Cornière, Paul Diegert, Ben Handel, Bruno Jullien, Patrick Rey and Chuqing Jin for valuable comments and discussion.

and Lizzeri (2003), and Herring and Pauly (2006)). The question of how firms respond to selective exit in settings with market power and what their response means for consumers is less examined.

This paper studies how insurers price and retain individuals in long-term contracts in health insurance. Long-term contracts protect individuals against *reclassification risk*, the risk that premiums rise after a bad health shock, by specifying a premium schedule in advance (Ghili, Handel, Hendel, and Whinston (2024)). However, commitment is enforced only on the insurer's side, so healthier individuals often leave as their alternatives improve once their health is revealed. This generates *dynamic adverse selection*: as low-risk individuals leave, the average cost of the plan increases. With perfect competition, the literature studies front-loaded premiums, charging premiums above expected cost early on to finance protection later, as the response to this problem (Hendel and Lizzeri (2003) and Ghili, Handel, Hendel, and Whinston (2024)). Yet many health insurance systems rely on managed competition, in which a few insurers offer differentiated plans under price regulation and exert market power by charging consumers a mark-up over marginal costs. When regulation also sets a plan's price at introduction, insurers cannot reduce the premium on that plan to retain healthier individuals once their health is revealed, but they might use an alternative retention strategy: introduce new plans to retain healthier individuals who might otherwise switch to competitors. We empirically investigate this strategy in a regulated setting with market power.

How do new plan introductions affect pricing, retention and consumer welfare? We study these questions in Chile's private health insurance market, one of the few settings where long-term contracts and imperfect competition among insurers coexist. The market consists of six competing insurers offering contracts that specify coinsurance rates, maximum reimbursement amounts, and a base premium that is fixed at introduction[1]. Commitment is one-sided: insurers cannot reprice individual contracts or cancel coverage, but consumers can switch plans at any time[2]. Some individuals switch, so insurers earn margins on the individuals who stay. Those who switch are mainly low-risk, which creates dynamic adverse selection.

Under perfect competition, as in the front-loaded contracts of Hendel and Lizzeri (2003), Ghili, Handel, Hendel, and Whinston (2024), and Atal, Fang, Karlsson, and Ziebarth (2025), insurers earn zero expected profits over the contract. Individuals pay more than current expected cost early on, and these payments finance protection against reclassification risk. In Chile, this front-loading mechanism is limited, and insurers have market power. The price of a plan is set at introduction and then follows a regulated age-gender weight[3]. With market power, insurers can also finance part of this protection from positive margins, reducing their reliance on front-loading through prices paid by entrants, who are more responsive to price. Low-risk individuals carry the largest margins and are also the most likely to leave. Under perfect competition, there is no positive margin to preserve when these individuals leave for a better offer, but here insurers have

[1]The base premium is adjusted over time only according to regulated rules. Our model treats the base premium of a plan as fixed at introduction.

[2]There are constraints for switching, including pre-existing conditions. Details are in Section 2.

[3]Unlike in Germany, nothing an individual pays builds up as a personal balance that carries over from year to year or determines the price they will pay in the future (Atal, Fang, Karlsson, and Ziebarth (2019)). In Herring and Pauly (2006), guaranteed renewable contracts also have fixed prices.

an additional reason to retain them. The response we document with the data is that insurers introduce new plans to keep them within their portfolio. How insurers set prices in this market is therefore an empirical question: how large markups are relative to average cost, and how much of the price set at introduction reflects the margins the insurer expects to recapture when individuals move to its own future plans.

We use detailed administrative data from Chile's health insurance regulator (*Superintendencia de Salud*) covering 5 million enrollees between 2016 and 2022, including plan characteristics, individual demographics, and medical claims. A distinctive feature of this market is that insurers maintain thousands of plans but regularly introduce new ones while closing older ones, generating rich variation in the choice set consumers face over time. Rich information on individuals' medical expenditures observed in different health insurance plans with different prices allows us to measure how the medical costs of the individuals who respond most to prices differ from those of the average individual, a *selection wedge channel*. Other data sources include insurers' financial statements, enrollment and cost data from the public option (*FONASA*), the 2017 CASEN household survey, and health cost growth series from the regulator and the medical costs CPI from the Central Bank of Chile.

We begin by documenting four empirical patterns. First, low-risk individuals are more likely to switch plans, contributing to a deterioration of the risk pool in the plans they leave. Second, plans with higher exit among low-risk enrollees have higher average costs in subsequent years, consistent with dynamic adverse selection. Third, we show that the introduction of new plans increases within-insurer retention. We measure the entry intensity a plan faces as the similarity-weighted sum of new plans introduced by the same insurer. Since insurers choose entry while observing retention, we instrument it with a shift-share design that interacts each plan cell's 2016 demographic composition with national cost growth by demographic group. The assumption is that, absent new plans, enrollees in cells with different compositions would have been equally likely to stay with their insurer over time. This is plausible because enrollees choose plans based on their characteristics, not on who else is enrolled. Moving from zero to mean entry intensity raises the probability that a switcher stays with the same insurer by 4 percentage points, consistent with plan introduction serving a retention role. Lastly, we provide descriptive evidence on the price-cost margins insurers earn from individuals who switch plans and those who stay in their current plan. Switchers who stay within the same insurer pay on average lower prices in the destination plan, and the insurer earns a lower margin on them than in their previous plan. Together, these patterns suggest that insurers use new plans to retain low-risk individuals, and accept lower margins to do so.

These patterns show that retention matters how insurers compete in this market. But how does new-plan introduction affect prices, consumer choices, and insurer profits? We develop and estimate a model of supply and demand to assess the incentives insurers face when setting prices and to quantify the consequences of alternative market rules. On the demand side, individuals choose among plans based on prices, expected out-of-pocket costs and their variance (through which risk aversion enters), plan characteristics, and a switching cost that generates inertia. Price

sensitivity varies with age and income. On the supply side, insurers set the price of a plan once, when they introduce it, and take into account how today's enrollment affects future enrollment and costs.

Estimating demand requires separating preferences from inertia, since individuals who rarely switch may either value their plan or face high switching costs. Individuals entering the private system have no previous plan, so their choices reveal how they value plans, and the choices of incumbent enrollees then reveal switching costs. We estimate the model with the two-step method of Berry and Haile (2024), which combines maximum likelihood estimation of individual coefficients with two-stage least squares to address price endogeneity. The estimates show high inertia, and that older individuals are more price sensitive and higher-income individuals less so.

The supply side poses the main challenge. Since the price of a plan applies for its whole life, the insurer must forecast how many enrollees the plan will keep, how costly they will be, and which plans its rivals and the insurer itself will introduce. Inertia makes this forecast dynamic: a lower price attracts enrollees today and keeps them in the plan for many years, so the effect of the price on profits is a sum over the full future path of enrollment, not a single elasticity. The cost side has the same structure: those who leave are lower-risk than those who stay, so the price also changes who remains in the plan. With six insurers maintaining thousands of plans and introducing new ones every year, a dynamic game in which each insurer tracks every rival plan, its price, and its pool composition is infeasible to solve or estimate. We exploit the pricing regulation, which makes this problem tractable: since the price of an existing plan cannot change, the only future decisions the insurer must anticipate are plan introductions, its own and its rivals'. Rather than solving the game, we estimate how plans gain and lose enrollees as a function of plan age and require each observed price to satisfy the insurer's first-order condition given these forecasts.

We use the model for three analyses. First, we ask how much of the observed prices our model can explain. In typical structural IO models, a first-order condition is used to recover marginal costs. Here, we observe individuals' costs and, therefore, infer what the insurer forecasts to set prices. Second, we analyze what happens if we eliminate the retention instrument and do not allow for introduction of new plans. The third analysis asks whether two-sided commitment, which achieves the first best under perfect competition and forward-looking individuals (Ghili, Handel, Hendel, and Whinston (2024)), also benefits individuals when insurers have market power. We compare switching with two-sided commitment, which removes dynamic adverse selection but locks individuals into the plan they first choose.

We rationalize observed prices using the dynamic first-order condition of the supply model. Because the panel includes plans introduced before our sample begins, we can estimate how state variables transition over time and construct the laws of motion the insurer uses to forecast a plan's future. The first-order condition is approximated with a future term that depends on how the market structure changes as rivals and the insurer itself introduce new plans. This term combines the estimated transitions with derivatives from the structural model. The transitions describe the composition of enrollees entering a plan, from outside the system, from rivals, or from the same insurer. We find that for the median plan, 58% of the base price comes from the future terms,

which capture profits after the plan is introduced and priced, and about one third of this future component comes from margins the insurer expects to recapture through its other plans.

The second analysis compares the status quo, in which insurers continue to introduce new plans, with a counterfactual in which no new plans are introduced in the future. We model future plan introductions as one new option per insurer in each period after the plans we price enter the market. These new options attract both enrollees from existing plans and individuals entering the private system, and disproportionately attract healthier individuals. When future plan introductions are not allowed, those who leave for other existing plans or the outside option are, on average, less healthy than those who would have switched to new plans. This changes risk selection because the individuals most responsive to prices are no longer the healthiest. We compute the first-order condition for plans introduced in 2018, taking into account that no new plans will be introduced in the future while holding the prices of older plans fixed. We find that prices increase when future plan introductions are not allowed. The first-order condition decomposes the price of a plan into the average cost of its pool, a selection wedge, a markup, and the margins the insurer expects to recapture on its other plans. The decomposition shows that average cost falls, while the selection wedge and markup together increase because those who leave are less positively selected[4]. Under perfect competition, the reduction in average cost would translate into lower prices. With market power, prices instead increase by 13% , raising exit to the public system by 10 percentage points approximately.

In the third analysis, we use the estimated model to compare a scenario with switching, where consumers can switch plans each period, with a two-sided commitment counterfactual in which individuals choose a plan once upon entering the private sector and remain enrolled for life. The choice set is fixed in both cases and contains a large number of observed plans that represent the variety of the industry. We find that switching provides higher consumer surplus than commitment by US$235 (5.5 UF[5]) per year per individual, approximately two months of premiums for a 26-year-old. Decomposing this difference reveals that the flexibility value of re-optimizing as health needs change is large, but it is partially offset by a price effect that works against consumers under the switching scenario. The price effect operates almost entirely through a market-power channel rather than through higher costs from a deteriorated risk pool. Under switching, consumer inertia gives insurers a partial source of market power over existing enrollees, under commitment, insurers must compete for the entire discounted lifetime value of the individual at the moment of entry, pushing markups below switching levels. Because Ghili, Handel, Hendel, and Whinston (2024) assume full insurance, switching plans has no value in their framework, but in ours, plans differ in coverage, and we measure the value of switching against the market power it creates.

Related literature. Our analysis connects to the literature on plan design and risk selection in health insurance. Existing work documents how insurers shape the composition of their enrollee pool through the design of their products: differences in drug formularies (Geruso, Layton, and

[4]The selection wedge is the difference between the marginal cost of individuals who respond to price and the average cost of the pool. The sum of the selection wedge and the markup equals the conventional markup scaled by the response of average cost to price, as in Mahoney and Weyl (2017).

[5]UF is the inflation-indexed unit of account used in Chile, mainly for financial contracts (1 UF corresponds to US$43 in 2018).

Prinz (2019)), coverage generosity (Cao and McGuire (2003) and Carey (2017)), and network composition (Shepard (2022) and Serna (2025)) all serve as screening devices that attract favorable risks. A large number of plans can amplify opportunities for selection-driven differentiation (Veiga and Weyl (2015) and Chade, Marone, Starc, and Swinkels (2022)) and make consumer decision-making more difficult (Ericson and Starc (2016)), motivating standardized plan categories in ACA Marketplaces (Handel and Kolstad (2022)) that can strengthen price competition (Ericson and Sydnor (2017) and Handel, Kolstad, and Spinnewijn (2019)). Cuesta and Tebaldi (2025) quantify this tension, comparing a procurement auction that reduces variety but limits selection with a regulated oligopoly that offers more variety but suffers more from selection. Outside insurance, Bourreau, Sun, and Verboven (2021) show firms in the telecom industry introduce new products strategically to respond to competition. This literature focuses on plan design as a static screening instrument. We study a dynamic counterpart: insurers in our setting introduce new plans not only to attract favorable risks, but to retain them over time, using product entry as a strategic response to the compositional deterioration that one-sided commitment generates in existing plans.

We also contribute to the literature on long-term contracts in insurance. Hendel and Lizzeri (2003) show that in the US, life insurance contracts with more front-loading lose fewer consumers, because their future prices are better than the spot contracts available to them. Ghili, Handel, Hendel, and Whinston (2024) characterize the optimal long-term health insurance contract and show that it eliminates most of the welfare loss from reclassification risk, and Atal, Fang, Karlsson, and Ziebarth (2025) show that German contracts, which front-load through individual old-age provisions, achieve similar welfare to the optimal contract, although the observed front-loading is too large when there are rising income paths resulting in welfare losses of at most 4%[6]. We build on this work by studying these contracts in a setting with market power and multiproduct firms, in which insurers have an additional incentive to retain healthy individuals because they generate positive margins[7], and multiproduct insurers can do so by moving them across plans within their own portfolio. Rather than quantify the welfare gains from protection against reclassification risk, we study insurers' pricing strategies and how higher prices distort the provision of private insurance. We characterize a long-term contract that differs from the German system. Atal (2019) also studies the long-term contracts setting in Chile, analyzing how detrimental insurer lock-in is for individuals, but we highlight the dynamic incentives for pricing, and provide evidence that insurers use plan introduction as a retention tool. Aizawa and Ko (2025) study US long-term care insurance, where insurers can reprice contracts after aggregate shocks, and find that limiting repricing increases concentration with little benefit to consumers. In our setting, prices cannot be revised after introduction, so the insurer must account for the whole future of a plan, including

[6]In Germany, this is possible because the system allows for an old-age provision that is individual specific and therefore premiums in the future are a function of what was paid over the lifecycle. Front-loading effects can change after the portability policy, see Atal, Fang, Karlsson, and Ziebarth (2019). Chile has no equivalent individual prepayment. Premiums depend only on the plan and on the enrollee age and gender, so any front-loading operates through the regulated age gradient rather than through accumulated individual reserves.

[7]In both settings, low-risk individuals are more likely to face attractive outside offers. Under perfect competition, when such an outside offer becomes relevant, the incumbent can match it at zero continuation profit and is therefore indifferent between retaining and losing the individual. With market power, retaining the individual also preserves a positive margin.

the risk revealed within its pool, in a single price.

Our supply model relates to the literature on firm competition under consumer inertia (Dubé, Hitsch, and Rossi (2010), Ericson and Starc (2016), Brown, Egan, Jeon, Jin, and Wu (2025), and Mackay and Remer (2026)). Fleitas (2017) studies how dynamic pricing affects consumer welfare in Medicare Part D, where firms set low initial prices to build market share and raise them over time, exploiting the fact that inertia discourages switching (Klemperer (1995) and Farrell and Klemperer (2007)). Our setting features a key institutional difference: prices are fixed at the time of plan introduction and cannot be adjusted afterwards. Insurers therefore face a similar trade-off under a single decision: rather than exploiting inertia through dynamic repricing, they must set a single price that balances the need to mitigate dynamic adverse selection while harvesting the non-responsive, with the need to keep prices low enough to attract young and healthy enrollees. This constraint gives rise to a pricing first-order condition in which the optimal price internalizes the derivatives that propagate current enrollment into all future shares through inertia, distinguishing our framework from both standard static Bertrand models and dynamic pricing models such as those of Fleitas (2017) or Brown, Egan, Jeon, Jin, and Wu (2025), whose dynamic pricing collapses to a steady-state.

Our supply model also relates to the literature on estimating dynamic games. Insurers decide which plans to introduce and their prices. We model only the price as the strategic choice, while plan introduction follows a stochastic process. Insurers hold beliefs about how the market will evolve, including the process of plan introduction, and these beliefs determine the price they set at introduction. We follow the two-step logic of Hotz and Miller (1993), Bajari, Benkard, and Levin (2007), and Pakes, Ostrovsky, and Berry (2007): in a first stage we estimate how the market objects move over time from the observed data, and in a second stage we impose that each observed price satisfies the firm's first-order condition given those estimated processes, without solving the full game. Many dynamic-pricing papers compute rivals' best responses in the future. What is convenient from our setting is that the price of an existing plan cannot change, so the only future actions the insurer needs to forecast are future plan introductions, the insurer's own and its rivals'. The insurer's beliefs therefore reduce to a process for plan introductions, which we estimate from the data under a stationarity assumption. Insurers forecast future plan menus based on a coarse state vector (Weintraub, Benkard, and Van Roy (2010) and Ifrach and Weintraub (2017)) and, following the experience-based equilibrium of Fershtman and Pakes (2012) and Dubois and Pakes (2026), beliefs are the estimated transitions themselves, required to be consistent with the outcomes on the states the data visit. Counterfactuals that remove plan introductions change these beliefs, so we re-solve prices by simulating the lifecycle of each plan under the structural first-order condition.

The rest of the paper proceeds as follows. Section 2 describes the Chilean institutional setting and the data. Section 3 documents the empirical patterns of switching behavior, dynamic adverse selection, and the effect of plan introduction on retention. Sections 4 and 5 develop and estimate the supply and demand model. Section 6 presents two counterfactuals: no introduction of new plans in Section 6.1 and two-sided commitment in Section 6.2. Section 7 concludes.

2 Institutional Framework and Data

Health insurance systems around the world aim to provide accessible health care to their citizens. Most countries combine a public option with a private market. While the US is characterized by managed competition insurance, France covers the whole population with a public option, offering private insurance as supplemental insurance. A large literature in economics has studied health insurance markets and their inefficiencies, providing evidence on how to better design these markets. Most systems today use community rating, so that no individual is risk-rated, which reduces the problems associated with the loss of risk protection. Before the ACA's rating reforms took effect in 2014, however, health status could affect premiums in the individual and small-group markets, which left many people uninsured because they could not afford their new premiums. When this risk of being reclassified can be insured, the market can accommodate more affordable prices.

2.1 Institutional Framework

We study long-term health insurance contracts in the Chilean setting. Chile is one of the few countries that use this type of contract, which allows us to measure inefficiencies generated from dynamic adverse selection. In addition, the oligopolistic structure of the industry provides a setting to analyze the tools insurers use to retain individuals and protect profits, such as introducing new plans and setting prices.

The Chilean health insurance system combines a public system and a private system. Approximately 80% of the population is enrolled in the public sector. Workers must contribute at least 7% of their income to health insurance, and choose between the public and the private system. To join the private system, individuals often need to contribute more than the required 7% of their income.

In the public sector, health insurance benefits depend on an individual's income group. Enrollees have access to public hospitals and providers participating in the public system. Waiting times are generally longer, and the quality of services may be lower in some situations.

The private sector consists of six competing insurers that attract enrollees and charge a premium for a contract that specifies coinsurance rates, maximum reimbursement amounts, and any restricted provisions, as psychology or certain surgeries. There are no deductibles. These contracts are considered long-term because their key features remain defined over time, enrollees can stay in the same plan as long as they continue paying for it, and price changes from year to year are determined by a specific rule set by the regulatory agency.

Contracts cannot be priced individually, and the premium is calculated as a base price multiplied by a weight that depends on the individual's age and gender. Factor weights are bounded by regulation and usually increase with age, except for female maternity years, which have the largest weights. If the individual continues to pay for the health insurance contract, the insurer cannot deny benefits. This structure is similar to a guaranteed-renewable contract in the sense of Herring and Pauly (2006).

When choosing a health insurance plan in the private sector—for the first time or when moving from another insurer—individuals have to declare pre-existing conditions, and insurers could deny coverage based on these conditions. Thus, even if the price of a health insurance plan is the same for all enrollees within the same demographic group, there can still be denials of coverage.

2.2 Data

Our main dataset comes from the Health Insurance Regulatory agency in Chile (*Superintendencia de Salud*) from 2016 to 2022. It includes information on contracts between insurers and individuals in the private health insurance system. It has information on the health insurance plans and their characteristics, and information on enrollees, their demographics, and their medical claims. We combine this dataset with information from the public sector (*FONASA*), which contains the number of enrollees across different demographic groups and also information on health expenses for these groups, but we cannot wed these two datasets by individual information.

We restrict attention to insurers that sell plans in the market, in contrast to insurers that serve a single company or a few companies. We keep information only for the capital city of Chile, Santiago. We focus on plans that are bought by individual people and do not include plans bought for a group of employees in a company. Regarding family composition, our analysis focuses only on the policy holder of the health insurance plan, although we have information on children and spouse, we do not treat them as separate individuals to compute costs or prices. In Appendix A we detail the construction of our sample.

The final sample consists of six insurers with 5 million individuals enrolled between 2016 and 2022. There are four big insurers with a market share close to 20% and two smaller insurers with 5% and 9% resulting in a market HHI of 1971 (in 2018). One insurer left the industry in 2017, but was bought by a new insurer, and we treat those as the same insurer in the analysis of this paper.

Health Insurance plans are contracts that detail the coverage for different medical services. This can be summarized by an inpatient generosity rate and an outpatient generosity rate. It also includes the maximum amount covered in a year. Once the total expenditure for an individual is larger than this threshold, the generosity rate goes down to a regulated minimum. The prices of the contracts also depend on age and gender group, and their annual adjustment is regulated.

The number of enrollees per health insurance plan is around 37 on average but ranges from 1 to 7,574 individuals. The number of health insurance plans by insurer and year is between 114 and 6,269. And on average there are 128 new health insurance plans per insurer every year.

The market is characterized by a large and rapidly evolving set of plans. At any point in time, insurers maintain thousands of plans, but most have short lifespans: the majority of plan-individual observations correspond to plans that have been offered for fewer than five years. Insurers regularly introduce new plans while closing older ones, resulting in a high number of plans (Appendix Figure 8). Newly introduced plans tend to have lower insurer costs than existing plans in the same coverage category, consistent with new plans attracting a healthier enrollment pool at introduction (Appendix Figure 10). This pattern of frequent plan entry and exit is a central

feature of the market that motivates our analysis of plan introduction as a strategic retention tool.

For more information on the institutional framework, Cuesta, Noton, and Vatter (2019) have an extensive description of the market. They use the same dataset from the regulatory agency in Chile to study the welfare effects of vertical integration between insurers and hospitals. They use the data on medical claims to estimate costs and assess how hospital prices are changing. Atal (2019) also describes the setting in detail.

3 Empirical Facts and Motivating Evidence

In this section, we use the data to study how individuals behave under the rules of the system (Section 2) and what their behavior implies for insurers. Plans are offered every year, with a base premium set at introduction that can later change only within regulated limits. An enrollee whose health deteriorates therefore keeps the price associated with her plan. However, enrollees can leave at any time. Whether they stay and which enrollees are more likely to leave is an empirical question.

We find that enrollees do leave selectively. Low-risk enrollees switch at higher rates than high-risk enrollees, and switching is more common in plans that receive larger price adjustments. This selective exit changes the composition of the remaining pool: plans with higher exit among low-risk enrollees have higher average costs in subsequent years. These patterns capture the two parts of the mechanism described in Section 2: who leaves the plan and how their departure affects the risk pool that remains.

Insurers cannot respond by lowering the price of the existing plan, since its price was set when the plan was introduced. Instead, they can introduce new plans. Insurers do so frequently, and these new plans affect where switching enrollees go. We find that when an insurer introduces plans that are close substitutes for an existing plan, enrollees who leave that plan are more likely to stay with the same insurer rather than switch to a rival. Moving from no entry to the average entry intensity increases the probability of staying with the insurer by four percentage points.

We show this retention comes at a cost. Enrollees who move to a new plan within the same insurer pay lower prices than before, and the insurer earns a lower margin on them than it did in their previous plan.

These results show how selective departures and new plan introduction interact in the data. To determine how much insurers adjust introduction prices in anticipation of future departures, and what consumers and insurers would obtain if enrollees could not leave, we need counterfactual prices and enrollment that we do not observe. We address these questions with the model in Section 4.2.

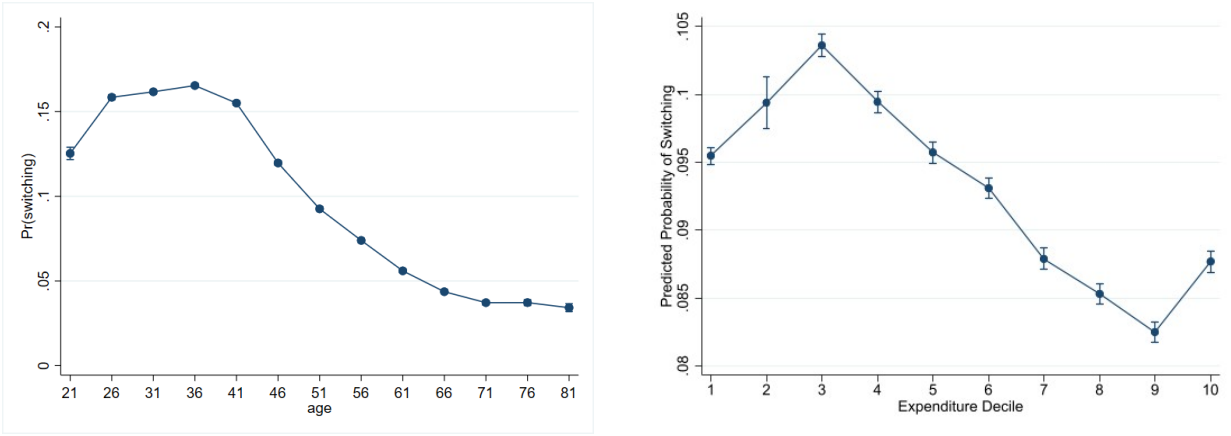
3.1 Switching behavior

Individuals might want to switch plans for different reasons, including a change in family composition, a new disease, or looking for a more suitable plan in terms of characteristics. Another important reason may be that the price of the plan is too high. This leads to dynamic adverse

selection in this market. In the long-term contracts setting with one-sided commitment only on the insurers' side, individuals can leave their plan whenever is convenient for them. When the price of the plan becomes too high for them, potentially because their costs are lower than the average cost of the plan, they leave their plan, which leads to a higher average cost of the plan over time. In Table A2 we specify a logit model for switching to another health insurance plan within the private sector. We show switching is positively correlated with the price adjustment of the current plan. We also observe that low-risk people have higher probability of switching, and having maternity-related expenditures increases probability of switching. In contrast, people that used services related to diabetes or cardiology have a lower probability of switching.

Figure 1 shows the predicted probabilities of switching by age (Panel (a)) and by expenditure decile (Panel (b)), based on a multinomial logit model. We find the probability of switching is large for ages before 41, and monotonically decreases after age 46. For medical expenditures, we find the probability increases until the 3rd decile, suggesting the first two deciles and be composed by young individuals that are joining the system, and decreases after it.

Figure 1: Predicted probability of switching



(a) By age

(b) By expenditure decile

Notes: Predicted probabilities of switching from a logit model, evaluated at sample means of the remaining covariates. Panel (a) controls for price adjustment, chronic conditions and low-risk dummy. Panel (b) controls for coverage rate of the new plan. Both include age and gender of the individual, and year and insurer fixed effects. Sample: 2016-2020.

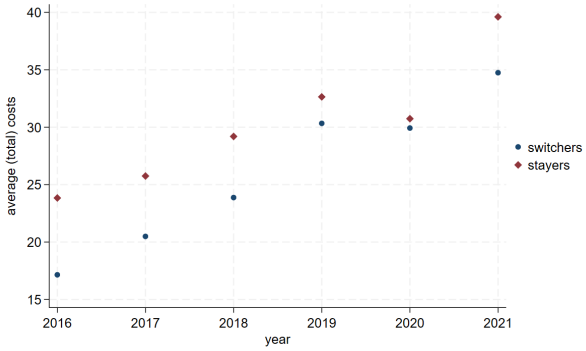
3.2 Selective switching and average costs

Switchers are mainly low risk relative to the average cost of their plan, as shown in Figure 2.

We show what dynamic adverse selection looks like by measuring how the average costs of plans change over time when the lapsation rate—the share of previously enrolled individuals who leave a plan over a given period—is higher for low-risk individuals. We only keep one cohort in the exercise. We define low-risk individuals as those who have average total costs over the years less than the median total cost of the sample.

For each horizon $h \in \{-2, -1, 0, 1, 2, 3\}$, we estimate:

Figure 2: Average costs for plans introduced in 2015



Notes: Mean over total costs of individuals enrolled in plans introduced in 2015 and appearing every year after. Total costs are shown separately for those who stayed in the same plan the next year and those who switched away (left the system or switched to another plan). Lower-risk individuals are more likely to switch.

$$\bar{c}_{j,t+h} = \theta_h \, \Delta LR_{jt}^{L-H} + \gamma_h \, X_{jt} + \zeta_{f(j)} + \zeta_t + \varepsilon_{j,t+h} \tag{1}$$

where $\bar{c}_{j,t+h}$ is the average total cost per enrollee[8] in plan j at time $t+h$, and $\Delta LR_{jt}^{L-H} = LR_{jt}^{L} - LR_{jt}^{H}$ is the differential lapsation rate, defined as the difference between the low-risk and high-risk lapsation rates at time t. The vector X_{jt} includes plan-level controls, $\zeta_{f(j)}$ are insurer fixed effects, and ζ_t are year fixed effects. Standard errors are clustered at the plan level. The coefficient θ_h captures the effect of differential lapsation on average costs h periods ahead. When low-risk enrollees exit at higher rates than high-risk enrollees ($\Delta LR_{it}^{L-H} > 0$), the risk pool of the plan gets worse, increasing average costs. We expect $\theta_h > 0$ for $h \geq 0$, with the effect increasing in h as the compositional deterioration compounds. Estimates of θ_h for $h < 0$ work as a pre-trend test: current differential lapsation should not predict past costs. Table 1 presents the pooled results.

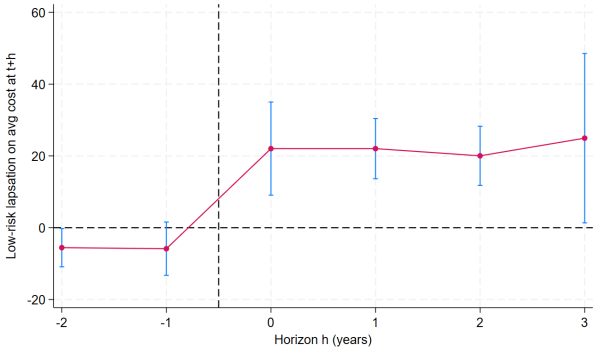
Table 1: Effect of differential lapsation on average plan costs

	Average cost of the plan		
	(1)	(2)	(3)
	No controls	Year FE	Year + Insurer FE
$LR_{it}^{L} - LR_{it}^{H}$	21.61***	22.15***	22.03***
	(6.326)	(6.587)	(6.601)
Observations	2,287	2,287	2,287

Notes: 2016 plan cohort, original enrollees only. Standard errors clustered at the plan level. $LR_{it}^{L} - LR_{it}^{H}$ is the low-risk minus the high-risk lapsation rate. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

[8]Total claims divided by the number of enrollees in the sample. This measure does not represent the insurer's cost; results are similar when we use the insurer's cost.

Figure 3: Dynamic Adverse Selection: higher low-risk lapsation rate increases average costs



Notes: Coefficients from separate regressions of average costs at different horizons on the differential lapsation rate (low-risk minus high-risk). All specifications include insurer and year fixed effects. Standard errors clustered at the plan level. 2016 plan cohort, original enrollees only.

3.3 Effect of introduction of plans in retention rate

In the previous section we showed the decision to switch health insurance plans depends on multiple factors that include family composition, price adjustment or changes in health conditions. Individuals can also switch to similar coverage plans when these are more attractive than their current one. In this section, we study this last explanation using the entry intensity an individual faces, defined as a similarity-weighted measure of new plans relative to their own plan. We provide evidence on how the introduction of plans can motivate individuals to stay within the same insurer, even when switching plans.

We investigate the introduction of plans by insurers every year by defining an entry intensity variable that captures a distance metric of similarity to new plans within an insurer. The question is how many individuals an insurer can retain with new, similar plans. Individuals usually switch for different reasons but if similar attractive plans are offered by the insurer, they might want to stay. To evaluate this empirically, we specify a linear equation of the decision to stay within insurer conditional on switching plans, as a function of the entry intensity faced by their current plan, chronic-condition variables, age, gender, and insurer variables. We also control for similar-plan fixed effects that capture the characteristics of the current plan. Entry intensity could be endogenous to unobserved variables that affect the decision to switch, such as costs or environmental or market factors, for example, the exit of one insurer. We instrument for entry intensity using a shift-share specification that uses differential exposure of demographic groups to similar plans. We assume the shares are exogenous conditional on individual characteristics and class attributes, as the decision to stay within a group of similar plans is not driven by the aggregate demographic composition of the group of similar plans but by the characteristics of the insurer, the product and the individual. Following Goldsmith-Pinkham, Sorkin, and Swift (2020), we present validity tests for the instrument and decompose the identifying variation across demographic groups using Rotemberg weights.

Insurers decide to introduce new plans based on the characteristics of their existing pool of enrollees and the screening needs within each group of plans. A group of similar plans with a

higher concentration of certain demographic groups faces different incentives to redesign its plan offerings, as the composition of enrollees determines the degree of adverse selection risk within each group. We exploit this differential exposure of demographic group across groups of similar plans (similar-plan cells) to construct the instrument. The share component captures the predetermined enrollment patterns of each demographic group k across similar plans, while the shift component captures the national growth rate of average medical costs for each demographic group. Similar-plan cells more exposed to demographic groups with higher relative costs face stronger incentives to introduce new plans, generating variation in plan introductions that is plausibly orthogonal to the individual decision to stay within insurer, conditional on product characteristics.

The instrument Z_{ct} is constructed as the sum over demographic groups of the product of the predetermined shares and the shift. The shares represent the demographic group shares within a similar-plan cell c in 2016 (predetermined year). The shift is the average cost growth rate by demographic group at the national level[9].

There are six insurers f with a total of 266 similar-plan cells c (group of plans classified according to similar attributes) over three years, while demographic groups k consist of 31 age and gender cells. Individual characteristics X_{it} include age, gender and health conditions including diabetes, maternity and cardiology. Plan attributes $\zeta_{c(i)}$ are represented by similar plans FE. The regressor entry intensity is the similarity-weighted sum of the insurer's new plans, where similarity to the individual's current plan is measured in coverage rate, maternity restriction, and price.

$$\mathbf{1}\{\text{stay same insurer}\}_{it} = \beta_1 X_{it} + \beta_2 \,\text{entry-intensity}_{it} + \zeta_{c(i)} + \zeta_{f(i)t} + \zeta_{k(i)t} + \epsilon_{it} \tag{2}$$

Table 2 presents the main results. A one-standard-deviation increase in entry intensity raises the probability of staying within the same insurer by 2 percentage points[10]. Moving from zero entry to mean entry intensity increases retention by 4 percentage points. OLS effect in Column (2) is also positive but smaller. We also present the unconditional results: how entry intensity affects the decision to stay or switch in the whole sample. In Table 3, Columns (2)–(4) are mutually exclusive categories. Entry intensity has no significant effect on staying in the same plan (Column (2)) or on switching within the same insurer (Column 3), which is in line with the interpretation that entry is not triggering the decision to switch. However, we see in Column (4) that the probability to switch insurers decreases when entry intensity within the same insurer is larger (the probability to move to a rival decreases by 7%). We compute Rotemberg weights following Goldsmith-Pinkham, Sorkin, and Swift (2020). Because the 2016 shares do not vary within a similar-plan cell, the relevant unit is a demographic group by year cell: cluster fixed effects absorb cross-sectional share variation. The decomposition has 90 cells, pooled to 72 for reporting by merging the sparse tail bins (16-20 into 21-25, 81+ into 76+). The weights sum to one and $\sum_{k,t} \hat{\alpha}_{kt}\hat{\beta}_{kt} = 0.029$ recovers the two-stage least squares coefficient[11].

[9] Information computes costs of health-related expenses. Information given by Superintendencia de Salud, Chile.

[10] The mean of the outcome variable, stay conditional on switching, is 46.5 percent. Entry closeness is measured in standard deviations, so the coefficient is the effect of a one standard deviation increase. The mean of entry closeness is 1.75 standard deviations, so at mean closeness the effect is aronud 4 percentage points.

[11] Individual controls of the main equation are aggregated to means in the just-identified regressions, and so it

Thirty-three cells carry positive weight, summing to 5.11, and 39 carry negative weight, summing to −4.11. Negative weights account for 45% of total absolute weight, so the estimate is not a convex combination of group-specific effects: it nets +0.029 from −0.054 in positive-weight cells and +0.083 in negative-weight cells. In Table A5 we see the decomposition at the group level, 15 of 24 groups have net positive weight and 9 have net negative weight. Weight concentrates on mid-age males, consistent with these enrollees being a primary target of insurer screening through plan design. Males 76+ carry the third-largest weight, their shares matter because entry responds to the cost profile of the cluster's full enrolled book. The cell weights ($\hat{\alpha}_{kt}$) and just-identified coefficients ($\hat{\beta}_{kt}$) are dispersed[12]. The $\hat{\alpha}$-weighted mean $\hat{\beta}$ is −0.011 across positive-weight cells and −0.020 across negative-weight cells.

Despite the large negative weights, the estimate is stable. The leave-one-out analysis in Table A3 shows that dropping any single demographic group moves the estimate between 0.017 and 0.028 around a baseline of 0.022, well within one standard error. The only k estimates for well-identified groups (first-stage $F > 15$) range from 0.015 to 0.023, tightly centered on the baseline. The only-k column does show an age gradient, with positive estimates for younger groups and negative for 71+, which points towards some heterogeneity. This is the one case where negative weights can be a concern. Table A4 reports additional robustness checks under alternative specifications, including cluster-specific time trends and richer individual-level controls, as well as a placebo test that regresses the 2016 staying-within outcome on the post-2016 instrument in Table A6.

An important caveat concerns spillovers across insurers. New plans introduced by one insurer may attract enrollees from rivals, affecting retention decisions of individuals not directly exposed to the new plans. These cross-insurer effects are not captured in the main equation and represent a threat to SUTVA. Although this exercise provides a useful benchmark under the assumption that own-insurer plan introductions are the primary driver of within-insurer retention, we address spillovers in the structural model of the following sections, where cross-insurer competition is explicitly modeled. The counterfactual in Section 6.1.1 includes the new plans of every insurer in every choice set and then removes all new plans at once, so its effects include spillovers across insurers. That analysis uses a moment that measures, among enrollees who move to a newer plan, the share who stay with their own insurer (Figure 20).

3.4 Reclassification evidence

We compute margins as price minus insurer cost at the destination plan, and compare switchers vs stayers effects on margins for within-insurer switching and cross-insurer switchers. For stayers, the destination margin is measured in the switchers' origin plan. Regressions control for plan of origin and year fixed effects, together with health state proxied by health costs before moving. Table 4 presents reclassification results for moving individuals, across and within the same insurer. When an individual switches to another plan within the same insurer, the margin gains on that person are reduced around 9.98 UF per year (Δm_t, omitted category is stayers),

does not recover the exact coefficient of the individual-level regression.

[12]Their standard deviations are 0.22 and 0.35. This dispersion reflects weak first stages in individual cells, not influential shocks (see the notes to Table A5). No single cell contributes more than 44% of the total.

Table 2: Effect of Plan Entry on Within-Insurer Retention

	(1) First stage	(2) OLS	(3) IV
Shift-share Instrument	103.22782***		
	(20.95884)		
Entry intensity		0.00994**	0.02166**
		(0.00415)	(0.01092)
KP F-statistic			24.26
Similar plan FE	✓	✓	✓
Insurer × Year FE	✓	✓	✓
Age × Female × Year FE	✓	✓	✓
Observations	251,717	251,717	251,717
N clusters	260	260	260

Notes: Standard errors clustered at the aggregate-plan level. All specifications control for individuals' chronic conditions and an indicator for having dependents. KP: Kleibergen-Paap F-statistic. The mean of the outcome is 46.5; entry intensity has mean 1.75 (in standard deviation). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Effect of Plan Entry on Within-Class Retention

	(1) Switch within[a]	(2) Same plan	(3) Switch within	(4) Switch across
Entry intensity	0.02166**	-0.00354	0.00655	-0.00301**
	(0.01092)	(0.00566)	(0.00539)	(0.00142)
KP F-statistic	24.26	32.72	32.72	32.72
Similar plan FE	✓	✓	✓	✓
Insurer × Year FE	✓	✓	✓	✓
Age × Female × Year FE	✓	✓	✓	✓
Observations	251,717	2,040,000	2,040,000	2,040,000
N clusters	260	266	266	266

Notes: Standard errors clustered at the aggregated-plan level. All specifications control for individuals' chronic conditions and an indicator for having dependents. Columns (2)–(4) are unconditional. KP: Kleibergen-Paap F-statistic. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
[a] Conditional on switching. Columns (2)-(4) are unconditional.

probably explained by switching to a cheaper plan (Column (2)). However, when the individual is moving across insurers, the rival firm gains 3.5 UF less than what the origin insurer was receiving in the previous plan. Column (1) of Table 4 shows the margins of the switching individual at the new plan (m_{t+1}) compared to the stayers in his origin plan. For own-firm switchers, margins in the new plan are 10.3 UF lower than those of stayers in the original plan, while cross-firm switchers show no significant margin difference relative to stayers. Given $\Delta m_t = m_{t+1} - m_t$, the results imply that own-firm switchers had the same average margins as stayers in the original plan (m_t), whereas cross-firm switchers had original margins about 3.4 UF larger than those of stayers. Cross-firm switchers were more profitable than the average enrollee in the pool before the move. These results imply that switchers are, on average, positively reclassified, and more so within the same insurer.

These results condition on origin-plan fixed effects, year, and pre-move health claims, which control for health status. Retention costs the insurer about 10 UF per mover, whereas a rival earns on the mover about the same average margin as the origin plan. In Appendix Table A7, we show descriptive statistics on margins gains of insurers by price tercile of the origin plan, and Table A8 compares the margin on own-insurer switchers with the margin they would have generated had they stayed.

Table 4: Margins on movers vs. stayers

	(1) Margin at t+1 (level)	(2) Δ margin (change)
Own-firm switcher	-10.301***	-9.989***
	(1.102)	(1.216)
Cross-firm switcher	-0.172	-3.532***
	(0.911)	(1.054)
Pre-move claims	-0.426***	0.032***
	(0.025)	(0.002)
Observations	1,312,520	1,312,520
Clusters	69	69

Notes: Cluster-robust standard errors (insurer × aggregate-plan) in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Sample: stayers and switchers with pre-move claims (origins 2017-2018); mover coefficients compare against stayers of the same origin plan × year. Ommitted category is stayers.

4 Model

Insurers use new plan introductions to reprice new cohorts of individuals, and the facts in Section 3 provide suggestive evidence that they also use them to retain individuals who would otherwise switch to other insurers. We want to know how much of the price set at introduction reflects the margin the insurer expects to recapture on retained enrollees in its own future plans, and use this to evaluate alternative regulatory rules. To decompose insurers' incentives for reten-

tion and repricing, we present a model of insurer competition that takes the demand model as an input. Demand is set up as a discrete choice model of health insurance plans.

The model has two parts. Consumers choose among plans given prices, expected out-of-pocket costs and their variance, plan characteristics, and a switching cost. Insurers set the price of a plan once, at introduction, and cannot change it afterwards. The insurer therefore balances the profit the plan earns today against the enrollment it carries into every future period, because an enrollee who joins today stays unless she incurs the switching cost to leave. The insurer's first-order condition is written in terms of choice probabilities and their derivatives with respect to price, derived from the demand model. Section 4.1 sets out demand and Section 4.2 sets out the insurer's problem.

4.1 Demand Model

Every year, individuals choose a health insurance plan from their choice set, which includes plans from the six insurers competing in the private sector and the outside option which is choosing the public sector. Private plans are diverse and have multiple characteristics, which makes different plans suitable for different people. The price of the plan is a relevant characteristic and different groups of individuals might have diverse preferences over prices. For example, high-income individuals might have a lower price elasticity, whereas older individuals might be more price-elastic as health expenditures are an important part of their expenses. Premiums vary substantially across plans and individuals, and individuals differ in their sensitivity to prices. This market is also characterized by high inertia, as individuals usually stay in their previously chosen plan.

We specify a discrete choice model in which individual i chooses a plan j from choice set J_{it}, which includes private-sector plans available to the individual and an outside option corresponding to enrollment in the public system. The choice set is individual-specific and includes plans that are closed to new enrollment but remain available to their current enrollees. Individual i belongs to demographic group g. The indirect utility for each plan j is given by:

$$\begin{aligned} u_{ijt} = & \underbrace{\alpha(z_{it})\, p_{jg(i)t}}_{\text{price}} + \underbrace{\beta_1 x_j}_{\text{attributes}} + \underbrace{\xi_{jg(i)t}}_{\substack{\text{unobs.}\\ \text{attributes}}} + \underbrace{\omega_0\, EOP_{ijt} + \omega_1\, VOP_{ijt}}_{\substack{\text{OOP}\\ \text{beliefs}}} \\ & + \underbrace{\gamma_1 \mathbf{1}\{j = j_{i,t-1}\} + \gamma_2 \mathbf{1}\{f(j) = f(j_{i,t-1})\}}_{\text{inertia}} + \varepsilon_{ijt} \\ u_{ijt} = & \delta_{g(i)jt} + \mu_{ijt}(z_i, j_{i,t-1}) + \varepsilon_{ijt} \end{aligned} \tag{3}$$

where $f(j)$ denotes the insurer offering plan j, so γ_1 captures plan-level inertia and γ_2 captures insurer-level inertia. δ_{gjt} represents mean utility common to individuals in demographic group g and μ_{ijt} captures preference heterogeneity driven by observable characteristics.

Individual i cares about the price of the plan j defined for their demographic group g, observable characteristics of the plan x_j and unobserved quality of the plan ξ_{jgt}. When choosing a plan, the individual also takes into account the future out-of-pocket costs she will pay for medical

expenses, that is, the amounts not covered by the insurer. The distribution of future costs is represented in this utility by the expected out-of-pocket cost EOP_{ijt} and its variance VOP_{ijt}. This captures the individual's risk aversion toward uncertain future costs, which is important when choosing a plan [13]. Individuals are allowed to choose a plan every year, however, large switching costs in this industry make individuals less likely to switch plans even when alternative options are more suitable. Inertia has been studied extensively in complex financial markets (Abaluck and Gruber (2011) and Brown, Egan, Jeon, Jin, and Wu (2025)) and more specifically in health insurance markets, where some articles focus on identifying its sources, which include pecuniary and non-pecuniary costs (Handel and Kolstad (2015) and Heiss, McFadden, Winter, Wuppermann, and Zhou (2021)) or its welfare consequences (Handel (2013)), while others assume that individuals may not study all the plans offered in a market (Lucarelli, Prince, and Simon (2008), Coughlin (2023), Abaluck and Adams-Prassl (2021) Atal (2019)). Staying in the same plan is represented by $\mathbf{1}\{j = j_{i,t-1}\}$ which indicates that individual i chose j in period $t-1$. The extra-utility amount that corresponds to staying in the same plan is $\gamma_1 + \gamma_2$, and only γ_2 if the individual switches to another plan within the same insurer.

Individuals have heterogeneous preferences for prices that vary with observable demographics z_i, including age, gender, and income. This specification allows price sensitivity to differ systematically across individuals. Finally, the individual faces an idiosyncratic shock for each plan j at period t, ε_{ijt}. We assume this shock has Type I extreme value distribution, and therefore the choice probability for option j, conditional on previous plan k can be written as:

$$s_{ijt}(k) = \frac{\exp(\,\delta_{g(i)jt} + \mu_{ijt}(z_i, k)\,)}{\sum_{l \in J_{it}} \exp(\,\delta_{g(i)lt} + \mu_{ilt}(z_i, k)\,)} \tag{4}$$

and the market share of plan j for demographic group g at period t is given by:

$$s_{jgt} = \int \sum_{k \in J_{i,t-1}} s_{ijt}(k) \Pr_{t-1}(k \mid z_i)\, dF_{gt}(z_i) \tag{5}$$

$$\Pr_{t-1}(k \mid z_i) = s_{i(g)k,t-1} = \sum_{m \in J_{i,t-2}} s_{i(g)k,t-1}(m) \Pr_{t-2}(m \mid z_i) \tag{6}$$

Equation (6) links choice probabilities across periods. The recursion is initialized at the individual's entry into the market: the first choice is made with no previous plan, so both inertia terms are zero for every option in the entry choice set.

The demand model delivers plan choice probabilities and price elasticities that are used to evaluate consumer responses to premium changes and plan characteristics. These demand estimates are a key input for the supply-side analysis, where insurers' pricing and plan design decisions depend on expected enrollment and consumer risk composition.

[13] Abaluck and Gruber (2011) specify a linear indirect utility from a CARA utility, in which the risk aversion parameter is $\frac{2\omega_1}{\omega_0}$.

4.2 Supply

Insurers set the price of each plan once, when the plan is introduced, and regulation prevents them from repricing it later. The insurer chooses the price of a new plan to maximize the discounted profits the plan earns over its life. Because demand has inertia, the price affects profits in every future period through enrollment and its composition, not only through this period's margin. Insurers face a dynamic trade-off when setting prices and introducing new plans. Higher prices increase current margins but may accelerate the exit of low-risk individuals, worsening the future risk pool through dynamic adverse selection. Conversely, introducing new plans at lower prices can retain profitable low-risk consumers, even if these plans generate lower margins. Firms therefore balance static pricing incentives with dynamic considerations related to consumer retention and the future composition of enrollment.

Six insurers compete to attract individuals to their health insurance plans. In each period, insurers simultaneously decide which new plans to introduce and set their prices. They take previously introduced plans as given and know the plans introduced by their rivals at t, consistent with the equilibrium choices of all insurers. They choose which plans to introduce and their prices taking into account the future stream of profits generated by each plan.

A multi-product insurer f introduces plan j at date t and sets its base price p_j once. Individual i pays $p_j \times rw_j$, where the rating weight rw_j is a regulated function of age and gender[14]. The insurer's cost of covering individual i is c_{ij}, which is the individual's total cost multiplied by the coverage rate of plan j[15]. Let $\mathcal{J}_t$ denote the menu of plans at t, with information on plans' attributes and prices, and let$\mathcal{J}_{f,t} \subseteq \mathcal{J}_t$ denote the firm's own portfolio. Let $\mathcal{I}$ denote the population of individuals in the market, including those currently enrolled in the public system, and let $s_{ij,t} = \sum_{k \in J_{i,t-1}} s_{ijt}(k) \Pr_{t-1}(k \mid z_i)$ be individual i's choice probability for plan j, unconditional on the previous plan. Integrating over $\mathcal{I}$ aggregates these probabilities into plan enrollment, $\int_{\mathcal{I}} s_{ij,t}\, di = \sum_g N_{gt}\, s_{jgt}$, where N_{gt} is the number of individuals in demographic group g. Demand has inertia, so the choice probabilities $s_{im,t+\tau}$ delivered by the demand model depend on p_j both directly and through i's enrollment history. The state variables of the firm at time t include the menu of plans $\mathcal{J}_t$ and the distribution of individuals in all previous plans $\mathbf{s}_{t-1} = \{s_{im,t-1}\}_{i,m}$, which records who holds each plan and so has information on individuals' characteristics; however, individuals demographics, health and costs evolve exogenously and do not depend on prices. The firm's value from the pricing decision is

$$\Pi_f(p_j; \mathcal{J}_t, \mathbf{s}_{t-1}) = \mathbb{E}\left[\sum_{\tau \geq 0} \beta^\tau \sum_{m \in \mathcal{J}_{f,t+\tau}} \int_{\mathcal{I}} (p_m - c_{im,t+\tau})\, s_{im,t+\tau}\big(\mathbf{s}_{t+\tau-1}, \mathcal{J}_{t+\tau}\big)\, di \;\middle|\; \mathcal{J}_t,\ \mathbf{s}_{t-1} \right] \quad (7)$$

where the expectation integrates over rival entry and exit $\mathcal{J}_{-f,t+\tau}$, the insurer's own future portfolio actions (introductions, prices, closures) decided after a private *iid* random shock $\epsilon_{f,t+\tau}$[16],

[14] We suppress rw_j in the notation. Every empirical counterpart is computed with the observed rating table.

[15] The individual's cost c_i follows the simulated cost process of Section 5.3.1.

[16] The private *iid* shock to the entry decision represents random costs or benefits of introducing new plans,

and demand and cost shocks. At the pricing date the firm observes the entire current menu, but the future path of the menu is uncertain[17].

A change in p_j affects profits in every period of the plan's life. Contemporaneously, it moves revenue on the retained pool and demand at the margin, and over time, it operates through the state: enrollment acquired today is carried forward by inertia, the composition of the retained pool shifts, and enrollees displaced from j reappear in the firm's later plans. Future prices defined by the firm and the rivals are set given a menu that includes (j, p_j), so they are also functions of today's decision. We differentiate equation (7) with respect to the price of the plan p_j and use two assumptions to reduce this dependence to a tractable set of terms (the derivation is in Appendix E). First, the firm's own future actions are chosen optimally against the same objective, so by the envelope theorem their adjustment to p_j has no first-order effect on the value. Future plans therefore enter the condition at their optimized prices, only through the direct demand cross-effect of p_j on their shares (Assumption 5[18]). Second, incumbent rival plans cannot reprice, because by regulation the price is set once, so the remaining rival response is future entry. Entry and the prices of entering plans depend on class-level aggregates of the state, such as the enrollment and cost of each coverage class and the number of plans in it. Because plans are small (Assumption 6), the price of one plan does not move these aggregates, and the menu process can be treated as invariant to p_j. The first-order condition for a plan introduced at t is

$$0 = \mathcal{F}_j^{(0)} + \sum_{\tau \geq 1} \beta^\tau \, \mathbb{E}\Big[\mathcal{F}_j^{(\tau)} \,\Big|\, \mathcal{J}_t,\ \mathbf{s}_{t-1} \Big], \tag{8}$$

where the period-τ marginal-profit contribution of p_j decomposes into three blocks,

$$\mathcal{F}_j^{(\tau)} = \underbrace{\int_{\mathcal{I}} s_{ij,t+\tau}\, di}_{\text{(A) level}} + \underbrace{\int_{\mathcal{I}} (p_j - c_{ij,t+\tau}) \frac{ds_{ij,t+\tau}}{dp_j}\, di}_{\text{(B) own derivative}} + \underbrace{\sum_{k \in \mathcal{J}_{f,t+\tau} \setminus \{j\}} \int_{\mathcal{I}} (p_k - c_{ik,t+\tau}) \frac{ds_{ik,t+\tau}}{dp_j}\, di}_{\text{(C) cross, incl. recapture}} \tag{9}$$

Block (A) is the inframarginal term: raising p_j earns revenue on every staying enrollee, including non-responsive individuals, who could be captive due to large switching costs. Block (B) captures individual margins, price minus cost, weighted by individuals' price sensitivity. Captive enrollees have lower sensitivity than individuals choosing for the first time. Static selection appears at every period and it is adverse if marginal enrollees with large willingness to pay have higher individuals costs. In selection markets with imperfect competition, if higher prices translate into larger average costs, the markup term (price minus average cost) is lower (Mahoney and Weyl

and makes the value of introduction differentiable in p_j. The shock plays the role of the private information for the payoffs in Doraszelski and Satterthwaite (2010) that guarantees existence in dynamic discrete games. Here it makes the expected continuation value differentiable in p_j, and it makes the introduction policy a conditional choice probability. In the estimation, plan introduction enters the transitions of Section 5.3.1 through entry intensity.

[17] Beliefs assume a stationary market: no long-run change in technology, costs or the composition of the population. Information about consumers' costs is symmetric across insurers and time, and insurers predict each person's cost with the same Markov process.

[18] The envelope theorem applies to future prices because they are chosen at interior optima. We transform the discrete plan-introduction and closure decisions into cutoff rules by introducing private payoff shocks, $\epsilon_{f,t+\tau}$. At the margin, these decisions correspond to indifference conditions, so the envelope argument continues to apply. The argument is in Appendix E.

(2017)). Over the life of the plan, the enrolled composition can change the average cost: as the plan ages, the enrolled pool becomes older. Also, the pool average cost can increase as healthier enrollees exit for other reasons not related to price, and this enters the condition through the cost path $c_{ij,t+\tau}$ at which every future margin is evaluated. Whether weighted marginal cost sits below or above the pool average is an empirical question about who is mobile, and is one of the objects the supply estimation recovers. Block (C) captures the cross-effects on the insurer's other plans, including its own future plans represented in the recapture margin: enrollees priced out of j who land in the firm's newer plans are not lost, only re-margined.

The derivatives in (9) are total: the choice probability depends on prices and on last year's choice probabilities. Let $\Pr_{i,t}(j \mid m) \equiv s_{ijt}(m)$ be the demand model's probability, that i chooses j at t having held plan m (or the outside option) the year before. The choice set J_{it}, and therefore entry and exit of plans, enters through its denominator. Enrollment evolves by the law of total probability over where the individual was,

$$s_{ij,t+\tau} = \sum_{m} \Pr_{i,t+\tau}(j \mid m)\, s_{im,t+\tau-1}, \tag{10}$$

and persistence implies that $\Pr_{i,t}(j \mid j) > \Pr_{i,t}(j \mid m)$ for $m \neq j$. Differentiating (10), we obtain the own-price derivative, which is recursive and computed forward:

$$\frac{ds_{ij,t+\tau}}{dp_j} = \underbrace{\sum_{m} \Pr_{i,t+\tau}(j \mid m)\, \frac{ds_{im,t+\tau-1}}{dp_j}}_{\text{persistence propagation}} + \underbrace{\sum_{m} s_{im,t+\tau-1}\, \frac{\partial \Pr_{i,t+\tau}(j \mid m)}{\partial p_j}}_{\text{direct: this year's margins}}, \tag{11}$$

initialized at the derivative in the year of introduction, $ds_{ij,t}/dp_j$. A price cut at introduction raises enrollment today, and persistence carries the acquired enrollees forward. Enrollees made marginal by p_j in earlier years are still in the pool, one year older and probably costlier. Enrollment is carried forward by the stay probability $\Pr_{i,t+\tau}(j \mid j)$, while the response to the price is carried forward by the stay probability net of the displaced who would have returned to j anyway (Appendix E). The cost term of the FOC, which is $c_{ij,t+\tau}$ weighted by (11) gives the response of the plan's total cost to its price, $d\,\mathrm{TC}_{j,t+\tau}/dp_j$, which splits into an inherited term carried through time, the leaving margin of the pool the plan has built, and the joining margins (Appendix E, equation (64)).

When the conditioning is irrelevant, $\Pr_{i,t}(j \mid j) = \Pr_{i,t}(j \mid m)$ for all m, there are no switching costs and no persistent selection into the plan. The propagation term then drops, because last year's responses sum to zero across destinations, (11) collapses to the contemporaneous derivative, and (8) reduces to a discounted sequence of static Bertrand conditions. The difference between the two is the dynamic persistence motive. When there is inertia but prices are chosen every period, the setting is one of dynamic pricing Dubé, Hitsch, and Rossi (2010) and Fleitas (2017), with an invest then harvest mechanism: a low price today attracts enrollment that is captive next period, which allows the firm to raise its price later. Here the insurer cannot reprice, so it has to combine both the invest and the harvest motives into one price at introduction. When insurers decide on the introduction of plans and prices, the equilibrium concept to model the

market structure would be given by a Markov Perfect Equilibrium: a strategy of entry or not and its price that maximizes profits given beliefs over rivals and the future conditional on the information set that includes menu of plans, distribution of consumer types, tastes and costs. The price decision satisfies the FOC of the firm (optimality). The beliefs are given by the observed transitions in the market as a function of the states, reduced to some moment-based state variables (Ifrach and Weintraub (2017)). For the off-path beliefs, this could be derived from the structural discrete choice model of demand with expectations over market structure. If the estimated law of motion is a result of all firms best responding to them, that would require a fixed point leading to a moment-based equilibrium. We do not show the estimated transitions are the results of best responses. This relates to experienced-based-equilibrium (Fershtman and Pakes (2012) and Dubois and Pakes (2026)) as it satisfies optimality given beliefs, and beliefs are consistent with outcomes at repeatedly visited states. Regarding multiplicity of equilibria for estimation, we estimate conditions that any equilibrium consistent with the data would satisfy.

4.3 Prices under market power with one price for life

This section rewrites the price implied by the first-order condition (8) as an equilibrium object that can be decomposed into different terms. The decomposition is useful to interpret the results of the counterfactuals in Section 6. Summing (9) over the life of plan j, introduced at t, and suppressing the expectation over the future menu and the rating weights, the price is

$$p_j = \underbrace{\frac{\sum_\tau \beta^\tau \int c_{ij,t+\tau} \frac{ds_{ij,t+\tau}}{dp_j} di}{\sum_\tau \beta^\tau \int \frac{ds_{ij,t+\tau}}{dp_j} di}}_{\text{marginal cost}} + \underbrace{\frac{\sum_\tau \beta^\tau \int s_{ij,t+\tau}\, di}{-\sum_\tau \beta^\tau \int \frac{ds_{ij,t+\tau}}{dp_j} di}}_{\text{markup}} + \underbrace{\frac{\sum_\tau \beta^\tau \sum_{k \in \mathcal{J}_{f,t+\tau} \setminus \{j\}} \int \left(p_k - c_{ik,t+\tau}\right) \frac{ds_{ik,t+\tau}}{dp_j} di}{-\sum_\tau \beta^\tau \int \frac{ds_{ij,t+\tau}}{dp_j} di}}_{\text{recapture}} \tag{12}$$

$$p_j = \underbrace{\frac{\sum_\tau \beta^\tau \int s_{ij,t+\tau}\, c_{ij,t+\tau}\, di}{\sum_\tau \beta^\tau \int s_{ij,t+\tau}\, di}}_{AC_j^{PV}} + \underbrace{\frac{\text{Cov}_q(c,\, r)}{\bar{r}_j}}_{\text{selection wedge}} + \text{markup} + \text{recapture} \tag{13}$$

$$\begin{aligned} p_j = {} & AC_j^{PV} + \underbrace{\sum_\tau q_\tau \frac{\text{Cov}_{q,\tau}(c,\, r)}{\bar{r}_j}}_{\text{health selection}} \\ & + \underbrace{\sum_\tau q_\tau \frac{\left(AC_{j,t+\tau} - AC_j^{PV}\right)\left(\bar{r}_{j,t+\tau} - \bar{r}_j\right)}{\bar{r}_j}}_{\text{selection across ages}} + \text{markup} + \text{recapture} \end{aligned} \tag{14}$$

In (12), the marginal cost is the cost of the enrollees the price moves over the plan's life, the markup is lifetime enrollment over its lifetime response to the price, and the recapture term is the margin the insurer earns on the enrollees the price pushes into its other plans. Equations (13) and (14) further decompose the marginal cost. In this decomposition we consider a single cohort that enrolls when the plan is introduced, so that τ indexes both the age of the plan and the age band of its enrollees. Enrollee i in age band τ accounts for a share $q_{i\tau} = \beta^\tau s_{ij,t+\tau} \big/ \sum_{\tau'} \beta^{\tau'} \int s_{ij,t+\tau'}\, di$ of

the plan's discounted enrollment, and $\mathbb{E}_q$ and Cov_q are the mean and the covariance over enrollees and age bands with these enrollment shares. The responsiveness of enrollee i in age band τ is $r_{ij,t+\tau} = -\big(ds_{ij,t+\tau}/dp_j\big)/s_{ij,t+\tau}$, the proportional fall in their probability of being in the plan when the price rises. Let $\bar{r}_j \equiv \mathbb{E}_q[r_{ij,t+\tau}]$ denote the plan's mean responsiveness over all enrollees and age bands,it is a single number and does not vary with τ. With these shares, AC_j^{PV} in (13) is the plan's lifetime average cost, the present value of the plan's costs divided by the present value of its enrollment, which equals the mean cost $\mathbb{E}_q[c_{ij,t+\tau}]$, and the marginal cost is $\mathbb{E}_q[r_{ij,t+\tau}c_{ij,t+\tau}]/\bar{r}_j$, the average cost weighted by responsiveness. Their difference is the covariance in (13). Each cost is paired with the response in the same band. The bands are linked through the derivatives, which are total, and through the enrollment shares. In (14), $q_\tau = \int q_{i\tau}\, di$ is the share of the plan's discounted enrollment in age band τ, $AC_{j,t+\tau}$ and $\bar{r}_{j,t+\tau}$ are the average cost and the mean responsiveness of the plan's enrollees in that band, weighted by enrollment $s_{ij,t+\tau}$, and $\mathrm{Cov}_{q,\tau}$ is the covariance between cost and responsiveness among the same enrollees, with the same weights. The covariance of (13) splits into a part within each age band and a part across age bands.

The selection wedge is negative when the enrollees who respond to the price are cheaper than the average. It has two parts. Health selection is the covariance between cost and responsiveness among enrollees in the same age band. It is negative under adverse selection, when the healthy are the ones who respond to the price. Selection across ages is the covariance, across age bands, between the average cost of the plan's enrollees and their average responsiveness. Average cost rises with age, while responsiveness falls. In the young bands many enrollees have just joined and can still move, while in the old bands most are held by inertia. In the young bands, average cost is below its lifetime average and responsiveness is above its mean. In the old bands, average cost is above its lifetime average and responsiveness is below its mean. Both give a negative product, so the term is negative. It is selection on age, not on health, and it comes from charging all ages one base price, with age factors that rise much less than costs.

In the static condition of selection markets with market power of Mahoney and Weyl (2017), the price is the average cost plus a selection wedge plus a markup, in one period. Our condition has the same structure, with two differences: the averages and the responses are taken over the life of the plan, and the wedge includes selection across ages, which comes from charging one base price at every age. Rating weights reduce this term, because payments also rise with age, and it would vanish if payments rose in proportion to costs. Here rating weights rise about two times over the life cycle while costs rise about seven times.

The comparison with perfect competition. Under perfect competition, profit over the life of the plan is zero and the price is the lifetime break-even price, $p_j = AC_j^{PV}$. The pool that enrolls in equilibrium is different, so this is not the same AC_j^{PV} evaluated at our prices. Under market power, (12) to (14) give

$$p_j - AC_j^{PV} = \underbrace{\text{health selection} + \text{selection across ages}}_{\text{selection wedge}} + \text{markup} \;+\; \text{recapture}.$$

The markup pays for the two selection discounts, and what is left above them is the margin over lifetime break-even. When the markup no longer covers the discounts, the plan prices at the break-even floor, $p_j = AC_j^{PV}$. A change in costs that affects every enrollee raises the average and the marginal cost by the same amount, and the price rises by that amount. A change in who is in the pool that is not caused by the price changes the average cost but not the marginal cost, so it changes the margin and not the price, until the price reaches the break-even floor. Appendix I compares the assumptions to contrast this case with the long-term contracts of Hendel and Lizzeri (2003) in a two-period version.

5 Estimation and Results

5.1 Demand Identification and Estimation

Identification. We have information on multiple health insurance plans with different attributes and benefits that can be bought by different demographic groups, which provides variation that allows us to identify parameters at both the plan level and the individual level. The choice set construction for each individual is explained in Appendix B.

To identify preferences related to risk aversion, we use expected out-of-pocket costs and their variance that are constructed at the individual and plan level usinga prediction model that maps individual health history and demographics into expected spending given each plan's observable features, including coverage rates and hospital provider indicators (see Appendix (B))[19]. While plan level averages of predicted out-of-pocket costs may be correlated with unobserved plan quality, identification of risk preferences relies on within-plan, across-individual variation in predicted out-of-pocket costs driven by observed differences in health risk. Conditional on observable plan characteristics, insurer fixed effects, and demographic-group controls, the individual-specific variation is not correlated with unobserved plan quality, ruling out moral hazard on unobservables: individuals with the same health status do not spend more when unobserved plan characteristics are more generous. Under this assumption, variation in EOP identifies ω_0 and VOP variation identifies ω_1.

Switching-costs parameters are identified from individuals choosing the same health insurance plan through time even though alternative plans offer higher indirect utility absent switching costs. Changes in choice set over time generate incentives to switch plans, which provide identifying variation for the switching cost parameters. Persistence in choices may also arise from spurious state dependence due to serially correlated taste shocks or from time-invariant preferences for specific plans. We do not allow for individual unobserved heterogeneity. Following the standard approach in the literature, we assume idiosyncratic taste shock ε_{ijt} are iid ruling out serial correlation in preferences. Time-invariant preferences are addressed by conditioning on a rich set of individual characteristics, plan attributes, and insurer fixed effects (Dubé, Hitsch, and Rossi (2010)).

[19] We use a gradient boosting regression model, which is a decision-tree machine learning algorithm similar to a random forest, but it minimizes residuals based on the previous prediction tree rather than independently minimizing residuals of multiple trees. The gradient boosting algorithm works better when multiple variables are categorical, as dummies for maternity, cardiologist and diabetes enter the model.

Our estimation method allows us to identify mean taste utilities without contamination from the specification of the inertia term. Mean utilities are identified from newcomers who face no state dependence. Once the plan quality for a demographic group is fixed, and conditional on other variables, such as expected-out-of-pocket costs, the total inertia term is identified by excess persistence of incumbents' choices relative to what newcomer-estimated plan qualities would predict. Comparisons between incumbents that are switching plans and those who don not allow us to pin-down the plan-inertia parameter. The insurer-inertia parameter is separately identified from the destination pattern of incumbents who do switch, the extent to which switchers disproportionately choose plans from their previous insurer, beyond what plan characteristics alone predict.

Prices of health insurance plans are potentially endogenous to unobserved characteristics of the plan as insurers set premiums in response to expected demand and unobserved plan quality. To identify the price coefficient, we exploit variation generated by changes in aggregate medical costs that shift insurers' pricing incentives. We use the national growth in medical costs, measured by the Medical CPI. Interacting this supply shifter with the predetermined potential market size outside the region under study generates variation across plans, demographic groups and time. Under the assumption that national cost shocks and predetermined market exposure outside Santiago do not directly affect plan preferences within Santiago conditional on controls, this variation identifies the price coefficient. More information on the construction of the instrument can be found in Appendix B.

Estimation method. We estimate the discrete choice model with a two-step method.First, we use maximum likelihood to estimate the individual-specific parameters. Then, we estimate the plan-level coefficients by two-stage least squares to address price endogeneity (Berry, Levinsohn, and Pakes (1995) and Berry and Haile (2024))[20]. We exploit our setting to better estimate different parameters by splitting the sample into newcomers, who have just arrived in the system and have no state dependence structure, and incumbents, who do have inertia and are choosing an offered plan. The mean taste parameters are cleanly recovered from the newcomers sample who are active choosers and whose observed shares should reflect the quality of the plan. Conditional on these mean taste parameters δ_{gjt}, we can recover the inertia parameters using the incumbent sample in the likelihood, together with individual-specific coefficients (Θ).

$$
\begin{aligned}
u_{ijt} = & \underbrace{(\alpha_0 + \alpha_1 \text{age} + \alpha_2 \text{inc})\, p_{jg(i)t}}_{\text{price}} + \underbrace{\beta_1 x_j}_{\text{attributes}} + \underbrace{\xi_{jg(i)t}}_{\substack{\text{unobs.} \\ \text{attributes}}} + \underbrace{\omega_0\, EOP_{ijt} + \omega_1\, VOP_{ijt}}_{\substack{\text{OOP} \\ \text{beliefs}}} \\
& + \underbrace{\gamma_1 \mathbf{1}\{j = j_{i,t-1}\} + \gamma_2 \mathbf{1}\{f(j) = f(j_{i,t-1})\}}_{\text{inertia}} + \varepsilon_{ijt} \\
u_{ijt} = & \delta_{g(i)jt} + \mu_{ijt}(z_i, j_{i,t-1}) + \varepsilon_{ijt}
\end{aligned}
$$

[20] We prefer a two-stage procedure to account for endogenous prices in a linear specification rather than estimating all parameters jointly with a control function approach, which can be sensitive to mispecification.

In the first step of the estimation we recover $\alpha_1, \alpha_2, \omega_0, \omega_1, \gamma_1, \gamma_2$ (Θ) and δ_{gjt} maximizing the following log-likelihood function:

$$L(\Theta, \delta) = \sum_i \sum_{j \in J_{it}} \mathbf{1}_{ijt} \log s_{ijt}$$

The individuals i in the likelihood include: newcomers, whose inertia terms are zero, market-plan incumbents, who carry both plan and insurer inertia and closed plans holders who switched to a market plan, who carry insurer inertia only and their closed plan is excluded from the choice set, conditional on having switched out of that plan. To recover δ_{gjt} we match the observed market shares among newcomers by market with the theoretical market shares given by s_{gjt} and for a fixed Θ we compute δ_{gjt} through a contraction mapping function. There is one contraction-mapping fixed-point per market, that corresponds to year and age-gender group. Once the individual-specific parameters and the inertia parameters are consistently estimated, we iterate the contraction mapping for all offered plans by matching the observed market shares given by the pooling sample of newcomers and incumbents individuals, conditioning on the estimated preference and inertia parameters $\hat{\Theta}$. The resulting $\hat{\delta}_{gjt}$ is the one used in a second step. This improves on efficiency for the second-step parameters (see Appendix B).

In the second step, we estimate equation (15) by two-stage least squares instrumenting for price with an interaction instrument Z_{gjt}, accounting for the share of individuals outside the region of study (Santiago) for different population groups (defined by income, gender and age) in 2017. Consequently, this will not be correlated with unobserved demand shocks after 2017. For the shift, we use a cost-shifter of medical expenses at the national level. We use the yearly increase in medical costs for the Health CPI from the Chilean Central Bank[21]. Other instruments include insurer-specific balance-sheet information in the year the plan is introduced, such as liquidity and debt. We also include the maximum price adjustment rate the insurer could apply in the year the plan was introduced, as set by the regulator. Another instrument is the potential total cost of the plan outside the individual's demographic group

$$\delta_{gjt} = \beta_1 x_j + \alpha_0 p_{jgt} + \xi_{jt} \tag{15}$$

The standard BLP contraction mapping is inconsistent in the presence of inertia: observed market shares reflect both plan quality and inertia, and the contraction attributes both to δ_{gjt} biasing mean utilities upward for plans with large incumbent bases. We address this by restricting the initial contraction mapping to newcomers, who face no state dependence by construction, resulting in δ_{gjt} that are not contaminated by the inertia term, similarly to Ericson and Starc (2016). This allows us to consistently estimate the individual-specific parameters. Given the estimated inertia parameters, we then re-compute δ_{jgt} from a full-sample contraction mapping that conditions on the consistently estimated inertia parameters. This second contraction incorporates incumbents' choices while netting out their estimated state dependence, delivering more efficient

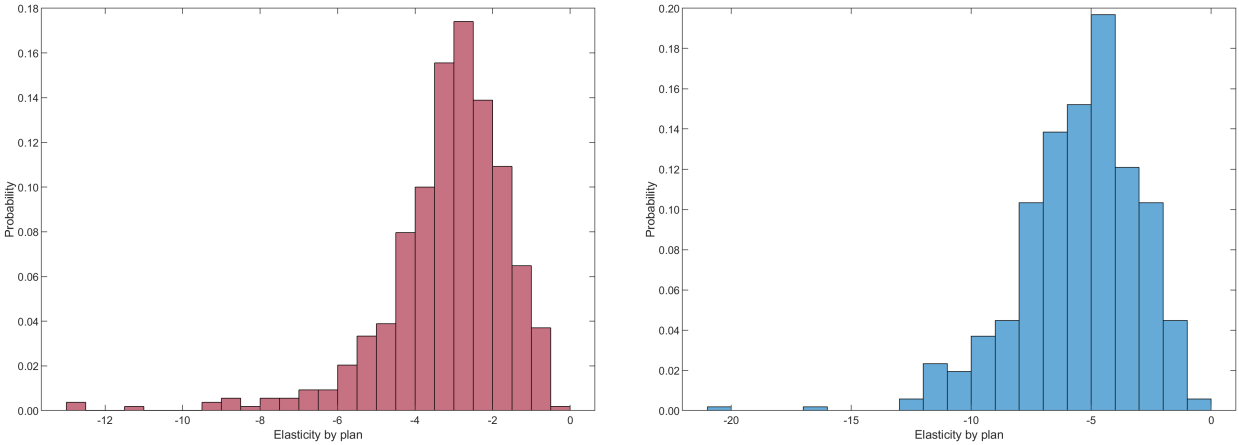
[21] Cuesta, Noton, and Vatter (2019) address endogeneity by instrumenting each insurer's negotiated hospital prices with off-sample closed insurer rates, plan premiums with the rivals' average actuarially fair premiums.

estimates of mean utilities for the instrumental variables regression. The idea follows the logic of the nested pseudo-likelihood (NPL) algorithm of Aguirregabiria and Mira (2002) with two main differences. First, a first step based on conditional choice probabilities (CCPs) would require estimating transition probabilities across approximately 300 prior plans, 300 destination plans, and each demographic cell, most of which contain no observations, whereas the newcomer-based contraction avoids estimating these objects entirely, as newcomers' state variable is degenerate by construction.Second, although our second step mirrors the structural update in NPL (mean utilities are re-derived from the model at the estimated parameters rather than from raw shares), we stop after one update: since consistency is achieved in the first step, further iterations would affect only efficiency.

5.2 Demand Estimation Results

Individuals are price sensitive, and this sensitivity increases with age and decreases with income. They prefer plans in which their expected out-of-pocket costs are smaller, and they also dislike large variance of out-of-pocket costs. Individuals have large switching costs if they want to move to a different insurer. They prefer larger coverage rate plans. Details of the demand estimation results are in Appendix B.

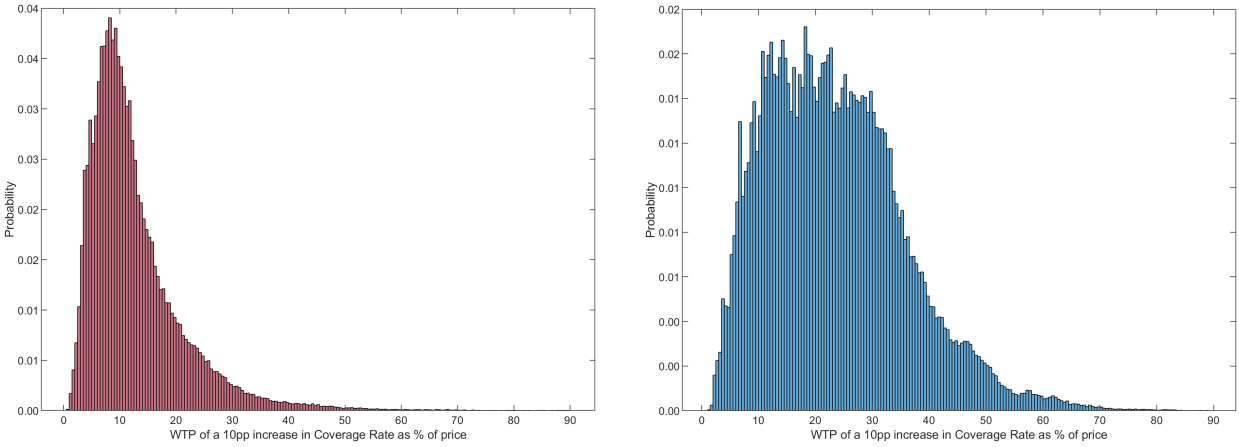
Figure 4: Own-price elasticities per plan.



(a) Female. (b) Male.

Notes: Own-price elasticities per plan aggregated over demographic groups. Elasticities computed from dynamic choice probability taken into account previously chosen plan.

Figure 5: Willingness to pay for Coverage Rate



(a) Female. (b) Male.

Notes: Willingness to pay for a 10-percentage-point increase in the coverage rate, as a percentage of the price, computed for each individual as the coverage-rate coefficient divided by the negative of the price coefficient.

5.3 Supply: constructing the first-order condition

To evaluate the first-order condition (8), the insurer must forecast three objects over the life of the plan: how its enrollment decreases as the plan ages and new plans enter, how the cost of the remaining pool drifts as healthy enrollees leave, and how many of the enrollees a lower price attracts today are still in the plan in later years. These forecasts depend on the future menu of plans, which is the outcome of the dynamic game among insurers. Solving that game would require each insurer to track the identities, prices and pools of thousands of rival plans, which does not describe how prices are set. We use only the necessary condition: at the observed equilibrium, each introduction price satisfies the insurer's first-order condition. We estimate the three objects as transitions on a coarse state of the plan, which describes how the industry moves (Assumption 7), and evaluate the condition along them. This follows the two-step approach of Hotz and Miller (1993), Aguirregabiria and Mira (2007), and Bajari, Benkard, and Levin (2007): transitions come from the data in a first step, and optimality is imposed in a second step. Summarizing the future in a coarse state is close to the oblivious and moment-based approximations to dynamic oligopoly of Weintraub, Benkard, and Van Roy (2010) and Ifrach and Weintraub (2017), although here what is observed at the pricing date, the current menu and the plan's enrollment, is used as it is, and only the future is summarized.

Under Assumption 7 (Appendix F), insurers forecast the plan's future with the estimated transitions, and Assumption 8 closes the far future. Beliefs only need to be consistent with outcomes at the states the market visits, as in the experience-based equilibrium of Fershtman and Pakes (2012) and Dubois and Pakes (2026). We impose optimality on each insurer's own decisions: by Assumption 5, its introduction policy is a cutoff rule in a private shock, and plan introduction enters the estimated transitions through entry intensity. We do not impose that the forecasts are a fixed point of all insurers' best responses. The cost of this approach is that a counterfactual that changes how plans are introduced also changes the forecasts. The commitment counterfactual

therefore holds the menu fixed, and the counterfactual without new plans replaces the forecasts with a model in which future plans are an explicit option in each consumer's choice set, which can be removed (Section 6.1).

Section 5.3.1 writes the approximated condition and explains how its elements are estimated, combining laws of motion estimated in the data with responses to price from the demand model. Section 5.3.2 reports what the condition implies for observed prices.

5.3.1 The approximated condition

As in Section 4, we suppress the rating weights in the integrals. Under Assumption 7, the expectation in (8) conditions on the coarse state $\mathcal{S}_t$, and the first-order condition, written by plan age τ, is

$$0 = \sum_{\tau \geq 0} \beta^\tau \, \mathbb{E}\Big[\underbrace{\int s_{ij}(\tau)\, di}_{S_j(\tau)} + \underbrace{\int (p_j - c_i) \, \frac{ds_{ij}(\tau)}{dp_j} \, di}_{p_j \Omega_j(\tau) - G_j(\tau)} + \underbrace{\sum_{k \in \mathcal{J}_{f,t+\tau} \setminus \{j\}} \int (p_k - c_{ik}) \, \frac{ds_{ik}(\tau)}{dp_j} \, di}_{X_j(\tau)} \;\Big|\; \mathcal{S}_t \Big] \quad (16)$$

$S_j(\tau)$ is the plan's enrollment, $\Omega_j(\tau) = \int ds_{ij}(\tau)/dp_j \, di$ is the response of its enrollment to its base price, $G_j(\tau) = \int c_i \, ds_{ij}(\tau)/dp_j \, di$ is the same response weighted by each enrollee's cost, and $X_j(\tau)$ is the margin the insurer gains or loses in its other plans when the price of j changes. To evaluate them exactly, the insurer would need the menu and every enrollee's choice probabilities in every future year.

The approximated condition replaces each of these objects by a path over the plan's life, built from profiles of plan age that we estimate in the data. Writing the rating weights explicitly and adding the cost loading λ_f, the condition is

$$0 = \sum_{\tau=0}^{\bar{\tau}} \beta^\tau \Big[\underbrace{\widehat{rw}_j(\tau)\, S_j(\tau)}_{\text{revenue on enrollees}} + \underbrace{p_j \, \hat{\rho}_j^{rw}(\tau)\, \Omega_j(\tau)}_{\text{revenue lost at the margin}} - \underbrace{\lambda_f \, G_j(\tau)}_{\text{cost saved at the margin}} + \underbrace{\hat{X}_j(\tau)}_{\text{recaptured margin}} \Big], \quad (17)$$

where $\widehat{rw}_j(\tau)$ is the mean rating weight of the plan's enrollees, $\beta = 0.95$, and the horizon is $\bar{\tau} = 20$ years. A higher base price raises the payment of each enrollee in proportion to their rating weight, so the revenue lost at the margin weights each enrollee who responds to the price by their rating weight. $\hat{\rho}_j^{rw}(\tau)$ is the average rating weight of the enrollees who respond. It differs from the plan's mean weight because the enrollees who respond to the price are mostly young. The individual cost c_i is the claim paid by the plan for enrollee i in the claims data. The insurer's paid benefits in its accounts exceed the claims recorded at the individual level. We scale individual costs by a loading λ_f for each insurer, the slope of the insurer's accounting loss ratio on the loss ratio computed from the claims data, which ranges from 1.08 to 1.33 across insurers (Table A15, column 1). Here the insurer takes the recaptured margin fully into account, and Appendix F.6 estimates how much of

it insurers internalize. The condition is linear in p_j, so it implies a price,

$$p_j^* = \frac{\sum_{\tau=0}^{\bar{\tau}} \beta^\tau \left[\widehat{rw}_j(\tau)\, S_j(\tau) - \lambda_f\, G_j(\tau) + \hat{X}_j(\tau) \right]}{D_j}, \qquad D_j = -\sum_{\tau=0}^{\bar{\tau}} \beta^\tau \hat{\rho}_j^{rw}(\tau)\, \Omega_j(\tau) > 0. \tag{18}$$

Because (18) is a sum over the plan's life, the implied price splits into the price that the year of introduction alone would set at $\tau = 0$, and a future part (Appendix F.7).

Each path follows a law of motion over the age of the plan, written out in Appendix F.1. Enrollment this year is the share of last year's enrollees who stay, plus the new enrollees of the plan: enrollees new to the private system, enrollees from rival insurers, and enrollees from the insurer's other plans. The response to the price has two parts. Part of last year's response is carried forward, net of the enrollees the price displaced who would have come back to the plan anyway. Because the enrollees who respond to the price are more mobile than the average enrollee, the response is carried forward less than enrollment is. On top of it comes the response of this year's choices: every new enrollee chooses the plan actively and responds to the price, while among holders only those close to leaving respond. The cost saved at the margin is the cost of the enrollees the price moves, which is the average cost of the pool plus the difference between the marginal and the average enrollee. We measure this difference with the cost of the enrollees who switch plans relative to the peers they leave behind in the same plan. Appendix Table A14 reports the average gap: switchers cost 3.7 UF a year less than the peers they leave behind. The condition uses its profile by plan age, from -1.7 UF in the first two years of the plan to -11.3 UF at ages 18 and above, estimated by interacting the switch indicator in column (1) with plan-age bins (not reported). The price is therefore disciplined by the cost of the marginal enrollee, not by the average cost of the pool. Finally, the recaptured margin is the margin the insurer earns on the enrollees the price pushes into its other plans.

The inputs of these laws of motion come from two sources. The levels, that is, how many enrollees a plan keeps, how many new enrollees it receives from each origin, how costly they are and where its leavers go, are estimated from the plan-year panel. We assume that every plan follows the same shape over its life and differs only in its level: we estimate the shape for a typical plan and move it to each plan's own observed values (Appendix F.2). Tables A16 and A17 report the estimated profiles, and Figure 13 shows them for one plan. The share of enrollees who stay rises with the age of the plan, new enrollees fall sharply after the first years, the average cost of the pool rises, and the share of leavers who stay with the insurer is highest in the first year. The responses to price come from the demand model, computed from the individual choice probabilities of the 210,742 individuals observed in 2018 and 2019 (Appendix F.3). A plan's price is set once and never changes, so the data contain no variation in a plan's own price over its life, and the response of future enrollment to the price cannot be read from the data. The levels, by contrast, are observed for every plan and every age, and using them directly avoids importing the demand model's errors in levels, such as the rate at which enrollees leave the system, into the price. The static pricing condition that Berry, Levinsohn, and Pakes (1995) and Nevo (2001) use to recover markups combines observed market shares with price derivatives computed from the

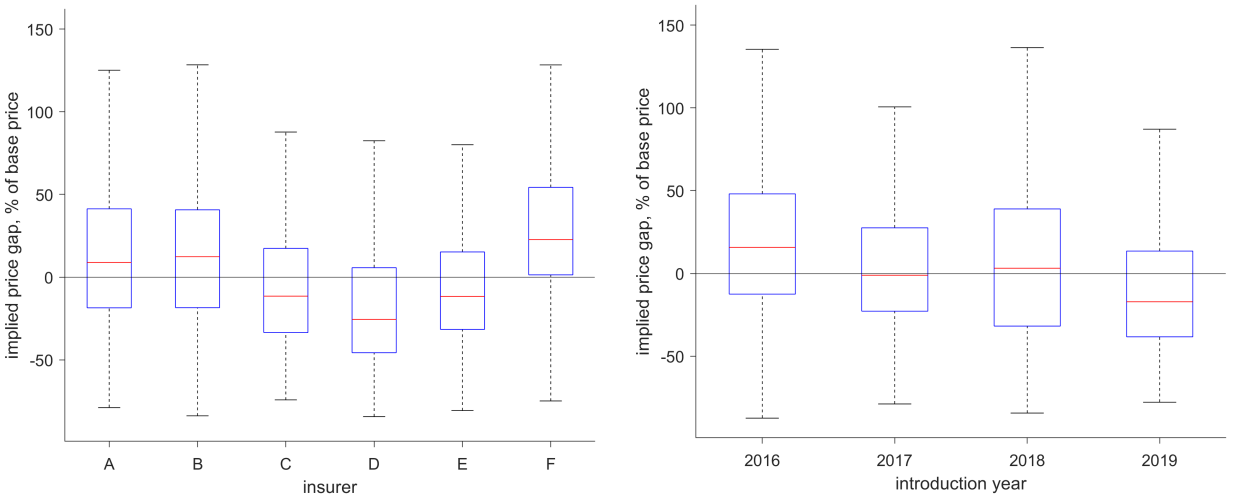
demand model. Our condition needs the plan's enrollment and costs in every future year, which are not observed when the price is set, so we forecast them with the estimated profiles.

To validate the construction, we rebuild the paths of the plans introduced in 2016, which we observe during their first four years, using only their enrollment at introduction and the estimated profiles, and compare them with the observed rows (Figure 13). We also test that the forecast errors of the paths are uncorrelated with the information available at introduction, and the test does not reject ($J = 52.1$, $p = 0.84$, Appendix F.6). The individual transitions computed from the demand model reproduce the observed 2019 choice probabilities of individuals whose 2018 plan is still offered.

5.3.2 What observed prices reveal

We compute the implied price (18) for every plan introduced between 2016 and 2019, and express the difference between the implied and the observed base price as a percentage of the observed price, the price gap. The future accounts for a large part of the price. Splitting the implied price into the part set by the year of introduction and the part set by later years, the later years account for 58% of the base price of the median plan. From this future terms 31% comes from margins recaptured in the insurer's other plans. The condition fits the median plan but not the dispersion across plans, and the difference is larger for higher prices (Figure 6). The approximated condition implies prices that are too high for cheap plans and too low for expensive plans. Two features of the approximation explain the pattern, and both motivate the model of Section H. For cheap plans, the approximation counts the enrollees a higher price sends to the insurer's other plans, but not the enrollees it removes from the pool who would have moved to those plans in later years. This omitted effect is largest for cheap plans, whose enrollees are the most price sensitive and leave for newer plans at the highest rate (Appendix F.6). For expensive plans, the logit demand model cannot rationalize larger margins with an elasticity that only depends on age and income, a limit that remains in the equilibrium model of Section 6.1.2. The profiles are estimated in the observed market and describe enrollment under the current pattern of plan introduction. They are the right objects to rationalize observed prices, but not to predict prices in a market without new plans, where these paths would change. For that counterfactual, Section 6.1.2 solves for prices in a model in which the flows into new plans are generated by choices of individuals.

Figure 6: Approximated FOC gap: observed price vs implied price.



(a) Gap by insurer. Horizon to 20 years.

(b) Gap by year

6 Counterfactual Analysis

6.1 No introduction of new plans

We construct a counterfactual that bans introducing new plans in the future to compute how this affects the plans being priced in 2018. We want to evaluate what happens when we take out the main retention strategy in terms of prices and exit rates. We build a model that reproduces the behavior we documented with the approximated first-order-condition. That analysis showed two patterns. First, plans lose enrollees over time, because enrollees gradually move to newer plans. Second, most individuals who join the industry start in a recently introduced plan. We define a baseline model, different from the approximated condition of Section 5.3.1, that allows us to compare two scenarios: one with future entry and one without future entry. We present the baseline details in Section 6.1.1 and the counterfactual results in Section 6.1.2.

6.1.1 Baseline

We model the future non-observed plans by modifying the choice set of the individuals: we add one option per insurer and per period that represents all the new plans in that period, which we call *new-plan option*. The new-plan option should capture how much of the enrollment of the plan being priced is taken away by new plans. In this way, in the counterfactual we are able to take out these options from the choice set, to see how price channels change and how the exit rates to the outside option vary.

The indirect utility of the new-plan options combines an insurer-specific f mean taste parameter δ_{0f}, with an age-specific a term that represents the period in which the new plan is introduced g_{ia}, a price term built from the measured mean introduction base price of insurer f's new plans, a health-state term capturing that switching to a new plan is easier when healthy $\varphi\,\tilde{c}_{ia}$, and an

inertia term with the demand model's own-insurer coefficient plus an insurer-specific retention parameter $g_{O,f}$:

$$u_{ifa}^{new} \;=\; \delta_{0f} \;+\; g_{ia} \;+\; \hat{\alpha}_i \, b_f^{new} \, \overline{rw}_a \;+\; \varphi \, \tilde{c}_{ia} \;+\; \left(\hat{\gamma}_2 + g_{O,f}\right) \mathbf{1}\{f = f_i\} \;+\; \varepsilon_{ifa}, \tag{19}$$

with $g_{ia} = g_a$ if i already enrolled at a, $g_{ia} = h_a$ i entering the system at a. Demand model parameters $\hat{\alpha}_i$ and $\hat{\gamma}_2$ are fixed at their estimated values from Section 4.1.

Estimation. Holding the demand parameters fixed at their estimates, we estimate $\theta^{new} = \{\delta_{0f}, g_{ia}, \varphi, g_{O,f}\}$ by indirect inference on the observed entry and switching flows. The auxiliary statistics are regression coefficients computed with the same regression on the data ($\hat{m}_k$) and on model-generated flows ($m_k(\theta^{new})$). The estimator is classical minimum distance with diagonal weights:

$$\min_{\theta^{new}} \sum_k \left[w_k (m_k(\theta^{new}) - \hat{m}_k) \right]^2, \qquad w_k = 1/\mathrm{se}_k. \tag{20}$$

We estimate 22 parameters using 26 moments. The moments summarize the observed entry flows in four auxiliary regressions. The first group measures how often enrollees leave their plan for a newer plan each year, and how this depends on the age of the plan and on the insurer. The second group measures, among enrollees who move to a newer plan, the share that stays with their own insurer. The third group measures the share of new joiners that start in a recently introduced plan. The last regression measures how much cheaper the enrollees who move to a newer plan are relative to the enrollees who stay, which disciplines the health term of the new-plan option. Each moment is a regression coefficient estimated on the data, and we weight each one by the inverse of its standard error. The moments act as penalties rather than constraints, and the remaining gaps are informative about what the model cannot reproduce. The main gap is that the model sends too many movers to their own insurer's new plan, so the recapture of enrollees by their own insurer is somewhat overstated. The retention share in the new-plan option between the data and the model has the largest difference in the early years of the plan. The auxiliary regressions are explained in detail in Appendix (H).

6.1.2 Counterfactual: No introduction of new plans.

With the new-plan options parameters fixed at their estimates, we solve for equilibrium prices twice. In the entry scenario, the new-plan options are present, and we solve for the base prices of the plans introduced in 2018. In the no-entry counterfactual we remove the new-plan options and solve for the optimal prices again. In both worlds each insurer sets the base price of its plans to maximize the present value of profits, taking into account that a higher price today pushes some enrollees to other plans, to the new-plan option when it exists, or out of the industry. When an enrollee moves to the insurer's own new-plan option, the insurer keeps part of that enrollee's future margin, and the pricing condition accounts for this recaptured margin. The effect of banning entry is the comparison between price under entry and no entry. Individual costs are simulated from a Markov model of risk states (Appendix G).

We compare the prices computed under the entry scenario with the new-plan option and the observed prices from the data. The model-implied prices are generally lower than the observed prices over all the priced plans. The ratio is the largest in the cheapest tercile (1.13), further away in the middle (0.78) and top (0.59). For the counterfactual exercise, the fit is not important as both sets of prices come from the same model.

Removing entry raises prices. Across the plans priced in both scenarios, no-entry prices are 13.4 percent higher than entry prices when plans are weighted by their discounted enrollment over the life of the plan, and the median plan rises by 16.7 percent. The increase is broad: plans holding 57 percent of enrollment raise their price by more than 5 percent, and plans holding 36 percent by more than 20 percent, while only 9 percent of enrollment is in plans whose price falls. The increase is similar across the price distribution: 16 percent in the cheapest tercile of observed prices, 22 percent in the middle tercile, and 12 percent in the most expensive. Enrollees also leave the system more often. Among women, exit rises by 4.5 percentage points in the first band in which enrollees can leave, and by 2 points in the last band, and among men by 13 to 14 points in the first bands and 2 to 4 in the last. The payment of a 26-year-old rises by 11.9 percent on average. Most of the increase comes from the composition of the marginal enrollee. The markup term, enrollment over its response to price, rises slightly in 86% of plans but falls for the largest ones, by 0.12 UF per month weighted by enrollment. Figure 23 plots each plan's price under no entry against its price under entry.

6.1.3 Policy discussion

The effects of a market without new plans are mainly through prices or less options in the choice set. For a 26-year-old entering the private system, the immediate effect is a higher price: a transfer of 0.274 UF per month would, on average, offset the increase relative to the price they would have paid with new plans. Current enrollees face the higher price as well as the loss of the option to move to a new plan when their current plan becomes relatively unattractive, so compensating them for the price increase alone does not restore what they lose from the smaller choice set. Participation responds on both margins. Among women in plans aged one to four, exit to the public system rises by 4.5 percentage points, with most of this increase coming from the loss of new options rather than from higher prices. Meanwhile, the average cost of existing plans falls by about 4 percent because healthier enrollees who would have moved to new plans remain in their current plans. This reduction in average cost does not translate into lower prices because, prices depend on the cost of individuals who respond to price and on markups rather than only on the average cost of the pool. One possible substitute for new plans is more front-loading, with relatively higher prices at younger ages and lower prices later so that remaining in the plan becomes more attractive over time. We evaluate this by making the regulated age weights larger for 26 years-old by the same percentage and recomputing prices. We do not allow the insurer to choose the age profile because having prices and rating weights as instruments will not be possible to solve. Under our model, the larger rating weight for 26 years old does not translate to lower base price, and as a result increases payments for younger individuals, reduces entry into the

private system, and lowers industry profits by about 7 percent. Front-loading is less effective in this setting because individuals place substantial weight on current prices, so higher payments at young ages discourage entry even when they imply lower prices later. Our results also help explain why insurers continue to introduce new plans. Plan introduction protects each insurer separately but can be costly for the industry as a whole. Insurers introduce new plans to retain healthy enrollees who might otherwise move to rivals. In the model, when one insurer stops introducing plans while its competitors continue, switching falls only slightly, but most of the individuals who would have moved to its new plans instead switch to rival plans. When all insurers stop introducing plans, this competition for healthy enrollees disappears and prices rise by about 13 percent. Profits on existing plans increase for every insurer, but total profits fall for insurers that derive substantial value from new plans and rise for those with little plan introduction This has a direct implication for regulation. Limiting the number of new plans is often proposed as a way to reduce product complexity, but in our model such a limit also weakens competition for healthy enrollees and raises prices.

The shift-share results of Section 3.3 shows that enrollees are more likely to stay with their insurer when the insurer's new plans are close to the plan they hold. The shift-share estimates imply own-insurer new plans at mean closeness lower rival switching by about 7 percent compared with no close own-insurer new plans. The model captures this retention force but represents each insurer's new plans as a single option available to every enrollee, without the closeness dimension. Therefore, introducing new plans when all other rivals are introducing decreases the probability of enrollees switching to rivals in 50%. It therefore places more of enrollees' attachment to the insurer on that option: when it is removed, few switchers remain with the insurer, whereas in the data 57 percent of switchers to older plans remain with the same insurer.

6.2 Two-sided commitment

We compute the equilibrium prices generated by two models in order to compute consumer surplus under each scenario. We then decompose the change in consumer surplus into a flexibility effect and a price effect, and further decompose the price effect into a markup channel and a cost channel, which helps us understand the mechanisms through which the two scenarios generate costs and benefits. We compare the baseline switching environment with a two-sided commitment environment. In the baseline case, individuals may switch plans at each age group while facing a fixed choice set. In the commitment case, individuals choose a plan once, when entering the private sector, and remain enrolled in that plan in all subsequent periods. Individuals face the same initial choice set as in the switching environment, with the outside option also available.

6.2.1 Individual Behavior

Baseline: Switching We assume the choice set J_0 is fixed every period but is large enough that it captures the multiple combinations of characteristics of the different health insurance plans. Individuals are forward-looking: they take the future into account when deciding today, and the previous plan enters utility through the inertia term, so today's choice affects future payoffs and

the individual problem is in general dynamic:

$$V_{i0}(j_{-1}) = E_\epsilon \left[\max_{j \in J_0} \{U_{ij0} + \epsilon_{ij0} + \beta V_{i1}(j)\} \right]. \tag{21}$$

Appendix (D) shows that under assumptions on individuals' beliefs [22] the continuation value $V_{i1}(j)$ decomposes into components that are either common across plans or vary only at the plan and demographic group level. Under these assumptions, the individual dynamic decision is approximated by a static one: the individual behaves as if maximizing current-period utility,

$$\max_{j \in J_0} \; u_{ij0}, \tag{22}$$

with the plan and demographic group component of the continuation value entering mean utility, the individual component scaling the heterogeneous coefficients by a common present-value factor, and γ_1 the per-period cost of switching.

Utility from choosing plan j is $u_{ijt} = U_{ijt} + \varepsilon_{ijt}$, where ε_{ijt} is an idiosyncratic taste shock drawn each period and U_{ijt} is:

$$U_{ijt} = \alpha(z_i)p_{jgt} + \beta_1 x_j + \xi_{jgt} + \omega_0 EOP_{ijt} + \omega_1 VOP_{ijt} + \gamma_1 \mathbf{1}_{\{j = j_{i,t-1}\}} \tag{23}$$

Counterfactual: Commitment In the commitment case, individuals choose a plan when entering the private sector and remain enrolled in that plan in all future periods. Individuals do not have perfect foresight over their future health realizations, but they know the health transition process. For each simulated health path $h^r = \{h^r_\tau\}_{\tau=0}^T$, the discounted value of committing to plan j is

$$W^r_{ij0} = \sum_{\tau=0}^{T} \beta^\tau u_{ij\tau}(h^r_\tau). \tag{24}$$

We make more assumptions on the beliefs of the individuals when choosing their health plan only once, and we argue these are consistent with the assumptions made in the baseline case. We assume individuals have rational expectations about their health status path for their whole lifetime. This translates to the same expected out-of-pocket costs variable as in the switching case[23]. For each future period along simulated path r, utility from plan j is evaluated using the expected out-of-pocket costs implied by health status h^r_τ. The idiosyncratic shock faced by individuals is $\epsilon_{ijt} \sim T1EV$, however, when making the decision at 0 there is a naive persistence belief: at the moment of commitment the agent observes ϵ_{ij0} and believes $\epsilon_{ijt} = \epsilon_{ij0}$ for all $t > 0$, leading to an error term with Type 1 extreme value distribution with scale parameter $\sum_\tau \beta^\tau$. Utility from staying in plan j in period τ is $u_{ij\tau}(h^r_\tau) = U_{ij\tau}(h^r_\tau) + \varepsilon_{ij}$, and:

[22] Individuals expect the current choice set to persist, expectations of future plan-specific utility are common within demographic groups, individual deviations driven by persistent health conditions are stable, and plan choice does not affect future health costs

[23] There are other possible assumptions here: compute one discounted EOP for the whole life given the health status path. Or assume individuals are myopic and $EOP_t = EOP_0$.

$$U_{ij\tau}(h_\tau^r) = \alpha(z_{i\tau})p_{jg\tau} + \beta_1 x_j + \xi_{j\tau} + \omega_0 EOP_{ijg\tau}(h_\tau^r) + \omega_1 VOP_{ijg\tau}(h_\tau^r) \tag{25}$$

The inertia term disappears from the commitment scenario, because there is no re-optimization. The value of committing to plan j is computed by taking expectations over simulated health paths:

$$W_{ij0} = E\left[W_{ij0}^r\right] \tag{26}$$

The individual chooses the plan that maximizes the expected value of commitment:

$$\max_{j \in J_0} W_{ij0} \tag{27}$$

6.2.2 Insurer Pricing

Insurers take the menu of plans as given and choose prices once to maximize discounted profits over the lifetime of enrolled individuals. The effective premium paid at age τ is $P_{j\tau} = p_j \cdot rw_j(\tau)$, where $rw_j(\tau)$ is a deterministic age-rating factor that scales the base premium p_j across age groups. Insurers form expectations over individual costs, future market shares, and the choice set, which we treat as fixed throughout the counterfactual exercise. Predicted costs follow individuals over their lives, with individual health and costs simulated from a Markov model of risk states constructed following Appendix G.

The profit-maximizing first-order-condition for the price of plan j is:

$$\begin{aligned} 0 \; = \mathbf{s}_{jt} \; + \; & \int_{\mathcal{I}} (p_j - c_{ij}) \frac{ds_{ijt}}{dp_j} \, di \; + \; \sum_{m \in \mathcal{J}_f \setminus \{j\}} \int_{\mathcal{I}} (p_m - c_{im,t}) \frac{ds_{imt}}{dp_j} \, di \\ & + \; \sum_{\tau \geq 1} \beta^\tau \left[\mathbf{s}_{j,t+\tau} \; + \; \int_{\mathcal{I}} (p_j - c_i) \frac{ds_{ijt+\tau}}{dp_j} \, di \; + \; \sum_{m \in \mathcal{J}_f \setminus \{j\}} \int_{\mathcal{I}} (p_m - c_{im,t}, t+\tau) \frac{ds_{imt+\tau}}{dp_j} \, di \right] \end{aligned} \tag{28}$$

The first line is the static term: current market share plus own and cross-plan margin weighted derivatives. The second line adds discounted future terms that arise because the single base price p_j has implications in the future market shares. In the status-quo switching scenario, choice probabilities depend on the past choice probabilities due to inertia but also insurers need to predict future choice probabilities. In the commitment counterfactual, the FOC reduces to a discounted sum because the choice probabilities are fixed forever, and the cost path is also known.

The total costs of each plan are endogenous as they are defined by the enrolled individuals. Individual level costs c_i are taken as given, but the cost to the plan depends on who enrolls, which depends on the choice probabilities s_{ij} from the demand model. A lower price attracts a different mix of individuals, changing the plan's expected cost, which feeds back into margins. Equilibrium prices and shares are therefore jointly determined, and we solve for the Nash-Bertrand equilibrium by iterating between the FOC and the demand model until convergence.

The derivative $\partial S_{k,t+\tau}/\partial p_j$ has two channels in the baseline switching environment. The first

is the standard direct price sensitivity: a higher price reduces demand, holding past enrollment fixed. The second operates through inertia. Lower prices today raise current enrollment, which, because inertia makes switching less likely, raises enrollment in all subsequent periods. The full Jacobian is therefore recursive:

$$\frac{\partial S_{k,t}}{\partial p_j} = \underbrace{w_j(t) \sum_i \alpha_i \, s_{ij,t} \big(\mathbf{1}_{k=j} - s_{ik,t} \big)}_{\text{direct price effect}} + \underbrace{T_t^{\text{agg}} \cdot \frac{\partial S_{k,t-1}}{\partial p_j}}_{\text{inertia propagation}} \tag{29}$$

where $T_t^{\text{agg}}(k,m) = \sum_i s_{ik,t} \, s_{im,t-1} / S_{m,t-1}$ is the aggregate conditional transition matrix, which measures how past enrollment in plan m translates into current enrollment in plan k through inertia.

The role of inertia in equilibrium pricing. Equation (29) decomposes the share Jacobian into two components: a direct price effect and an inertia-propagation channel. When $\gamma_1 = 0$, the second term drops out and $\partial S_{k,t+\tau}/\partial p_j$ reduces to the contemporaneous price elasticity at each future date. The FOC then collapses to a discounted average of standard Bertrand conditions: because p_j is set once and applies in every period, the optimal markup balances own and cross price elasticities at each future date, weighted by the age-rating factor and the discount factor. When $\gamma_1 > 0$, today's enrollment propagates into future shares through the aggregate transition matrix T_t^{agg}, so the price of plan j must also account for how a marginal change in p_j shifts the trajectory of future enrollment, not only contemporaneous choice probabilities. This propagation introduces an additional downward force on the optimal price: lowering p_j at introduction raises today's share and, through inertia, raises future shares for many periods, generating discounted future profits from each additional enrollee acquired today. The opposing force is that retained enrollees are partially captive, switching costs lower their effective elasticity in later periods, which raises the per-period margin the firm collects from them and pushes toward higher prices at introduction. The recursive structure of Equation (29) makes both forces explicit: the direct term is the standard elasticity at the moment of price-setting, while the inertia term captures the value of building a captive base whose later attachment to the plan is itself shaped by the same price. The net effect on equilibrium prices depends on which channel dominates and is in general ambiguous.

6.2.3 Welfare

To compute the consumer surplus difference between the status-quo and the commitment counterfactual, we allow individual characteristics to evolve exogenously so that U_{ijt} varies over time. Under assumptions 1-3, the Bellman equation for the switching case separates across periods: the continuation value does not depend on the current choice j and drops out of the maximization. Consumer surplus for individual i under switching is therefore the sum of per-period log-sum terms:

$$CS_i^{\text{switch}} = \sum_{t=1}^{T} \beta^{t-1} \log \left(\sum_j \exp \left\{ U_{ijt}\left(p^{\text{switch}}\right) \right\} \right) + \left(\sum_{t=1}^{T} \beta^{t-1} \right) \gamma_E \tag{30}$$

We set the inertia parameter γ_1 to zero when computing consumer surplus. Inertia is relevant for predicting choices, but we treat it as a behavioral friction rather than a true utility component, so it does not enter the welfare calculation. This follows the standard treatment of switching costs in the insurance-choice literature (e.g. Handel (2013)). For the committed individual, the choice at entry maximizes the discounted sum of per-period utilities. The option value of re-optimizing is lost, so the log-sum is taken over plans evaluated at their discounted average value rather than period by period:

$$CS_i^{\text{commit}} = \left(\sum_{t=1}^{T} \beta^{t-1}\right) \log\left(\sum_j \exp\left\{\frac{\sum_{t=1}^{T} \beta^{t-1} U_{ijt}\big(p^{\text{commit}}\big)}{\sum_{t=1}^{T} \beta^{t-1}}\right\}\right) + \left(\sum_{t=1}^{T} \beta^{t-1}\right) \gamma_E \tag{31}$$

Decomposition with endogenous prices. Prices differ across regimes because lock-in changes pricing incentives. Let p^{switch} and p^{commit} denote equilibrium prices under the status-quo and commitment. The welfare difference decomposes as

$$\Delta CS_i = \underbrace{CS^{\text{switch}}(p^{\text{switch}}) - CS^{\text{commit}}(p^{\text{switch}})}_{\text{(A) Flexibility value} \geq 0} + \underbrace{CS^{\text{commit}}(p^{\text{switch}}) - CS^{\text{commit}}(p^{\text{commit}})}_{\text{(B) Price effect, ambiguous}}. \tag{32}$$

Term (A) is the pure option value of re-optimizing when health changes. Term (B) captures the net price impact: markups and costs. Defining p^{mc} as the hypothetical price with commitment markups but status-quo costs, the price effect further decomposes into

$$\underbrace{CS^{\text{commit}}(p^{\text{switch}}) - CS^{\text{commit}}(p^{mc})}_{\text{(B1) Market power harm} \geq 0} + \underbrace{CS^{\text{commit}}(p^{mc}) - CS^{\text{commit}}(p^{\text{commit}})}_{\text{(B2) Cost saving benefit} \leq 0} \tag{33}$$

Results. We find the switching scenario provides higher consumer surplus than the commitment case. The total consumer surplus difference is 5.5 UF per year per individual, which is approximately 2 months of premiums for a 26 year old. The difference is mainly driven by the value of flexibility when enrollees can switch. As health conditions change, the value of moving to a better-matched plan is high, even when individuals have information about the future. The price effect goes in the other direction: it reduces consumer surplus under switching because prices are higher than in the commitment counterfactual. When we decompose this number, the savings from lower aggregate plan costs over the life-cycle are small, and the markup channel is the dominant force on the supply side. Although prices are set once in both cases, the demand elasticity insurers face at that moment differs across them. Under switching, consumer inertia gives insurers a partial source of market power over their existing enrollees: switching costs allow firms to maintain markups over customers whom rivals would have to compensate to attract. Under commitment, the cohort makes a single choice in period 0 and insurers must compete for the entire discounted lifetime value of the customer at that moment. This concentration of competition in the first period raises the elasticity of demand insurers face when setting prices, pushing markups

below the switching levels. The logic is the classical switching-cost mechanism (Klemperer (1995) and Farrell and Klemperer (2007)): inertia raises average prices in the switching scenario because firms can extract markups from locked in individuals after the first period, while under commitment enrollees are also locked in but their entire lifetime relationship is priced competitively in period 0, leaving no separate lock-in margin to extract later.

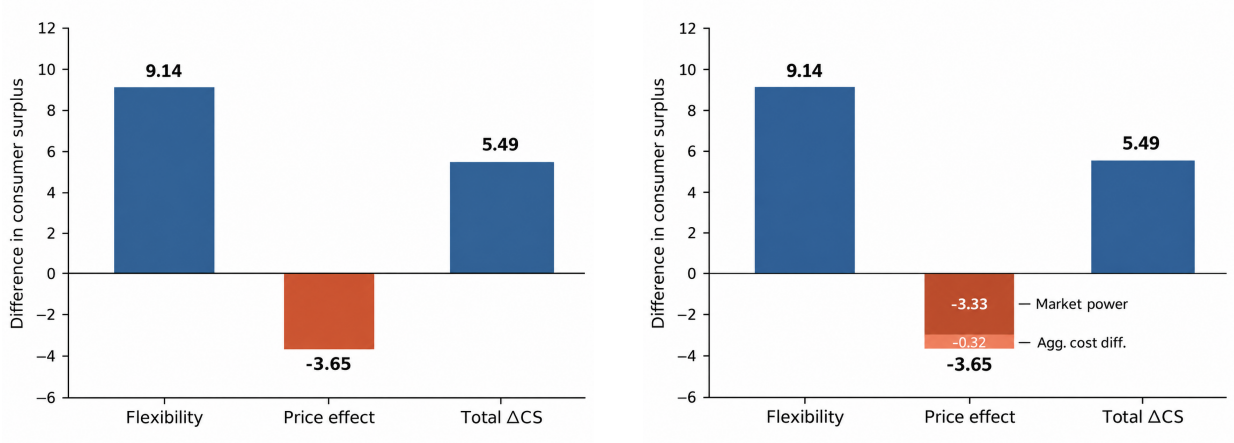


Figure 7: Consumer surplus decomposition

The role of beliefs. To understand the source of the flexibility gain, we recompute the decomposition under an alternative assumption on consumer beliefs. In the main specification, consumers are forward-looking over expected out-of-pocket costs: when choosing at age 26 they anticipate that EOP will evolve with age as their risk profile changes. In the alternative specification, for the commitment case, consumers treat their current EOP as permanent over the lifetime, removing forward-looking beliefs over health costs evolution while leaving all other elements of the model unchanged. The total welfare comparison flips sign across specifications: switching dominates commitment by 5.49 UF per year under forward-looking beliefs, but commitment dominates switching by 0.62 UF per year under no foresight. The flip is driven almost entirely by the flexibility term, which goes from 9.14 to 2.92 UF, while the price effect is essentially unchanged at around -3.5 UF. The option value of re-optimizing is large when consumers anticipate that their health needs will change. A forward-looking 26-year-old understands that committing locks them into a plan that may become poorly matched as their expected costs rise with age, so the ability to switch is highly valuable. Under myopic beliefs over EOP, consumers do not perceive commitment as risky, they evaluate a lifetime plan as if today's health state were permanent, and the implied cost of giving up re-optimization is therefore much smaller. The 2.9 UF of flexibility value that survives in the no-foresight specification is the residual Jensen option value generated by idiosyncratic taste shocks alone, while the additional 6 UF in the main specification reflects the information value of anticipating health costs evolution. The negative total in the no-foresight case then follows mechanically: when the perceived flexibility gain collapses, the largely supply-driven price effect dominates, and myopic consumers are better off under commitment because

they obtain lower markups without internalizing the value of the re-optimization they forgo.

7 Conclusion

This paper studies how long-term contracts shape competition and consumer outcomes in health insurance markets. While these contracts provide protection against reclassification risk, they also create incentives for dynamic adverse selection as low-risk individuals leave contracts over time. We document that in the Chilean market, insurers respond to this problem not only through pricing but also through product design, regularly introducing new plans to retain profitable enrollees.

We show that low-risk individuals are more likely to switch plans and that their exit increases the average cost of remaining enrollees, a pattern consistent with dynamic adverse selection. At the same time, we show that the introduction of new plans increases retention, suggesting that insurers use product proliferation as a tool to mitigate these compositional changes. These findings highlight an important mechanism through which firms manage risk pools in regulated environments where pricing flexibility is limited.

To quantify the implications of these behaviors, we develop and estimate a demand and supply model that incorporates switching costs, risk preferences, and insurers' dynamic incentives to introduce new plans. The model allows us to evaluate how different regulatory constraints affect both consumer welfare and insurer behavior.

Our analysis points to a central trade-off. Policies that limit switching can stabilize risk pools and reduce dynamic adverse selection, but they also reduce consumers' ability to adjust their coverage as their needs evolve. Conversely, allowing switching improves matching but exacerbates selection. The introduction of new plans partially offsets this tension by enabling insurers to retain low-risk individuals, but it may also lead to increased product complexity and potential distortions in plan design.

These results have important implications for regulation. Restricting insurers' ability to adjust prices or contracts does not eliminate selection forces; instead, it shifts them toward other strategic margins such as product proliferation. Effective policy design should therefore account for both pricing and product design incentives.

More broadly, our findings highlight the importance of considering dynamic interactions between consumer behavior and firm strategy when evaluating health insurance regulation. Long-term contracts can mitigate some inefficiencies, but their overall performance depends critically on how firms respond to the constraints they impose.

References

Abaluck, J., & Adams-Prassl, A. (2021). What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses*. *The Quarterly Journal of Economics*, *136*(3), 1611–1663.

Abaluck, J., & Gruber, J. (2011). Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program. *American Economic Review*, *101*(4), 1180–1210.

Abaluck, J., & Gruber, J. (2016). Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance. *American Economic Review*, *106*(8), 2145–84.

Aguirregabiria, V., & Mira, P. (2002). Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models. *Econometrica*, *70*(4), 1519–1543.

Aguirregabiria, V., & Mira, P. (2007). Sequential Estimation of Dynamic Discrete Games. *Econometrica*, *75*(1), 1–53.

Aizawa, N., & Ko, A. (2025). Dynamic Pricing Regulation and Welfare in Insurance Markets. *Journal of Political Economy*, *133*(8), 2371–2413.

Arcidiacono, P., & Miller, R. A. (2011). Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity. *Econometrica*, *79*(6), 1823–1867.

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

Atal, J. P., Fang, H., Karlsson, M., & Ziebarth, N. R. (2019). Exit, Voice, or Loyalty? An Investigation Into Mandated Portability of Front-Loaded Private Health Plans. *Journal of Risk and Insurance*, *86*(3), 697–727.

Atal, J. P., Fang, H., Karlsson, M., & Ziebarth, N. R. (2025). German Long-Term Health Insurance: Theory Meets Evidence. *Journal of Political Economy*, *133*(6), 1840–1885.

Bajari, P., Benkard, C. L., & Levin, J. (2007). Estimating Dynamic Models of Imperfect Competition. *Econometrica*, *75*(5), 1331–1370.

Berry, S., & Haile, P. A. (2024). Nonparametric Identification of Differentiated Products Demand Using Micro Data. *Econometrica*, *92*(4), 1135–1162.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, *63*(4), 841–890.

Bourreau, M., Sun, Y., & Verboven, F. (2021). Market Entry, Fighting Brands, and Tacit Collusion: Evidence from the French Mobile Telecommunications Market. *American Economic Review*, *111*(11), 3459–99.

Brown, Z., Egan, M., Jeon, J., Jin, C., & Wu, A. (2025). Why Do Index Funds Have Market Power? Quantifying Frictions in the Index Fund Market. *Working Paper*.

Cao, Z., & McGuire, T. G. (2003). Service-level Selection by HMOs in Medicare. *Journal of Health Economics*, *22*(6), 915–931.

Carey, C. (2017). Technological Change and Risk Adjustment: Benefit Design Incentives in Medicare Part D. *American Economic Journal: Economic Policy*, *9*(1), 38–73.

Chade, H., Marone, V. R., Starc, A., & Swinkels, J. (2022, October). *Multidimensional Screening and Menu Design in Health Insurance Markets* (Working Paper No. 30542). National Bureau of Economic Research.

Conlon, C., & Mortimer, J. H. (2021). Empirical properties of diversion ratios. *The RAND Journal of Economics*, *52*(4), 693–726.

Coughlin, M. (2023). Insurance Choice with Non-Monetary Plan Attributes: Limited Consideration in Medicare Part D. *Working Paper*.

Cuesta, J. I., Noton, C., & Vatter, B. (2019). Vertical Integration between Hospitals and Insurers. *Working Paper*.

Cuesta, J. I., & Tebaldi, P. (2025). Public Procurement vs. Regulated Competition in Selection Markets. (34141).

Doraszelski, U., & Satterthwaite, M. (2010). Computable Markov-perfect industry dynamics. *The RAND Journal of Economics*, *41*(2), 215–243.

Dubé, J.-P., Hitsch, G. J., & Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics*, *41*(3), 417–445.

Dubois, P., & Pakes, A. (2026, April). *Pharmaceutical Advertising in Dynamic Equilibrium* (Working Paper No. 35025). National Bureau of Economic Research.

Ericson, K., & Starc, A. (2016). How product standardization affects choice: Evidence from the Massachusetts Health Insurance Exchange. *Journal of Health Economics*, *50*, 71–85.

Ericson, K., & Sydnor, J. (2017). The Questionable Value of Having a Choice of Levels of Health Insurance Coverage. *The Journal of Economic Perspectives*, *31*(4), 51–72.

Farrell, J., & Klemperer, P. (2007). Coordination and Lock-In: Competition with Switching Costs and Network Effects. In M. Armstrong & R. Porter (Eds.). Elsevier.

Fershtman, C., & Pakes, A. (2012). Dynamic Games with Asymmetric Information: A Framework for Empirical Work. *The Quarterly Journal of Economics*, *127*(4), 1611–1661.

Fleitas, S. (2017). Dynamic Competition and Price Regulation When Consumers Have Inertia: Evidence from Medicare Part D. *Working Paper*.

Geruso, M., Layton, T., & Prinz, D. (2019). Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges. *American Economic Journal: Economic Policy*, *11*(2), 64–107.

Ghili, S., Handel, B., Hendel, I., & Whinston, M. D. (2024). Optimal Long-Term Health Insurance Contracts: Characterization, Computation, and Welfare Effects. *The Review of Economic Studies*, *91*(2), 1085–1121.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, *110*(8), 2586–2624.

Handel, B. (2013). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *American Economic Review*, *103*(7), 2643–82.

Handel, B., & Kolstad, J. (2015). Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare. *American Economic Review*, *105*(8), 2449–2500.

Handel, B., & Kolstad, J. (2022). The Affordable Care Act After a Decade: Industrial Organization of the Insurance Exchanges. *Annual Review of Economics*, *14*(Volume 14, 2022), 287–312.

Handel, B., Kolstad, J., & Spinnewijn, J. (2019). Information Frictions and Adverse Selection: Policy Interventions in Health Insurance Markets. *The Review of Economics and Statistics*, *101*(2), 326–340.

Harris, M., & Holmstrom, B. (1982). A Theory of Wage Dynamics. *The Review of Economic Studies*, *49*(3), 315–333.

Heiss, F., McFadden, D., Winter, J., Wuppermann, A., & Zhou, B. (2021). Inattention and Switching Costs as Sources of Inertia in Medicare Part D. *American Economic Review*, *111*(9), 2737–81.

Hendel, I., & Lizzeri, A. (2003). The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance*. *The Quarterly Journal of Economics*, *118*(1), 299–328.

Herring, B., & Pauly, M. V. (2006). Incentive-compatible guaranteed renewable health insurance premiums. *Journal of Health Economics*, *25*(3), 395–417.

Hotz, V. J., & Miller, R. A. (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. *The Review of Economic Studies*, *60*(3), 497–529.

Ifrach, B., & Weintraub, G. Y. (2017). A Framework for Dynamic Oligopoly in Concentrated Industries. *The Review of Economic Studies*, *84*(3), 1106–1150.

Klemperer, P. (1995). Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade. *The Review of Economic Studies*, *62*(4), 515–539.

Lucarelli, C., Prince, J., & Simon, K. (2008). *The Welfare Impact of Reducing Choice in Medicare Part D: A Comparison of Two Regulation Strategies* (NBER Working Papers No. 14296). National Bureau of Economic Research, Inc.

Mackay, A., & Remer, M. (2026). Mergers in Markets with Consumer Inertia.

Mahoney, N., & Weyl, E. G. (2017). IMPERFECT COMPETITION IN SELECTION MARKETS. *The Review of Economics and Statistics*, *99*(4), pp. 637–651.

Nevo, A. (2001). Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, *69*(2), 307–342.

Pakes, A., Ostrovsky, M., & Berry, S. (2007). Simple estimators for the parameters of discrete dynamic games. *The RAND Journal of Economics*, *38*(2), 373–399.

Rey, P., & Salanie, B. (1990). Long-term, Short-term and Renegotiation: On the Value of Commitment in Contracting. *Econometrica*, *58*(3), 597–619.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, *55*(5), 999–1033.

Serna, N. (2025). Determinants of Provider Networks: Risk Selection vs. Administrative Costs. (34141).

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

Veiga, A., & Weyl, E. G. (2015). Product Design in Selection Markets. *Microeconomics: Production*.

Weintraub, G. Y., Benkard, C. L., & Van Roy, B. (2010). Computational Methods for Oblivious Equilibrium. *Operations Research*, *58*(4-part-2), 1247–1265.

8 Appendix

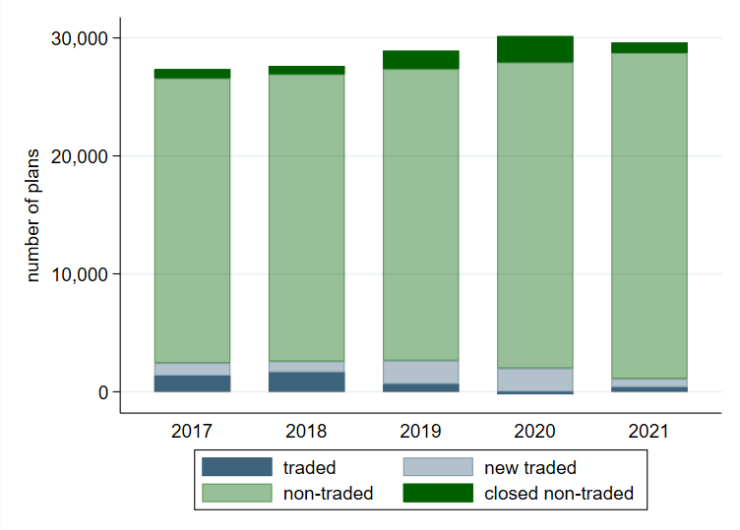
A Data and Descriptives

Table A1: Plan characteristics, 2018 cross-section

	Mean	SD	Min	Max
Coverage rate	0.6125	0.1535	0.0221	1.0000
Preferred network	0.6551	0.4753	0.0000	1.0000
Max coverage	3.44e+04	8.91e+04	510.0000	3.00e+05
Monthly premium (UF)	6.4100	5.7890	0.4674	97.7850

Notes: Coverage rate of the plan is percentage covered by the insurer. Preferred network is 1 if the plan has better coverage rates within a set of fixed providers. Max coverage is the maximum amount covered by year in UF. Monthly premium in UF.

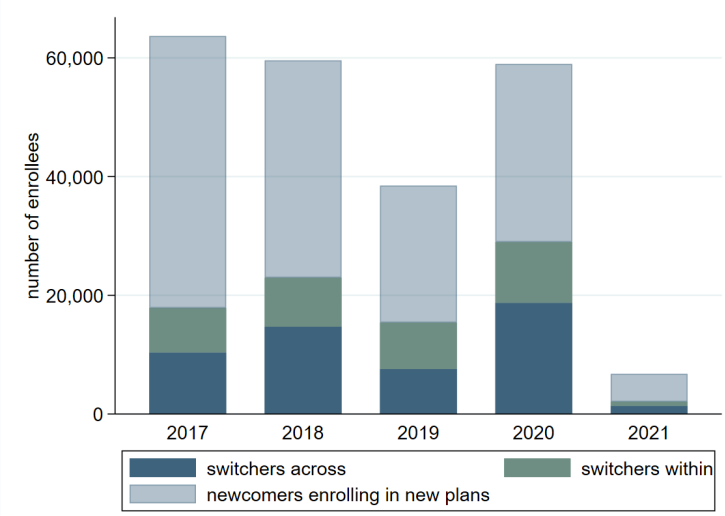
Figure 8: Number of plans by portfolio type



Notes: Number of plans by year and portfolio type: offered, offered and newly introduced, non-offered, and non-offered and closed in that year.

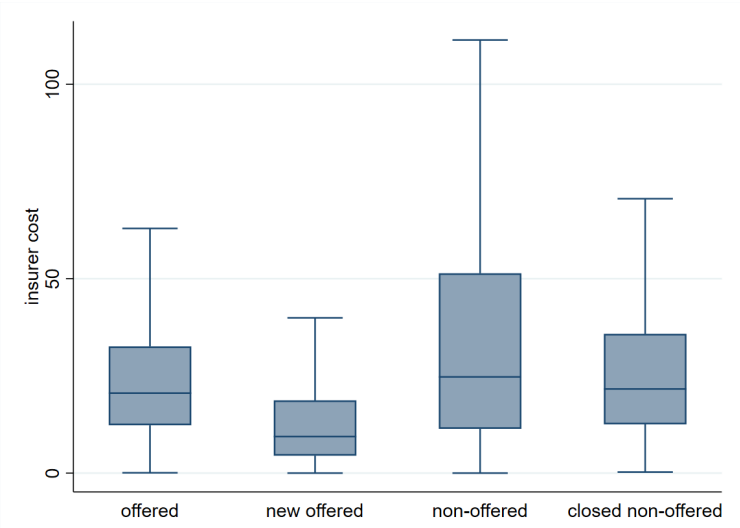
Figure (8) provides descriptive evidence on the evolution and persistence of plans in the market by portfolio type, distinguishing between offered plans, newly introduced offered plans, non-offered plans, and non-offered plans that are closed in a given year. This decomposition illustrates the extent of plan entry and exit over time. Figure (10) compares the distribution of insurer costs across these portfolio types before 2020. Together, these figures describe how plan availability, plan age, and cost distributions vary across offered and non-offered plans.

Figure 9: Number of enrollees in new plans: switchers (within and across insurers) and newcomers



Notes: Number of individuals enrolled in plans that are offered for the first time in the indicated year, separated for switchers within the same insurer, switchers across insurers and newcomers. 2021 data is only half a year.

Figure 10: Insurers cost by portfolio type



Notes: Distribution of insurers cost over plans by portfolio type of plan: non-new offered and new, non offered, and non traded and closed in that period. Only before 2020.

A.1 Empirical Facts

Table A2: Determinants of switching

	Pr(switching)
Price adjustment	0.227***
	(0.005)
Low risk	0.011***
	(0.001)
Maternity	0.048***
	(0.002)
Endocrinology	−0.042
	(0.037)
Gynecology	−0.022***
	(0.002)
Cardiology	−0.009***
	(0.001)
Cancer	−0.010
	(0.019)
Diabetes	−0.012***
	(0.001)
Renal	0.011
	(0.041)
Year FE	Yes
Insurer FE	Yes
Observations	2,454,990

Notes: Standard errors in parentheses. The specification also includes age dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.2 Shift-share results

In Table A3 each row re-runs the baseline two-stage least squares specification of Table 2, column (3), changing only the instrument. The outcome is staying with the same insurer conditional on switching, the controls are the dependents and chronic-condition indicators, and the fixed effects are aggregate plan, insurer × year and age × gender × year. Columns (1)–(3) drop group k from the instrument, $Z_{ct}^{(-k)} = \sum_{\ell \notin k} s_{\ell c,2016}\, g_{\ell t}$, and columns (4)–(6) use only group k, $Z_{ct}^{(k)} = \sum_{\ell \in k} s_{\ell c,2016}\, g_{\ell t}$, where ℓ indexes the age and gender bins and group k pools the bins reported in the row.

Table A3: Overidentification: Leave-One-Group-Out and Only-k Estimates, All Demographic Groups

	Leave group out			Only group k		
	$\hat{\beta}$	SE	F	$\hat{\beta}$	SE	F
31–35, Male	0.01748	0.00850	28.48	0.00880	0.00672	31.01
36–40, Male	0.02713	0.02029	16.90	0.01465	0.01276	29.79
76+, Male	0.02786	0.01273	18.06	-0.01417	0.01203	6.39
26–30, Male	0.01886	0.00889	28.13	0.01103	0.01332	13.78
36–40, Female	0.02085	0.01059	14.63	0.02289	0.01387	19.94
41–45, Female	0.02422	0.01214	21.91	0.01368	0.00990	21.08
71–75, Male	0.02508	0.01210	20.92	-0.00793	0.01174	9.99
56–60, Male	0.02322	0.01231	21.05	0.00556	0.00966	13.15
71–75, Female	0.02438	0.01170	23.26	-0.01567	0.02031	3.93
76+, Female	0.02189	0.01099	23.27	0.01497	0.01514	13.30
51–55, Male	0.02075	0.01013	26.28	-0.00471	0.02019	9.35
41–45, Male	0.02160	0.01157	25.27	0.02226	0.01107	15.15
16–25, Male	0.02041	0.01037	26.01	0.00621	0.01227	13.92
51–55, Female	0.02148	0.01127	26.91	-0.06167	1.02854	0.01
26–30, Female	0.02041	0.01149	26.51	0.03449	0.01550	3.27
31–35, Female	0.01865	0.00944	24.17	0.03927	0.02098	9.18
66–70, Male	0.02162	0.01043	24.34	0.02568	0.13547	0.63
56–60, Female	0.02118	0.01150	26.78	0.03241	0.02333	1.89
46–50, Male	0.02155	0.01125	23.10	0.09947	0.67284	0.02
16–25, Female	0.02194	0.01105	24.56	-0.00148	0.04081	1.33
61–65, Male	0.02132	0.01086	24.43	-0.00615	0.02046	4.09
46–50, Female	0.02146	0.01111	26.87	0.02817	0.02117	1.39
61–65, Female	0.02164	0.01118	24.84	0.02244	0.01773	4.45
66–70, Female	0.02178	0.01103	24.13	-0.01694	0.09388	2.21
Baseline	0.02166	0.01092	24.26	.	.	.

Notes: Each row is a regression of the SSIV, under a different instrument. The regression is staying within the same insurer conditional on switching on covariates as in baseline main regression. Columns 1–3 use $Z_{ct}^{(-k)} = \sum_{j \neq k} s_{jc,2016} g_{jt}$, which drops every elemental instrument inside pooled bin k; columns 4–6 use the single-group instrument $Z_{ct}^{(k)} = \sum_{j \in k} s_{jc,2016} g_{jt}$. Groups are ordered by Rotemberg weigths from Table A5. First-stage F is the Kleibergen-Paap statistic.

Table A4: Robustness of the Entry-Intensity Effect to More Demanding Specifications

	(1)	(2)	(3)	(4)
Entry intensity	0.02166**	0.01291	0.01375**	0.01519
	(0.01092)	(0.01200)	(0.00609)	(0.01003)
Kleibergen–Paap F	24.26	12.42	36.47	24.25
Cluster time trends	No	Linear	No	No
Class × year FE	No	No	Yes	No
Extra individual controls	No	No	No	Yes
Observations	251,717	251,717	251,716	251,717
Clusters	260	260	260	260

Notes: Two-stage least squares; entry intensity is instrumented with the shift-share instrument Z_{ct}. All columns include aggregate-plan, insurer × year and age × gender × year fixed effects, and standard errors are clustered at the aggregate-plan level. Column (1) is the baseline, with an indicator for having dependents and chronic-condition indicators. Column (2) adds aggregate-plan-specific linear time trends. Column (3) adds plan-class × year fixed effects. An aggregate-plan × year fixed effect cannot be included because the instrument varies at that level, and with three years an aggregate-plan quadratic trend is saturated. Column (4) replaces the chronic-condition indicators with chronic-condition × year fixed effects and adds tenure in the previous plan. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Rotemberg Weights for the Shift-Share Instrument

	Year t	$\hat{\alpha}_{kt}$	g_{kt}	$\hat{\beta}_{kt}$	$\hat{\alpha}_{kt}\hat{\beta}_{kt}$
Ten largest positive weights					
36–40, Male	2018	0.8842	0.0803	-0.0043	-0.0038
76+, Male	2019	0.6745	0.0874	0.0117	0.0079
36–40, Female	2019	0.5246	0.0590	0.0185	0.0097
76+, Female	2019	0.3655	0.0559	0.0147	0.0054
41–45, Female	2019	0.3591	0.0735	0.0043	0.0015
71–75, Male	2019	0.3006	0.0737	-0.0154	-0.0046
61–65, Male	2019	0.2432	0.0470	-0.0523	-0.0127
56–60, Male	2019	0.2360	0.0510	-0.0374	-0.0088
26–30, Male	2017	0.2141	0.0540	-0.0576	-0.0123
41–45, Male	2018	0.1607	0.0552	-0.0269	-0.0043
Five most negative weights					
31–35, Male	2019	-0.7027	0.0293	-0.0111	0.0078
26–30, Male	2019	-0.5386	0.0412	-0.0163	0.0088
36–40, Male	2019	-0.3635	0.0246	-0.0141	0.0051
76+, Male	2017	-0.2211	0.0560	-0.0380	0.0084
76+, Female	2017	-0.1971	0.0599	0.0405	-0.0080

Notes: Rotemberg weights (Goldsmith-Pinkham, Sorkin, & Swift, 2020) decompose the shift-share estimate into the contribution of each demographic group × year shift: $\hat{\beta} = \sum_{k,t} \hat{\alpha}_{kt}\hat{\beta}_{kt}$. Here $\hat{\alpha}_{kt}$ is the weight, $\hat{\beta}_{kt}$ is the just-identified estimate that uses only the exposure to group k in year t, and g_{kt} is the national cost growth of group k in year t. The last column is each cell's contribution to the estimate. Because $\hat{\alpha}_{kt}\hat{\beta}_{kt} = g_{kt}(Z'_{kt}\tilde{y})/(G'Z'\tilde{x})$, where Z_{kt} is the vector of 2016 shares of group k across aggregate plans, Z and G stack the shares and the shifts, and $\tilde{y}$ and $\tilde{x}$ are the outcome and entry intensity net of the controls, the cell's own first stage cancels from the product. Cells with weak first stages therefore have extreme $\hat{\alpha}_{kt}$ and $\hat{\beta}_{kt}$ without having large contributions. The decomposition is computed at the aggregate-plan × year level, partialling out aggregate-plan and insurer × year fixed effects and the cell means of the controls, so the contributions sum to 0.029 rather than the individual-level estimate of 0.022 in Table 2. Groups are five-year age bins by gender. The decomposition uses all 30 bins. For reporting, the 16–20 bin is pooled into 16–25 and the 81–85 and 86+ bins into 76+; pooling does not change the total, because the weights are additive. Shares are computed on all enrollees aged 16 and over, because entry responds to the costs of the whole aggregate plan, while the estimation sample is restricted to ages 21–75. The 76+ bins account for 18.5 percent of the variance of Z.

Table A6: Placebo: Post-2016 Cost Shifts and the Pre-Period Outcome

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Z	2.33440*	5.29368***	1.76691	4.82824***	2.91772*	1.24064	1.20389
	(1.31131)	(1.82018)	(1.98189)	(1.64037)	(1.67236)	(1.91141)	(2.01410)
Outcome sample	2017–2019	2016	2016	2016	2016	2016	2016
Shift year in Z	contemp.	2017	2018	2019	2017	2018	2019
Plan characteristics				No	Yes	Yes	Yes
Observations	252,937	75,540	75,540	75,540	75,540	75,540	75,540
Clusters	260	234	234	234	234	234	234

Notes: Reduced-form regressions of staying with the same insurer, conditional on switching, on the shift-share instrument. All columns control for an indicator for having dependents and chronic-condition indicators, and standard errors are clustered at the aggregate-plan level. Column (1) is the reduced form on the estimation sample with the baseline fixed effects. Columns (2)–(7) use the 2016 cross-section, whose outcome is the 2016–2017 transition, and an instrument that combines the 2016 shares with the cost growth of a later year. In a single cross-section aggregate-plan fixed effects would absorb the instrument, so these columns include insurer and age × gender fixed effects instead. Columns (5)–(7) add plan characteristics (mean coverage, annual price, coverage cap, maternity restriction and a preferred-provider indicator) to approximate the aggregate-plan fixed effects of the baseline. Column (2) is not a falsification test, because 2017 cost growth drives the 2017 plan introductions that 2016 switchers face; columns (3), (4), (6) and (7) are the informative ones. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3 Switching and reclassification

Table A7: Reclassification: margin decomposition by group and price tercile

			Mean change, $t \to t+1$ (annual UF)			
Group	Tercile	N	Δ price paid	Δ claims cost	Δ margin	Claims at t
Stayer	T1	646,785	0.81	1.29	-0.48	14.89
			(0.16)	(0.16)	(0.20)	
	T2	653,573	1.43	1.72	-0.29	21.38
			(0.33)	(0.16)	(0.30)	
	T3	670,566	2.87	2.17	0.70	39.51
			(0.58)	(0.22)	(0.54)	
Own-switcher	T1	31,332	4.56	8.20	-2.62	12.11
			(0.95)	(0.83)	(0.55)	
	T2	40,533	-2.68	5.77	-6.36	17.50
			(0.72)	(0.71)	(0.77)	
	T3	41,158	-21.57	0.19	-18.59	30.42
			(2.01)	(1.35)	(1.79)	
Cross-switcher	T1	43,595	7.08	4.23	3.52	7.57
			(0.52)	(0.38)	(0.43)	
	T2	47,675	0.31	2.89	-1.48	10.75
			(0.52)	(0.38)	(0.41)	
	T3	43,437	-12.05	0.57	-10.70	19.61
			(1.42)	(0.49)	(0.83)	
Exit	T1	97,655				10.50
	T2	77,618				16.77
	T3	64,272				31.16

Notes: Cell means by group and price tercile. Cluster-robust standard errors (insurer × aggregate-plan) in parentheses. Same individual, origin year $t \to t+1$, annual UF (12 × monthly paid price minus insurers costs). Price terciles are computed within (age band × gender × year t) cells, so T1/T3 are relative to enrollees of the same demographic cell. Origin years $t = 2016$–2018 only (destination $t+1 = 2017$–2019). Δ price paid: for stayers, the change in their own paid premium, for switchers, the destination price at $t+1$ minus the origin plan's price at $t+1$ (what they would have paid staying, observed from the origin plan's stayers in the mover's age band × sex, or the plan's base growth where that cell is empty). Δ claims cost: change in insurers' costs, Δ margin uses the raw own-price change. Exit: no panel row at $t+1$, the changes are undefined there by construction. Claims at t: the group's mean insurers costs before the transition.

Table A8: The recapture gain against the predicted-stay imputation

Origin tercile	N	Realized margin at destination plan	Predicted margin if stayed	True gain	Naive gain	Share recovered
T1 (cheap)	31,332	6.46	9.22	-2.76	-9.18	0.700
		(0.37)	(0.50)	(0.57)		
T2	40,533	10.27	16.97	-6.70	-6.45	0.605
		(0.57)	(0.64)	(0.80)		
T3 (expensive)	41,158	13.19	32.54	-19.35	-3.72	0.405
		(0.61)	(1.82)	(1.83)		

Notes: own-insurer switchers, origins 2016–2018, annual UF, cluster-robust standard errors (insurer × aggregate-plan) in parentheses. Predicted margin if stayer is the mover's own price and claims at t grown by the stayers' ratio-of-sums growth in their (age band × gender × year) cell, a data-based imputation (parallel-growth assumption), the only non-observed object in the table. True gains: realized minus predicted margins; naive gains: realized minus the cell's stayers' realized mean. Recovered percentage: realized/predicted-if-stayed, the share of the mover's stay-value the insurer retains by recapturing them, the alternative to recapture is losing the person, whose retained value is zero, so even the largest concession keeps 40% of the stay-value within the insurer (50% overall, weighting terciles by their totals). Price terciles are within (age band × gender × year) cells, relative prices for the person's demographics.

B Details on Demand Estimation

Aggregation of Health Insurance Plans. We only consider the health insurance plans from the insurers that are selling contracts to the public in general, in contrast to other insurers that only offer plans to specific companies. We keep six insurance companies and the plans they are selling to individuals, so we get rid of plans that may be directed to a company, that would be the same for a group of people. We aggregate health insurance plans together due to the large space of plans that are in these markets. Many plans are very similar, so we assume they are the same. For this, first we separate the plans according to the insurer that owns the plan, if the plan is larger or smaller than the median price for that year and insurer and if the plan has a better coverage in a defined network of providers. Once we have defined the groups, we use a K-means clustering algorithm for each group to classify plans depending on their distance to other plans in multiple characteristics: standardized price, standardized outpatient and inpatient coinsurance rates, standardized maximum reimbursement value (when reached, coinsurance goes down to a minimum regulated value), mean coverage rate of the plan computed from the claims data (actuarial value measure), and a dummy if there is restricted maternity coverage. To determine the number of clusters for each group we define the centroid as the groups of deciles of the mean effective coverage rate of the plan. And to determine the cluster we use the euclidean distance between the values of the variables. Finally, we create a variable for the aggregated plans that depends on the group assigned by the Kmeans and the year the plan was introduced, to distinguish new plans from old plans.

Definition of conditional expected out of pocket costs and conditional variance of out of pocket costs. To define the expected out-of-pocket (OOP) expenditures we build on Abaluck and Gruber (2016) and Atal (2019) who assume rational expectations over the expected out-of-pocket costs

that the individual will have in the following years from the start of the contract. To compute the expected out-of-pocket costs over the years, we specify a gradient-boosting[24] regression on the copayment of the individual as a function of initial out-of-pocket cost, and chronic conditions: diabetes, cancer, renal, maternity. The characteristics of the plan include year, preferred provider dummy, insurer dummies, average costs by age-gender group inpatient and outpatient multiplied by coinsurance rate. We train the model using the observed sample for the individuals and its chosen option, taking into account the average costs over the near future. Then, we predict the conditional expected out-of-pocket costs that the individual would have for different possible plans using the trained model.

For individual i facing plan j in year t, the prediction is

$$\widehat{\text{EOOP}}_{ijt} \;=\; \widehat{F}(x_{ijt}), \qquad \widehat{F} \;=\; \arg\min_{F\in\mathcal{F}} \sum_{(i,j,t)\in\mathcal{T}} \big(\bar{c}_{ijt} - F(x_{ijt})\big)^2, \tag{34}$$

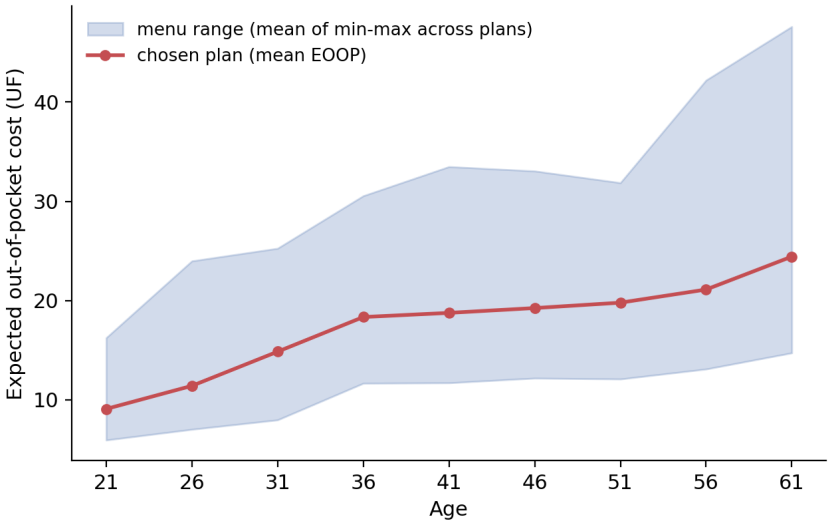
where $\bar{c}_{ijt}$ is the enrollee's mean yearly copayment over the years the contract is observed, x_{ijt} collects individual characteristics, $\mathcal{T}$ is the training sample of chosen plans, and $\mathcal{F}$ is the class of gradient-boosted tree groups described below. The variance of out-of-pocket costs, $\widehat{\text{VOOP}}_{ijt}$, is obtained by a linear projection the residuals squared on the same covariates trained on the squared out-of-fold residuals of (34). Because the model is trained on claims from the *ISAPRES* system only, it is not extrapolated to the public option: *FONASA* out-of-pocket costs are measured directly from *FONASA* claims as demographic-cell means. We evaluate the fit on a held-out sample split by individual and select the boosted model over a random forest and a linear benchmark on the R^2 and on the prediction fit against observed mean copayments across age, gender, insurer and year groups.

We find an out of sample $R^2 = 0.129$ in annual health spending, the model explains 13% of the variation. The most important variables for the change in MSE are initial cost, cancer dummy and cardiology dummy[25].

[24] Similarly to a random forest algorithm, it predicts the outcome variable by constructing trees, but it is done in a sequential way, minimizing the residuals squared of the previous tree.

[25] If we use a linear probability model to predict OOP cost we would assume no moral hazard because they spend the same in any plan (and prices are the same), like Abaluck and Gruber (2016) who assume claims would have been the same in any plan. However, using a decision tree regression we allow for behavioral responses of costs as it could be that the correlation between costs and coinsurance rate matters.

Figure 11: Constructed expected-out-of-pocket costs



Notes: Constructed expected out-of-pocket-costs from gradient-boosting regression. EOOP varies across plans for each individual. The figure shows in gray, the average of the minimum EOOP over individuals, and the average over the maximum. In red, it shows the mean of the chosen plan across individuals for each age group.

Table A9: Within-person dispersion of expected out-of-pocket cost across the choice menu, by age

Age	N	Menu min	Chosen plan	Menu max	Spread (max−min)	Spread (p90−p10)
21	2,951	5.91	9.06	16.23	10.31	3.51
26	29,162	7.00	11.39	23.98	16.98	4.53
31	46,807	7.95	14.83	25.25	17.30	6.07
36	39,366	11.64	18.33	30.55	18.91	7.24
41	30,660	11.69	18.75	33.49	21.80	7.50
46	27,247	12.16	19.23	33.05	20.89	7.77
51	22,119	12.07	19.78	31.86	19.80	7.75
56	19,182	13.07	21.10	42.21	29.13	9.34
61	15,098	14.71	24.42	47.61	32.91	11.10

Notes: Chosen plan is EOOP in UF average over chosen plans only. Menu min/max are means, across individuals in the age band, of each individual's own min/max EOOP across their choice set. Spread (p90−p10) is more robust to outliers than spread (max−min). Female, 2018 predict panel.

Outside option. The outside option is the public sector plan (called *FONASA*). The price is the 7% of income with a maximum value. We fix the generosity rate to 60% for inpatient and outpatient rates to capture a lower quality although the generosity rates might be different. We assume no preferred provider and the expected out-of-pocket costs and its variance is computed from a *FONASA* dataset that contains information on claims from *FONASA*) [26], and we can compute the mean and the variance over age and gender groups for different years. We also include individuals with high enough income from the *FONASA*) datasets that contains income, gender, and age (we only include segment D from individuals enrolled in *FONASA*). Individuals enrolled

[26]MLE is Modalidad de Libre Eleccion, which includes claims from partner providers with the public option.

in *Fonasa*) can also choose from the private sector portfolio of plans. The data from *Fonasa*) is cross section only because we cannot identify if the same person is in two different years, but we assume they never switch for the sample I have.

Choice sets. The individuals can choose a health insurance plan from a choice set that depends on the plans chosen by other people in the same gender-age group. We keep all the (aggregated) plans as defined in Aggregation of plans Section B that were bought by someone in the demographic group, and are offered at that point in time. We also include the outside option, and the non-offered plans that were chosen by the individual in the previous period. It can be seen as another option or as the actual chosen option. In that way we make sure that people that switch still have their previous option in their choice sets.

Instruments construction. We compute the potential number of people outside of Santiago that can buy a health insurance plan. For this, we take into account the number of people in each demographic group (age and gender), but also we restrict the potential market to individuals whose income allows them to pay the premium associated with each plan. In this way, we have variation across plans and demographic groups on the potential market size that comes from outside the region under study. The data comes from the CASEN survey for the year 2017. This measure reflects predetermined demographic composition and is fixed prior to the estimation period. Then we multiply this by a shift component, given by aggregate changes in medical costs over time, measured by the Health Consumer Price Index (CPI) published by the Central Bank of Chile. This national level cost shifter affects insurers costs and pricing decisions uniformly across regions. Interacting the predetermined demographic shares with the aggregate cost shifter generates time-varying, plan-specific variation in prices that operates through supply conditions and is plausibly orthogonal to unobserved demand shocks within Santiago.

$$Z_{jgt} \;=\; M^{\text{out}}_{jg,2017} \times \Delta\text{HealthCPI}_t, \tag{35}$$

Other instruments include information from the financial balance sheet of each insurer by year, including liquidity and debt, and the maximum price adjustment rate. Another instrument represents the potential total costs of the plan outside that demographic group. It is the potential total cost of the plan (sum across age groups) minus potential total cost of observed demographic group, whereas potential total cost is constructed as the average cost of the demographic group in the market by the number of individuals of that demographic group in the plan.

B.1 Estimation details.

We use a two step method to tackle concerns of endogeneity for prices due to unobserved characteristics of the health insurance plan, following standard methods as Berry, Levinsohn, and Pakes (1995) and Berry and Haile (2024). In the first step, we specify a Likelihood function that represents the probability of choosing the observed plan for each individual. The conditional choice

probability for the observed plan is a result of the discrete choice model with Type 1 Extreme Value distribution, and it depends on the parameters to be estimated and the variables of the individual. The indirect utility can be decomposed in a term which is plan and demographic group specific δ_{gjt} and an individual specific term μ_{ijt} as in equation (36). The Maximum Likelihood estimation recovers individual specific parameters Θ conditional on $\delta_{gjt}(\Theta)$. The mean taste parameter δ_{gjt} is recovered from a contraction mapping that matches the observed shares with the theoretical shares as a function of Θ. The contraction mapping for each market (demographic group - year) iterates until a fixed-point is reached within the likelihood estimation. In a second step, we regress the recovered mean utilities δ_{jgt} on multiple characteristics of the plan (coverage rate, provider fixed effects, insurer fixed effects and restriction of the plan fixed effects). We also control for the age group of the observation. We use a 2SLS specification to instrument price. Instruments for the second stage include the exposure with common aggregate shift IV. We also include information from the financial balance sheet of each insurer by year, including liquidity and debt, and the maximum price adjustment rate.

$$\begin{aligned} u_{ijt} =& \alpha(z_{it})\, p_{jg(i)t} + \beta_1 x_j + \xi_{jg(i)t} + \omega_0\, EOP_{ijt} + \omega_1\, VOP_{ijt} + \\ & \gamma_1 \mathbf{1}\{j = j_{i,t-1}\} + \gamma_2 \mathbf{1}\{f(j) = f(j_{i,t-1})\} + \varepsilon_{ijt} \\ u_{ijt} =& \delta_{g(i)jt} + \mu_{ijt}(z_i, j_{i,t-1}) + \varepsilon_{ijt} \end{aligned} \tag{36}$$

$$\delta_{jgt} = \beta x_j + \alpha_0 p_{jgt} + \xi_{jgt} \tag{37}$$

$$\beta x_j = \beta^1\, coverage_j + \sum_{k=1}^{K} \beta_k^2\, \mathbf{1}\{provider_k \in j\} + \lambda_{\text{ins}(j)} + \lambda_{\text{rest}(j)}.$$

With $\lambda_{\text{ins}(j)}$ insurance fixed effects and $\lambda_{\text{rest}(j)}$ plan's restrictions fixed effects. $coverage_j$ corresponds to the average rate of reimbursement from the insurer in plan j.

More details on the estimation procedure. Due to the high state-dependence on this industry, the contraction mapping to recover mean utilities needs to control for inertia because the observed shares are a combination of active choosers and non-active choosers, what can bias upward the mean utility if these are not taken into account. A joint estimation of δ_{gjt} and the inertia parameters is possible, but we take a step further to avoid contamination bias due to mispecification in the inertia term. If the inertia term in the utility is not correctly specified, it directly affects the contraction mapping. To avoid this, we are able to identify the mean taste parameters separately, by relying on individuals that enter the market, which we name newcomers. These individuals do not present any state dependence and their active choosing can correctly pin-down the mean taste utilities associated with quality only. We split individuals in newcomers, incumbents and legacy. Newcomers individuals are those that appear for the first time in the sample observed. Incumbents are those who are already in the system and are choosing an offered-plan at the period under analysis, while legacy individuals represent people that have an old plan which is not being offered anymore. This is important for the demand estimation because some closed plans, which are not offered anymore, have people buying from those, but the number of individuals that have these kinds of plans in their choice set is very thin. When specifying a contraction mapping these

plans are not part of the market as they are not in the choice of most of individuals, generating very large observed shares which produce biased mean taste parameters. In the likelihood function, we do include incumbents choosing offered plans as well, and this is the sample that identifies inertia parameters of the plan and the insurer. Then, using the consistent estimated parameters of the maximum likelihood step, we compute again with a contraction mapping the mean taste parameters using the total observed shares of the offered plans. Then, these recovered taste parameters are the ones regressed over characteristics of the plan in the second step. These types of procedures in two stages to identify state-dependence from other parameters are close to Aguirregabiria and Mira (2007), who solve a dynamic single-agent problem and in a first stage compute non-parametrically conditional choice probabilities which capture the state-dependence plus preferences, and conditional on these objects, then estimate the second stage parameters (Hotz and Miller (1993)). For efficiency, they re-compute the structural CCPs and iterate the two-steps, which is the nested pseudo-likelihood algorithm. In our case, the amount of plans in the market makes the non-parametric step hard to justify, and the newcomers specification results in precisely estimated free inertia mean utilities. Then, the efficiency is recovered by a second contraction mapping pooling individuals, conditional on the consistent estimates, which would correspond to re-iteration part of NPL after computing structural CCPs. Another paper identifying state-dependence is Mackay and Remer (2026) who do not have individual data and rely on specifying unobserved consumers types such that one has state-dependence and the other does not.

C Demand Results

Table A10: Demand Estimation in Two Steps

	Female	Male
Step 1: MLE		
Price × age (α_1)	−3.810	−4.794
	(0.097)	(0.148)
Price × income (α_2)	0.456	0.600
	(0.008)	(0.010)
EOP (ω_1)	−0.382	−0.511
	(0.029)	(0.078)
VOP (ω_2)	−0.091	−0.155
	(0.002)	(0.004)
Same plan (γ_{plan})	6.448	6.812
	(0.019)	(0.014)
Same insurer (γ_{ins})	1.232	1.016
	(0.011)	(0.008)
Observations	74,802,612	134,404,335
Step 2: TSLS		
Price	0.062	−1.781
	(0.060)	(0.379)
Coverage rate	3.988	14.918
	(0.590)	(2.350)
Provider FE	Yes	Yes
Insurer FE	Yes	Yes
Age FE	Yes	Yes
Restriction FE	Yes	Yes
Year of introduction	Yes	Yes
Observations	3,327	3,420

Step 1: SE are $\hat{\sigma}_k = \sqrt{\hat{V}_{kk}}$, $\hat{V}(\hat{\theta}) = H(\hat{\theta})^{-1}\left(\sum_i s_i(\hat{\theta}) s_i(\hat{\theta})^\top\right) H(\hat{\theta})^{-1}$, $s_i(\hat{\theta}) = \nabla_\theta \log \ell_i(\hat{\theta})$ the individual score.
Step 2: Robust standard errors in parentheses.

Table A11: First Stage

	Female	Male
	Price	
CPI × N_{demo}	−0.066	−0.003
	(0.005)	(0.001)
Liquidity × AC^{out}	−2.772	−1.257
	(0.406)	(0.317)
Loss ratio	−10.508	−7.224
	(1.253)	(0.946)
Debt	0.454	0.491
	(0.119)	(0.094)
Coverage rate	5.833	4.709
	(0.266)	(0.254)
Provider FE	Yes	Yes
Insurer FE	Yes	Yes
Age FE	Yes	Yes
Restriction FE	Yes	Yes
Year	Yes	Yes
F-stat (excluded instruments)	73.87	38.18
Observations	3,327	3,420

Notes: Robust (HC1) standard errors in parentheses. F-stat tests joint significance of the excluded instruments (CPI × N_{demo}, Liquidity × AC^{out}, Loss ratio, Debt).

Table A12: Willingness to pay and switching costs (female)

	Mean (wtd.)	Median	P10	P90
Coverage rate (0–1), UF	4.478	4.698	2.569	8.130
10 pp coverage increase, % of price	24.300	10.524	4.725	23.469
Switching cost (plan + insurer), UF	8.623	9.047	4.947	15.656
Same-insurer component, UF	1.383	1.451	0.793	2.511

Notes: WTP computed as $-\beta_x/\alpha_i$ on chosen plans with $\alpha_i < -0.15$; mean weighted by sample weights. Results based on demand estimates.

D Dynamic Foundations for Static Demand Estimation

We present a dynamic choice problem for individuals in which they take into account how the plan they choose today will affect the decisions in the future. This could be the case if individuals have an expectation on their health state path through the lifecycle, and considering switching costs are high they might choose a plan with higher coverage rates today. We present the general framework and state the assumptions under which a static estimation can be consistent with taking into account the future.

The Dynamic Choice Problem

Consider a finite horizon T. Individual i in demographic group g at time t with previous plan j_{t-1} solves:

$$V_{it}(j_{t-1}) = E\left[\max_{j\in\mathcal{J}}\{v_{ijt} + \gamma\mathbf{1}[j = j_{t-1}] + \epsilon_{jt} + \beta\, V_{i,t+1}(j)\}\right] \tag{38}$$

$V_{i,T+1} \equiv 0$, and flow utility absent inertia is:

$$v_{ijt} = \delta_{jgt} + \dot{\mu}_{ijt} \tag{39}$$

with mean utility δ_{jgt} common within group g and individual-specific deviation $\dot{\mu}_{ijt}$ (to differentiate with the main text in which μ_{ijt} includes inertia terms). The state at time t is the previously chosen plan j_{t-1} only. Health status affects utility solely through expected out-of-pocket costs in v_{ijt} [27] and therefore the chosen plan fully determines next period's state: the continuation value $V_{i,t+1}(j)$ in (38) depends on the current action j only, not on the inherited state j_{t-1}. The model therefore exhibits one-period finite dependence (Arcidiacono and Miller (2011)) by construction, which is what allows for the absorption results below.

Under rational expectations over $\{v_{ij\tau}\}_{\tau=0}^{T}$ (we suppress the expectation over future states[28]), the solution is obtained by backward induction. Under additive separability of ϵ_{jt}, conditional independence assumption (Rust (1987), ϵ_{jt} is iid across alternatives and over time, independent of the state j_{t-1}, so the shock is not a payoff-relevant state variable), and ϵ_{jt} iid Type 1 Extreme Value, by the logit inclusive value formula and using γ_E as the Euler-Mascheroni constant, the value function is:

$$V_{it}(j_{t-1}) = \gamma_E + \ln\sum_{j\in\mathcal{J}}\exp\left(v_{ijt} + \gamma\mathbf{1}[j = j_{t-1}] + \beta\, V_{i,t+1}(j)\right) \tag{40}$$

With J plans and T periods, this requires $J \times T$ logsum evaluations per individual, each over J alternatives, and this for every trial parameter vector in estimation. The period-0 choice is:

$$\max_{j\in\mathcal{J}}\{v_{ij0} + \gamma\cdot\mathbf{1}[j = j_{-1}] + \beta V_{i1}(j) + \epsilon_{j0}\} \tag{41}$$

The static demand model estimates:

$$\max_{j}\left\{\tilde{\delta}_{jg0} + \tilde{\mu}_{ij0} + \hat{\gamma}\cdot\mathbf{1}[j = j_{-1}] + \epsilon_{j0}\right\} \tag{42}$$

We seek conditions under which (42) recovers the correct choice probabilities from (41).

Following the inversion theorem of Hotz and Miller (1993) under the T1EV assumption, the value function can be written with any action a available in state j_{t-1} and its conditional choice

[27] Since plan enrollment does not affect the evolution of health, and Assumptions 1-4 below imply that individuals know the path of their expected costs, health is an exogenous deterministic trajectory absorbed into the time index of flow utility rather than a state variable.

[28] Assumptions 2-3 below imply that the conditional expectation of future flow utilities coincides with a deterministic path known to the individual

probability $P_t(a \mid j_{t-1})$.

$$V_{it}(j_{t-1}) = v_{iat} + \gamma \cdot \mathbf{1}[a = j_{t-1}] + \beta V_{i,t+1}(a) + \gamma_E - \ln P_t(a \mid j_{t-1}) \tag{43}$$

Stay and switch actions. At time t, an individual in state $k \equiv j_{t-1}$ faces a binary partition of actions: stay ($a = k$) or switch ($a = j \neq k$). Using the stay action:

$$V_{it}(k) = v_{ikt} + \gamma + \beta V_{i,t+1}(k) + \gamma_E - \ln P_t(k \mid k) \tag{44}$$

Using any switch action $j \neq k$:

$$V_{it}(k) = v_{ijt} + \beta V_{i,t+1}(j) + \gamma_E - \ln P_t(j \mid k) \tag{45}$$

Both representations give the same $V_{it}(k)$, which is given by the expected maximum value of the optimization problem at t. Equating (44) and (45) gives the value function difference between being an incumbent on plan k versus being an incumbent on plan j:

$$\beta \left[V_{i,t+1}(k) - V_{i,t+1}(j)\right] = (v_{ijt} - v_{ikt}) - \gamma + \ln P_t(k \mid k) - \ln P_t(j \mid k) \tag{46}$$

The left-hand side is the continuation value difference: how much more valuable it is to be locked into plan k versus plan j going forward. The right-hand side expresses this entirely in terms of current-period flow utilities and conditional choice probabilities, without requiring the value function itself. The probability of staying is:

$$\sigma_{ikt} \equiv P_t(k \mid k) = \frac{\exp(v_{ikt} + \gamma + \beta V_{i,t+1}(k))}{\sum_j \exp(v_{ijt} + \gamma \cdot \mathbf{1}[j = k] + \beta V_{i,t+1}(j))} \tag{47}$$

The correction term $-\ln \sigma_{ikt}$ in (44) captures the option value of switching:

$$-\ln \sigma_{ikt} = \ln \left[1 + \frac{\sum_{j \neq k} \exp(v_{ijt} + \beta V_{i,t+1}(j))}{\exp(v_{ikt} + \gamma + \beta V_{i,t+1}(k))} \right] \tag{48}$$

The numerator inside the ratio is the inclusive value of all switching destinations and the denominator is the payoff of staying. When γ is large, the staying payoff dominates, the ratio is small, and $-\ln \sigma_{ikt} \approx 0$.

Using the Hotz-Miller equation with the stay action at every period, we can solve $V_{i,t+1}(k)$

forward:

$$\begin{aligned} V_{i,t+1}(k) &= v_{ik,t+1} + \gamma + \beta V_{i,t+2}(k) + \gamma_E - \ln \sigma_{ik,t+1} \\ &= v_{ik,t+1} + \gamma + \gamma_E - \ln \sigma_{ik,t+1} + \beta \left[v_{ik,t+2} + \gamma + \gamma_E - \ln \sigma_{ik,t+2} + \beta V_{i,t+3}(k) \right] \\ &= \sum_{\tau=t+1}^{T} \beta^{\tau-t-1} \left(v_{ik\tau} + \gamma + \gamma_E - \ln \sigma_{ik\tau} \right) \end{aligned} \tag{49}$$

using $V_{i,T+1} = 0$. The switching option enters at every period through $-\ln \sigma_{ik\tau}$: at each future τ, the individual might switch away from plan k due to a sufficiently favorable taste shock on an alternative. The term $-\ln \sigma_{ik\tau}$ captures the expected value of this possibility.

The continuation value of choosing plan j at $t = 0$ is therefore:

$$CV_i(j, 0) \equiv \beta V_{i,1}(j) = \sum_{\tau=1}^{T} \beta^{\tau} \left(v_{ij\tau} + \gamma + \gamma_E - \ln \sigma_{ij\tau} \right) \tag{50}$$

Assumption 1 (Stable choice set) *Individuals expect the set of available plans $\mathcal{J}$ to remain unchanged over the horizon T. Newly introduced plans are unanticipated additions to the choice set.*

Assumption 2 (Group-level trajectory) *The group-level component of future plan-specific utility follows a trajectory that is common knowledge within group g:*

$$E_i[\delta_{jg\tau} + \bar{\mu}_{gj\tau}] = \delta_{jg\tau} + \bar{\mu}_{gj\tau} \qquad \forall\, i \in g,\ \forall\, j, \tau \tag{51}$$

where $\bar{\mu}_{gj\tau} = E[\dot{\mu}_{ij\tau} \mid i \in g]$ is the group mean of individual-specific utility. This is implied by the construction of EOOP, which is predicted from group-level observables (age, gender) and provides a common expected trajectory of out-of-pocket costs within groups.

Assumption 3 (Persistent individual deviations) *Individual-specific deviations from the group mean are approximately constant over the horizon:*

$$\dot{\mu}_{ij\tau} - \bar{\mu}_{gj\tau} \approx \dot{\mu}_{ij0} - \bar{\mu}_{gj0} \qquad \forall\, \tau = 1, \ldots, T \tag{52}$$

The primary source of within-group variation in $\dot{\mu}$ is expected out-of-pocket costs, which differ across individuals mainly through persistent health conditions (chronic diseases such as diabetes, cancer, renal conditions). These conditions change slowly relative to the contract horizon, so an individual's deviation from their group mean is approximately stable.

Assumption 4 (Observed heterogeneity) *All persistent individual-level determinants of utility are captured by observables: demographics z_i and predicted out-of-pocket costs. There is no persistent unobserved taste heterogeneity (ξ_{ij} or random coefficients). The only unobservable is the transitory shock ϵ_{jt}.*

This assumption separates state dependence from unobserved heterogeneity: conditional on observables, the previous plan depends only on past realizations of ϵ, which are independent of ϵ_0, so the inertia state is exogenous in the period-0 choice. Identification of γ then rests on switching responses to changes in the choice environment, in particular the entry of new plans. Using assumptions 2–4), we apply the decomposition $v_{ij\tau} = \delta_{jg\tau} + \bar{\mu}_{gj\tau} + (\dot{\mu}_{ij0} - \bar{\mu}_{gj0})$ to the exact continuation value (50):

$$\begin{aligned} CV_i(j,0) &= \sum_{\tau=1}^{T} \beta^\tau \left(\delta_{jg\tau} + \bar{\mu}_{gj\tau} + (\dot{\mu}_{ij0} - \bar{\mu}_{gj0}) + \gamma + \gamma_E - \ln \sigma_{ij\tau}\right) \\ &= \underbrace{\sum_{\tau=1}^{T} \beta^\tau \left(\delta_{jg\tau} + \bar{\mu}_{gj\tau} + \gamma + \gamma_E\right)}_{\text{group-specific, varies by } (j,g) \text{ only}} + \frac{\beta(1-\beta^T)}{1-\beta} \cdot (\dot{\mu}_{ij0} - \bar{\mu}_{gj0}) \; - \underbrace{\sum_{\tau=1}^{T} \beta^\tau \ln \sigma_{ij\tau}}_{\text{option value}} \end{aligned} \tag{53}$$

The continuation value has three components. The first is the group-level present discounted value of staying on plan j, including the group trajectory of quality, EOOP, and the switching cost bonus collected each period, which varies by (j, g) only. The second is the individual deviation from the group mean, proportional to the current-period deviation with factor $\frac{\beta(1-\beta^T)}{1-\beta}$. The third is the cumulative option value of potentially switching in future periods, this is the only term whose absorption into the static model requires further assumptions on $\sigma_{ij\tau}$.

Next, we provide different assumptions for the staying probabilities that can explain the absorption of option value terms. Denote $c \equiv \frac{\beta(1-\beta^T)}{1-\beta}$. The total utility of choosing plan j at $t = 0$ is:

$$\begin{aligned} \tilde{u}_{ij0} &= v_{ij0} + \gamma \cdot \mathbf{1}[j = j_{-1}] + CV_i(j,0) + \epsilon_{j0} \\ &= \underbrace{\delta_{jg0} + \sum_{\tau=1}^{T} \beta^\tau \left(\delta_{jg\tau} + \bar{\mu}_{gj\tau} + \gamma + \gamma_E\right) - c \cdot \bar{\mu}_{gj0}}_{\text{group-level, varies by } (j,g) \text{ only}} \\ &\quad + \; (1+c) \cdot \dot{\mu}_{ij0} + \gamma \cdot \mathbf{1}[j = j_{-1}] - \sum_{\tau=1}^{T} \beta^\tau \ln \sigma_{ij\tau} + \epsilon_{j0} \end{aligned} \tag{54}$$

Case 1: Staying beliefs. If individuals evaluate each plan as if remaining enrolled for the rest of the horizon, ignoring the possibility of future switching, they effectively set $\ln \sigma_{ij\tau} = 0$ for all j, τ. This is a behavioral assumption, not a parameter restriction: with full-support taste shocks $\sigma_{ij\tau} < 1$ in the data-generating process, but the individual plans as if they will stay. The option value vanishes from the subjective evaluation and total utility reduces to:

$$\tilde{u}_{ij0} = \underbrace{\delta_{jg0} + \sum_{\tau=1}^{T} \beta^\tau \left(\delta_{jg\tau} + \bar{\mu}_{gj\tau} + \gamma + \gamma_E\right) - c \cdot \bar{\mu}_{gj0}}_{\tilde{\delta}_{jg0}} + (1+c) \cdot \dot{\mu}_{ij0} + \gamma \cdot \mathbf{1}[j = j_{-1}] + \epsilon_{j0} \tag{55}$$

Under Assumptions 1–3 and staying beliefs, the dynamic choice (41) is exactly equivalent to the static choice (42): the group-level continuation value is absorbed into $\tilde{\delta}_{jg0}$, the individual-level parameters are scaled to $\tilde{\theta} = (1+c)\theta$, and $\hat{\gamma} = \gamma$. Future switching cost bonuses are collected unconditionally under staying beliefs, contributing the plan-invariant constant $c\gamma$ to $\tilde{\delta}_{jg0}$, the switching cost at $t = 0$ enters only through $\gamma \cdot \mathbf{1}[j = j_{-1}]$, unscaled.

Case 2: Forward-looking individuals with large switching costs. If individuals do anticipate future switching, the option value $-\sum_\tau \beta^\tau \ln \sigma_{ij\tau}$ enters utility, and the question is whether it varies across plans. Exact uniformity of $\sigma_{ij\tau}$ across j would require all plans to have identical total value, a degenerate case, but when γ is large, the staying probability is dominated by the switching cost rather than plan quality, and cross-plan differences in $\ln \sigma$ are of order $e^{-\gamma}$:

$$\sum_{\tau=1}^{T} \beta^\tau \left(\ln \sigma_{ij\tau} - \ln \sigma_{ij'\tau} \right) = O(e^{-\gamma}) \tag{56}$$

Since the logit depends only on utility differences, the option value cancels up to this error, and the conclusions of Case 1 hold with approximation error vanishing exponentially in γ. To the extent that residual cross-plan differences in staying probabilities are driven by group-level quality differences rather than individual deviations, the remaining error is group-specific and absorbed into $\tilde{\delta}_{jg0}$. The quality of the approximation can be assessed empirically from the cross-plan dispersion of staying rates.

What the static model recovers. Under both cases, the static estimates map to the dynamic primitives as follows. The mean utility $\tilde{\delta}_{jg0}$ is the present discounted value of group-level plan utility, absorbed by the contraction mapping. The inertia coefficient recovers the per-period switching cost, $\hat{\gamma} = \gamma$, without scaling. The individual-level parameters are inflated by the present-value factor, $\tilde{\theta} = (1+c)\theta$, this scaling cancels in choice probability ratios and is immaterial for elasticities and the supply-side analysis. It affects two objects only: welfare calculations requiring per-period willingness to pay, and switching costs expressed in currency units, $\hat{\gamma}/\tilde{\alpha}$, which understate the per-period cost by the factor $(1+c)$.

Implications for identification. Newcomers face no inertia, so the contraction mapping on their shares identifies $\tilde{\delta}_{jg0}$ uncontaminated by switching costs. Newly introduced plans are unanticipated under Assumption 1: switches to them respond to realized innovations in the choice set, variation orthogonal to the inertia structure, and this identifies γ under Assumption 4. Finally, the static model does not recover quality dynamics: future quality not captured by current observables enters ξ_{jg0} in the decomposition $\tilde{\delta}_{jg0} = x_j'\beta + \alpha p_{jg0} + \xi_{jg0}$, so $\hat{\alpha}$ reflects a present-value rather than per-period response to price.

E Supply model: derivation of the first-order condition

E.1 Bellman representation and envelope argument

The insurer's payoff-relevant state at t is $(\mathbf{s}_{t-1}, \mathcal{J}_t)$, where $\mathbf{s}_{t-1}$ represents the distribution of consumers over the plans they held, which carries enrollment persistence, and $\mathcal{J}_t$ is the realized menu, including the attributes and prices of all active plans. Consumer demographics, health and costs evolve exogenously and enter through the time index. Per-period profits are:

$$\pi_t = \sum_{m \in \mathcal{J}_{f,t}} \int_I (p_m - c_{im,t})\, s_{im,t}\big(\mathbf{p}_{\mathcal{J}_t}, \mathbf{s}_{t-1}\big)\, di, \qquad \mathbf{s}_t = s_t\big(\mathbf{p}_{\mathcal{J}_t}, \mathbf{s}_{t-1}\big), \tag{57}$$

and the menu evolves stochastically given the current state, through rival entry and exit and the firm's own portfolio actions. Since the continuation from $t+1$ depends on the past only through $(\mathbf{s}_{t-1}, \mathcal{J}_t)$ and the prices already set, the value of the firm pricing its newly introduced plans at t satisfies

$$V_t(\mathbf{s}_{t-1}, \mathcal{J}_t) = \max_{\{p_l :\, t_l = t\}} \Big\{ \pi_t + \beta\, \mathbb{E}\big[V_{t+1}(\mathbf{s}_t, \mathcal{J}_{t+1}) \,\big|\, \mathbf{s}_{t-1}, \mathcal{J}_t \big] \Big\}. \tag{58}$$

FOC. For plan j with $t_j = t$,

$$0 = \frac{\partial \pi_t}{\partial p_j} + \beta\, \frac{d}{dp_j}\, \mathbb{E}\big[V_{t+1}(\mathbf{s}_t, \mathcal{J}_{t+1}) \,\big|\, \mathbf{s}_{t-1}, \mathcal{J}_t \big], \qquad \frac{\partial \pi_t}{\partial p_j} = \int_I s_{ij,t}\, di + \sum_{m \in \mathcal{J}_{f,t}} \int_I (p_m - c_{im,t}) \frac{\partial s_{im,t}}{\partial p_j}\, di, \tag{59}$$

The expectation is over the next menu $\mathcal{J}_{t+1}$, which results from own and rival entry and exit given the current state, and over the exogenous consumer processes. The level term comes from $\partial\big[(p_j - c_{ij,t}) s_{ij,t}\big]/\partial p_j = s_{ij,t} + (p_j - c_{ij,t})\, \partial s_{ij,t}/\partial p_j$. The total derivative of the continuation value has three channels:

$$\frac{d}{dp_j}\, \mathbb{E}\big[V_{t+1}\big] = \underbrace{\mathbb{E}\left[\frac{\partial V_{t+1}}{\partial \mathbf{s}_t} \cdot \frac{\partial \mathbf{s}_t}{\partial p_j} \right]}_{\text{(i) enrollment state}} + \underbrace{\sum_{l:\, t_l > t} \mathbb{E}\left[\frac{\partial V_{t+1}}{\partial p_l} \cdot \frac{dp_l^*}{dp_j} \right]}_{\text{(ii) own future actions}} + \underbrace{\frac{\partial}{\partial p_j}\, \mathbb{E}_{\mathcal{J}_{t+1}}\big[\ \cdot\ \big]}_{\text{(iii) menu process}}, \tag{60}$$

where in channel (iii) the derivative acts on the distribution of $\mathcal{J}_{t+1}$, holding V_{t+1} fixed.

Assumption 5 (Private information shock and optimal own portfolio) *Future prices of own plans are interior optima of the continuation problem. Each discrete portfolio action (introduction, closure) is taken after the firm observes a private payoff shock ϵ_{ft}, iid over time and across firms, with a continuous, full-support distribution (Doraszelski and Satterthwaite (2010)), hence a marginal change in p_j has no first-order effect on the continuation value through own actions, and $E[V_{t+1}]$ is differentiable in p_j.*

Assumption 6 (Atomicity) *Conditional on the aggregate state , the law of motion of the menu: own and rival entry and exit counts and entrant attributes is invariant to p_j.*

Channel (ii) vanishes by the envelope theorem: the future actions p_l^* are functions of a state that p_j shifts, so $dp_l^*/dp_j \neq 0$ in general, but each such term is multiplied by $\partial V_{t+1}/\partial p_l$, which is zero at

an interior optimum of the continuation problem under Assumption 5. For the discrete portfolio decision of introduction and entry, Assumption 5 implies a cutoff rule. The threshold depends on the difference in continuation values with and without introduction. A marginal change in p_j changes this threshold and therefore changes the introduction decision only for values of ϵ_{ft} close to the cutoff. At the cutoff, the firm is indifferent between introducing and not introducing, so changing the decision has no first-order effect on value. Since the distribution of ϵ_{ft} is continuous, the probability of being in the region where the decision changes goes to zero with the change in p_j. The expected continuation value is therefore differentiable in p_j, and the change in the firm's own introduction decision does not contribute to $dE[V_{t+1}]/dp_j$ at first order. The shock therefore smooths the discrete introduction decision, and the envelope theorem applies. This is related to the role of private information shocks in Doraszelski and Satterthwaite (2010). The resulting introduction probability is a function of the threshold given by the shock, which gives the conditional choice probability representation of Hotz and Miller (1993). In the estimation, plan introduction enters the profiles of Appendix F through the coarse state. The effect of p_j on future outcomes is still present through channel (i). Only the first-order effect that works through changes in the firm's optimally chosen introduction decision disappears. What does not vanish is the direct demand effect of p_j on the shares of future plans, the enrollees displaced from j who land in the firm's later plans, which is part of channel (i) and remains as the recapture term. Channel (iii) vanishes by atomicity: the law of motion of the menu is invariant to the price of one plan.

Recursion. Applying (60) at $t+1, t+2, \ldots$ and expanding channel (i) through $\mathbf{s}_{t+\tau} = s_{t+\tau}(\cdot)$,

$$\beta \frac{d}{dp_j} \mathbb{E}\big[V_{t+1}\big] = \sum_{\tau \geq 1} \beta^\tau \, \mathbb{E}\left[\int_I s_{ij,t+\tau} \, di + \sum_{m \in \mathcal{J}_{f,t+\tau}} \int_I (p_m - c_{im,t+\tau}) \frac{ds_{im,t+\tau}}{dp_j} \, di \;\middle|\; \mathbf{s}_{t-1}, \mathcal{J}_t \right], \quad (61)$$

where at each τ the sum runs over all of the firm's plans active at $t+\tau$, including those introduced after t, whose prices enter at their optima, and the derivatives are total, the direct effect plus the propagation through the enrollment state. Substituting (61) into (59) gives equations (8) and (9) of the text.

Remark 1 (Set-once versus repriced) *If instead the firm re-optimized p_{jt} every period, the envelope theorem would apply to p_j itself at every future date, eliminating all continuation terms except the value of the enrollment state:*

$$0 = \int_{\mathcal{I}} s_{ij,t} \, di + \sum_{m \in \mathcal{J}_{f,t}} \int_{\mathcal{I}} (p_m - c_{im,t}) \frac{\partial s_{im,t}}{\partial p_{jt}} \, di + \beta \, \frac{\partial V_{t+1}}{\partial \mathbf{s}_t} \cdot \frac{\partial \mathbf{s}_t}{\partial p_{jt}},$$

the static Nash-Bertrand condition plus an investment motive for building share, as in dynamic-pricing environments such as Fleitas (2017). Set-once pricing removes this margin: the single price must balance the whole path of the plan at introduction, which is why the condition keeps the full sequence of future blocks.

E.2 The recursive derivative under persistence

At the individual level, past enrollment affects today's choice through the switching-cost term of the demand model and through the persistence of the tastes and health states of those who selected the plan. Enrollment therefore obeys the law of total probability over the previous plan, $s_{ij}(\tau) = \sum_m \Pr_{i,t+\tau}(j \mid m)\, s_{im}(\tau-1)$, where $\Pr_{i,t+\tau}(j \mid m)$ is the demand model's probability that i chooses j at plan age τ having held m the year before. The plan's own price and characteristics enter its numerator, while the state, in particular entry, lowers the stay probability $\Pr_{i,t+\tau}(j \mid j)$ through its denominator. Total differentiation gives the recursion reported in the text,

$$\frac{ds_{ij}(\tau)}{dp_j} = \sum_m \Pr_{i,t+\tau}(j \mid m)\, \frac{ds_{im}(\tau-1)}{dp_j} + \sum_m s_{im}(\tau-1)\, \frac{\partial \Pr_{i,t+\tau}(j \mid m)}{\partial p_j}, \tag{62}$$

where the first block is the propagation of all earlier responses, and the second is the direct effect at age τ holding last year's enrollment fixed, the only block a model without the dynamic link would have.

Cost-weighting. Weighting (62) by individual costs yields the response of the plan's total cost to its own price, $G_j(\tau) = d\,\mathrm{TC}_j(\tau)/dp_j$ with $\mathrm{TC}_j(\tau) = \int c_{ij}(\tau)\, s_{ij}(\tau)\, di$:

$$G_j(\tau) = \underbrace{\int c_{ij}(\tau) \sum_m \Pr_{i,t+\tau}(j \mid m)\, \frac{ds_{im}(\tau-1)}{dp_j}\, di}_{\text{(I) inherited}} + \underbrace{\int c_{ij}(\tau)\, s_{ij}(\tau-1)\, \frac{\partial \Pr_{i,t+\tau}(j \mid j)}{\partial p_j}\, di}_{\text{(II) leaving margin at age } \tau} \tag{63}$$

$$+ \underbrace{\int c_{ij}(\tau) \sum_{m \neq j} s_{im}(\tau-1)\, \frac{\partial \Pr_{i,t+\tau}(j \mid m)}{\partial p_j}\, di}_{\text{(III) joining margins}} \tag{64}$$

Term (I) does not appear in a static condition. The enrollees whom the price made marginal in earlier years are still in the pool at age τ, one year older and with cost $c_{ij}(\tau) > c_{ij}(\tau-1)$. Term (II) is static marginal selection: the covariance between cost and this year's response of the decision to leave, which is the difference between the cost of the enrollees the price moves and the average cost of the pool. Term (III) is selection into the plan through its new enrollees. The change in the level of costs, the remaining pool becoming sicker as the healthy leave at a fixed price, is a separate object, which enters through the cost profiles of Appendix F.1. Discounting over the plan's life, $d\,\mathrm{TC}_j/dp_j = \sum_{\tau \geq 0} \beta^\tau G_j(\tau)$, which is the object G_j of the approximated condition (17). Subsection E.3 builds the same derivative from two periods, term by term, which shows where the state dependence enters.

E.3 The transmission of the price response

This subsection follows the response at introduction into the following years, to show where state dependence enters the recursion (62). Fix individual i and plan j, introduced at $\tau = 0$ with price p_j paid in every period. Each probability $\Pr_{i,t+\tau}(j \mid m)$ is computed on the choice set at age τ, $J_{i,t+\tau}$, which differs from the one at introduction because plans enter and exit.

Two periods. Differentiating (10) at $\tau = 1$ by the product rule,

$$\frac{ds_{ij}(1)}{dp_j} = \underbrace{\sum_m \Pr_{i,t+1}(j \mid m)\, \frac{ds_{im}(0)}{dp_j}}_{\text{effect at introduction, propagated}} + \underbrace{\sum_m s_{im}(0)\, \frac{\partial \Pr_{i,t+1}(j \mid m)}{\partial p_j}}_{\text{new margins at age 1}}. \tag{65}$$

The propagated term runs over all the responses at introduction: the people the price at introduction pushed out of j did not vanish. They landed in specific alternatives, and each destination has its own probability of choosing j at age 1. Because choice probabilities over the full menu sum to one, $\sum_m ds_{im}(0)/dp_j = 0$, the displaced mass equals the lost mass, and the propagated term becomes

$$\Big[\underbrace{\Pr_{i,t+1}(j \mid j)}_{\text{stay probability}} - \sum_{m \neq j} w_{im} \underbrace{\Pr_{i,t+1}(j \mid m)}_{\text{return from } m} \Big] \cdot \frac{ds_{ij}(0)}{dp_j}, \qquad w_{im} = \frac{-\,ds_{im}(0)/dp_j}{ds_{ij}(0)/dp_j}, \tag{66}$$

where the weights $w_{im} \geq 0$ sum to one and are the destination shares of the displaced: the fraction of the enrollment the price moved that landed in m. The bracket is the transmission of a choice of the plan at age 0 into enrollment at age 1: the stay probability net of the probability that the same person would have come back anyway from wherever they went. Without state dependence and without persistent selection, $\Pr_{i,t+1}(j \mid j) = \Pr_{i,t+1}(j \mid m)$ for all m, the bracket is zero, and the price at introduction has no effect on enrollment at age 1. This is the exact form of the collapse to static Bertrand. The weights also show where recapture comes from: displaced enrollees who land in the firm's own plans (high w_{im} for $m \in \mathcal{J}_f$) remain in the firm's portfolio, so the same objects that transmit the derivative also affect the cross block (C).

The weights are diversion ratios. The w_{im} are the diversion ratios of Conlon and Mortimer (2021), computed individual by individual, with the outside option included. In a static model, diversion allocates today's lost demand and weights the margin recaptured in the cross block (C). Here it also determines the plan the displaced hold when they arrive at age 1, and state dependence makes that plan matter, so diversion has dynamic consequences. At later ages the weights are destination shares of all the mass displaced up to that age, so only the weights at introduction are those of Conlon and Mortimer (2021).

Later ages. At age 2 the same two steps apply to $s_{ij}(2)$. The margin created at introduction reaches age 2 multiplied by two transmissions of the form (66), the leaving and joining margins created at age 1 reach it multiplied by one, and a new pair of margins appears at age 2. In general, each margin created at an earlier age reaches the current age multiplied by the transmissions of all the years in between. Weighted by costs and summed over the plan's life, this gives (64): the transmission is the propagation block of (62) for the plan's own enrollment. When the insurer sets the price, it does not know the choice probabilities on future menus. The expectation in the first-order condition (8) is taken here, and under Assumption 7 the estimated profiles provide it.

F Supply approximation and estimation

To take the first-order condition derived in Appendix E to the data, we add two assumptions.

Assumption 7 (Coarse state and consistent beliefs) *The insurer's expectation of the future terms in* (8) *conditions only on a low-dimensional coarse state $\mathcal{S}_t$, which describes how the industry moves, together with the plan's own enrollment and its fixed characteristics. $\mathcal{S}_t$ is a known coarsening of the full state $(\mathcal{J}_t, \mathbf{s}_{t-1})$. It evolves according to a Markov transition law $\Gamma(\mathcal{S}' \mid \mathcal{S})$, which is the same for every plan and invariant to the price of a single plan. Future rival identities and attributes, and the identities of the insurer's own future plans, are not tracked: they enter only through $\mathcal{S}_t$. The insurer's beliefs about the evolution of $\mathcal{S}_t$ and of its plan's enrollment coincide with the true laws of motion. For its own portfolio, plan introduction and closure follow the cutoff policy of Assumption 5, so the envelope argument applies along the forecast path.*

Assumption 8 (Far-future regularity) *(i) Cross-plan price effects attenuate geometrically and are estimated from the panel and the model's impulse responses, so the truncation error beyond $\bar{\tau}$ is bounded. (ii) Plan-age profiles are stationary within cohort: the far future extrapolates the recent-cohort lifecycle.*

F.1 The laws of motion

This section presents the laws of motion behind the approximated condition (17). Each law of motion holds exactly for the plan's true enrollment, response to the price and costs. The approximation comes only from replacing their inputs with the estimated profiles. As in Section 5.3.1, τ is the age of the plan, and $\Pr_{i,t+\tau}(j \mid m)$ is the demand model's probability that i chooses j at age τ having held m the year before.

The enrollment of the plan follows the law of total probability over where each enrollee was the year before:

$$S_j(\tau) = r_j(\tau)\, S_j(\tau-1) + n_j(\tau), \qquad S_j(-1) = 0, \tag{67}$$

where $r_j(\tau) = \int s_{ij}(\tau-1)\Pr_{i,t+\tau}(j \mid j)\, di \,/\, S_j(\tau-1)$ is the probability of staying in the plan, and $n_j(\tau) = n_j^{\text{new}}(\tau) + n_j^{\text{rival}}(\tau) + n_j^{\text{own}}(\tau)$ are the new enrollees of the plan: enrollees new to the private system, enrollees coming from rival insurers, and enrollees coming from the insurer's other plans. The response of enrollment to the price follows

$$\Omega_j(\tau) = T_j(\tau)\,\Omega_j(\tau-1) + d_j(\tau), \qquad d_j(\tau) = \bar{\alpha}_j(\tau)\Big[n_j(\tau) + r_j(\tau)\, S_j(\tau-1)\big(1-\rho_j(\tau)\, r_j(\tau)\big)\Big]. \tag{68}$$

The first term carries the accumulated response forward. The transmission

$$T_j(\tau) \equiv \frac{\int \sum_m \Pr_{i,t+\tau}(j \mid m)\, \dfrac{ds_{im}(\tau-1)}{dp_j}\, di}{\int \dfrac{ds_{ij}(\tau-1)}{dp_j}\, di} \tag{69}$$

is the share of last year's response that is still in the plan this year, net of those the price displaced

who come back to the plan anyway. It is smaller than the stay probability $r_j(\tau)$ because the response is carried by the enrollees on the margin, who are more mobile than the average enrollee. The second term, $d_j(\tau)$, is the response of this year's choices. Every new enrollee chooses the plan actively, including those coming from the insurer's other plans, and responds to the price. Among last year's enrollees, only those close to leaving respond, and their share is defined by

$$1 - \rho_j(\tau)\, r_j(\tau) \equiv \frac{\int s_{ij}(\tau-1)\, \Pr_{i,t+\tau}(j \mid j)\big(1 - \Pr_{i,t+\tau}(j \mid j)\big)\, di}{\int s_{ij}(\tau-1)\, \Pr_{i,t+\tau}(j \mid j)\, di}. \tag{70}$$

This gives $\rho_j(\tau) = \mathbb{E}[P^2]/\mathbb{E}[P]^2 \geq 1$ by Jensen's inequality, where $P = \Pr_{i,t+\tau}(j \mid j)$ and the expectation is over last year's enrollees: the stay probabilities of the marginal enrollees are lower than the average one. $\bar{\alpha}_j(\tau) < 0$ is the average of the price coefficient $\alpha(z_i)$ times the rating weight of each new enrollee and each of last year's enrollees, weighted by their contribution to the response. It includes the rating weight because Ω_j is a response to the base price. The cost saved at the margin is

$$G_j(\tau) = \big(\bar{c}_j(\tau) + \Delta_j^{sw}(\tau)\big)\, \Omega_j(\tau), \qquad \Delta_j^{sw}(\tau) \equiv \frac{G_j(\tau)}{\Omega_j(\tau)} - \bar{c}_j(\tau), \tag{71}$$

where $\bar{c}_j(\tau) = \int c_i\, s_{ij}(\tau)\, di / S_j(\tau)$ is the average cost of the plan's enrollees and $\Delta_j^{sw}(\tau)$ is the difference between the cost of the enrollees on the margin and the average. Finally, the recaptured margin is approximated by

$$\hat{X}_j(\tau) = \overline{(p-c)}_j^{rec}(\tau)\, \Psi_j(\tau), \qquad \Psi_j(\tau) = \hat{r}(\tau)\, \Psi_j(\tau-1) + \hat{\pi}_j^{own}(\tau)\, M_j^{own}(\tau) + \hat{\pi}^{sh}(\tau)\, M_j^{ext}(\tau), \tag{72}$$

where $\Psi_j(\tau)$ is the stock of enrollees the price has pushed into the insurer's other plans, and $\overline{(p-c)}_j^{rec}(\tau)$ is the annual margin the insurer earns on each of them. The stock grows with the enrollment the price displaces at age τ. Displaced enrollees from last year, together with the accumulated response carried out by leavers, and displaced new enrollees from the insurer's own plans, $M_j^{own}(\tau)$, land in the insurer's plans at the share $\hat{\pi}_j^{own}(\tau)$ of the plan's leavers who stay with the insurer. Displaced new enrollees from outside the system and from rivals, $M_j^{ext}(\tau)$, land there at the share $\hat{\pi}^{sh}(\tau)$ of their alternatives that belong to the insurer. Recaptured enrollees survive in their new plans at the pooled retention $\hat{r}(\tau)$. This is an approximation: their survival is governed by the retention of the plans they move to, which we do not track.

We replace r_j, n_j, $\bar{c}_j$, Δ_j^{sw}, T_j, ρ_j, $\bar{\alpha}_j$ and the other inputs by the estimates of the transitions in Table A13, and run the recursions from each plan's observed enrollment at introduction, $S_j(0)$. The paths S_j, Ω_j, G_j and $\hat{X}_j$ are therefore specific to each plan, even where the profiles that propagate them are pooled.

F.2 Construction of the level profiles

Each level profile describes how a quantity, such as the stay probability or the average cost, changes over the life of a plan. The coarse state $\mathcal{S}_t$ of Assumption 7 collects the variables that

Table A13: Inputs of the approximated first-order condition

Input	Built from	Plan-specific?
Level profiles, from the plan-year panel		
$\hat{r}_j(\tau)$, stay probability	retention of last year's pool	plan level
$\hat{n}_j(\tau)$, new enrollees by origin	Poisson counts of new enrollees	plan level
$\hat{\bar{c}}_j(\tau)$, average cost	average cost of the pool, in logs	plan level
$\widehat{rw}_j(\tau)$, mean rating weight	mean rating weight of the pool	plan level
$\hat{\pi}_j^{own}(\tau)$, leavers kept by the insurer	own-insurer share of leavers	plan level
$\overline{(p-c)}_j^{rec}(\tau)$, margin on recaptured	price minus cost at destination	plan level
$\hat{\Delta}^{sw}(\tau)$, marginal minus average cost	mover cost regression, by plan-age bin	pooled
$\hat{r}(\tau)$, survival of recaptured	pooled stay probability	pooled
λ_f, cost loading	accounting loss ratios	insurer
Responses to price, from the demand model		
$\hat{\bar{\alpha}}_j$, price sensitivity	demand model, composition at introduction	plan composition
$\hat{T}_j(\tau) = \hat{\varsigma}(\tau)\,\hat{r}_j(\tau)$, transmission	demand model, 2018 and 2019 transitions	through $\hat{r}_j$
$\hat{\rho}(\tau)$, mobility of marginal enrollees	demand model, 2018 and 2019 transitions	pooled
$\hat{\pi}^{sh}(\tau)$, insurer share of alternatives	demand model, 2018 and 2019	pooled
$\hat{\rho}_j^{rw}(\tau)$, rating weight of responders	demand model, scaled by $\widehat{rw}_j(\tau)$	through $\widehat{rw}_j$

Notes: Plan level: the profile of a typical plan moved to the plan's own observed values, with observed years used directly. Pooled: one profile for all plans. The marginal minus average cost comes from the regression of Table A14 with plan-age interactions. Table A14 reports the average gap.

describe how the industry moves and on which these quantities depend: the entry intensity of each coverage class, the number of people entering the private system, the enrollment of rival insurers, the enrollment in the insurer's other plans, and the conditions common to the market each year, such as costs. The profiles are estimated over the life of plans that have already lived through the evolution of the industry, so a profile of plan age estimates the expected value of each input at every age, averaged over the paths of $\mathcal{S}_t$ that Γ generates. Under the stationarity assumption 8, the paths that older plans lived through are representative of the paths a plan introduced today will live through, so these estimated profiles are the insurer's beliefs. In the computation we evaluate the profiles at the average state of the industry in the sample, as in the oblivious equilibrium of Weintraub, Benkard, and Van Roy (2010). We build the profiles in three steps.

First, we estimate the profile of a typical plan. At each plan age, we take the median across plans in the 2017 to 2019 panel. We use the median because most plans are small, and an enrollment-weighted mean would describe the few largest plans instead. The panel observes plans up to age 11 well. Beyond that age, few plans are observed and most of them were introduced around the 2005 reform, so we extend the profile with a smooth function of age that follows the plan-age effects of the regression

$$\mathbb{E}\big[y_{jt}\big] = h\big(\phi_{a(j,t)} + \mu_t + \iota_{f(j)} + \kappa_{c(j)} + x_j'\gamma + \eta\, E_{c(j),t}\big), \tag{73}$$

where $a(j,t)$ is the age of plan j in year t, ϕ_a are plan-age effects, μ_t, ι_f and κ_c are year, insurer and coverage-class effects, x_j are the plan's characteristics at introduction, and $E_{c,t}$ is the entry intensity the plan faces (Tables A16 and A17). New enrollees are counts, estimated with a Poisson specification with the relevant population as exposure, so their plan-age effects are rates. Plans stop receiving new enrollees at age 15, as they close to new enrollment.

Second, we adapt the typical profile to each plan. Plans differ in size and in the cost of their enrollees, but we assume they follow the same shape over their life. We therefore move the typical profile up or down to match the plan's own observed values, and we use the observed values directly in the years we observe the plan.

Third, the inputs that are common to all plans are taken directly from their estimates. The difference between the cost of the marginal and the average enrollee comes from the regression of Table A14 with plan-age interactions (not reported), from -1.7 UF in the first two years of the plan to -11.3 UF at ages 18 and above. The cost loading of each insurer comes from Table A15, column 1.

All quantities are in UF per year. The condition is evaluated over the first 20 years of the plan's life.

Table A14: Cost of movers relative to the own-plan peers they leave behind

	(1)	(2)	(3)	(4)
Cost	Demographic-adjusted	Raw	Demographic-adjusted	Raw
Benchmark	Own-plan peers	Own-plan peers	Aggregate-plan peers	Plan × year FE
Switch	−3.728***	−2.891***	−3.846***	−3.763***
	(0.345)	(0.369)	(0.377)	(0.362)
Exit	−0.663	0.292	−1.403**	−0.745
	(0.680)	(0.721)	(0.704)	(0.698)
N	3,293,966	3,293,966	3,313,416	3,293,966
Clusters	283	283	284	283

Notes: Individual level, origin years 2016 to 2018 with the transition observed in the following year; costs are annual insurer claims in UF. In columns (1)–(3), the dependent variable is the person's cost in the year before the move, measured relative to the mean cost of the other members of the same peer group in that year. The peer mean excludes the individual. The regressors are indicators for switching and exiting the system, with staying as the omitted category, so each coefficient represents the mover minus the stayers of the plan they leave. Switch: the person moves to another plan and remains in the ISAPRE system. Exit: the person leaves the system. Column (1) is the primary estimate: costs are demographic-adjusted, that is, measured relative to the mean cost of enrollees of the same age group, gender and year, so the comparison holds demographics constant. Column (2) uses raw costs; the difference from column (1) measures how much of the gap is due to the demographic composition of the plans movers leave. Column (3) repeats column (1) with peers defined at the aggregate-plan level of the demand model. Column (4) uses raw costs and absorbs plan × year fixed effects instead of subtracting the peer mean; with the demographic fixed effects, it is the analogue of column (1). All columns absorb age group × gender × year fixed effects. Standard errors in parentheses, clustered by aggregate plan. Plans introduced before 2016 are included. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

F.3 Inputs from the demand model

The transmission, the mobility correction and the rating weight of the enrollees who respond are individual objects. The numerator of the transmission (69) pairs each person's response in the previous year, $ds_{im}(\tau-1)/dp_j$, with that same person's transition probabilities $\Pr_{i,t+\tau}(j \mid m)$. We compute it person by person and aggregate afterwards, rather than replacing the transitions with a population average, which would discard the covariance between who moves and who is price sensitive. We implement it on the 2018 and 2019 pair, for the 210,742 individuals we observe in both years and the $|\mathcal{J}| = 589$ plans in the menu. An individual's observed 2019 choice probabilities are their base utilities scaled by the multiplier that their observed 2018 plan induces through inertia, so dividing that multiplier out returns the base utilities up to an individual constant that cancels in every probability we form. Forming $\Pr_{i,t}(j \mid m)$ for every origin requires one logit denominator per origin, or $|\mathcal{J}|^2$ operations per person. The multiplier takes only three values: one when j is the plan held, a second when j belongs to the same insurer, and one otherwise. The denominator for origin m therefore equals a term common to all origins plus two corrections, one for the insurer of m and one for m itself, so each person's denominators cost $O(|\mathcal{J}|)$ operations while the transitions remain exact. Each individual contributes the outer product of two vectors with one entry per plan, so the sum over people is a matrix product. One pass delivers three aggregates: unweighted for Ω_j, weighted by rating weights for revenue, and weighted by cost for G_j.

From these aggregates we obtain the inputs of Table A13. The transmission is written as $\hat{T}_j(\tau) = \hat{\varsigma}(\tau)\,\hat{r}_j(\tau)$, where $\hat{\varsigma}(\tau)$, the ratio of the transmission to the stay probability, is measured at plan ages 0, 1 and 2 and held at the mean of ages 1 and 2 beyond. The mobility correction $\hat{\rho}(\tau) = \mathbb{E}[P^2]/\mathbb{E}[P]^2$ of (70) is measured on last year's enrollees of plans aged 1, 2 and 3 and held at its last value beyond. The insurer share of the alternatives of new enrollees, $\hat{\pi}^{sh}(\tau)$, is measured at ages 0, 1 and 2 and held beyond. The average rating weight of the enrollees who respond to the price, $\hat{\rho}^{rw}(\tau)$, is the ratio of the response weighted by rating weights to the unweighted response. It is measured at ages 0 to 4 and held at its last value beyond, and for each plan it is scaled by the plan's mean rating weight relative to the pooled one, $\hat{\rho}^{rw}_j(\tau) = \hat{\rho}^{rw}(\tau)\,\widehat{rw}_j(\tau)/\widehat{rw}(\tau)$. The price sensitivity $\hat{\bar{\alpha}}_j$ starts from the response at introduction per new enrollee, $\Omega_j(0)/n_j(0)$, averaged over plans at age 0 with weights given by their discounted enrollment. For each plan it is scaled by the mean of the demand model's $\alpha(z_i)$ over the age, income and sex composition of the plan's enrollees at introduction, relative to the average over plans, with the scale bounded between 0.5 and 2. The plan's price sensitivity is held at its value at introduction over the plan's life. The plan's elasticity still changes with its age, because the share of last year's enrollees, who respond only when close to leaving, grows as the plan ages.

F.4 Estimates and composed paths

Table A15 reports the cost loading by insurer. Tables A16 and A17 report the plan-age effects of (73), which give the shape of the profiles beyond age 11. Figure 13 shows the resulting profiles

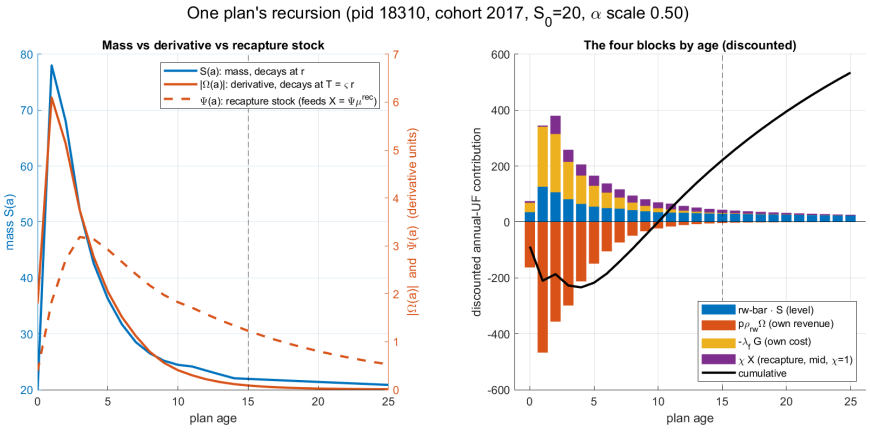
for one plan, and Figure 12 the paths and the blocks of the condition they produce for the same plan.

Table A15: Per-insurer cost loading λ_f

	(1) Paid benefits	(2) Incl. not-liquidated claims
Insurer A	1.218***	1.226***
	(0.011)	(0.010)
Insurer B	1.326***	1.344***
	(0.027)	(0.037)
Insurer C	1.125***	1.138***
	(0.013)	(0.021)
Insurer D	1.184***	1.182***
	(0.020)	(0.026)
Insurer E	1.082***	1.085***
	(0.025)	(0.024)
Insurer F	1.132***	1.140***
	(0.011)	(0.014)
Insurer-years	22	22

Notes: Through-the-origin regression of the accounting loss ratio on the micro loss ratio, one slope $\hat{\lambda}_f$ per insurer. Column (2) adds claims occurred but not yet liquidated to paid benefits. Robust standard errors in parentheses. Sample 2016 to 2019. Insurer D from 2018 (earlier years dropped). The model uses column (1).

Figure 12: FOC composed elements



Notes: FOC elements through time for one plan,

Table A16: Arrivals by origin: plan-age effects

	n^{new}	n^{rival}	n^{own}
Plan age 1	0.130	1.221***	1.104***
	(0.220)	(0.250)	(0.170)
Plan age 2	−2.722***	−0.548	−0.501
	(0.342)	(0.475)	(0.344)
Plan age 3	−4.112***	−2.109**	−2.252**
	(0.417)	(0.694)	(0.796)
Plan age 4	−4.587***	−2.898***	−1.465*
	(0.307)	(0.551)	(0.667)
Plan age 5	−4.719***	−3.221***	−1.675*
	(0.292)	(0.477)	(0.762)
Plan age 6	−4.670***	−2.995***	−1.575*
	(0.188)	(0.603)	(0.624)
Plan age 7	−5.587***	−3.698**	−3.493**
	(0.804)	(0.962)	(0.906)
Plan age 8	−7.988***	−6.995***	−5.282***
	(0.676)	(0.706)	(0.621)
Plan age 9	−8.091***	−7.836***	−5.917***
	(0.723)	(1.251)	(1.207)
Plan age 10	−7.841***	−9.302***	−7.031***
	(0.503)	(0.958)	(0.554)
Plan age 11	−7.822***	−10.284***	−7.157***
	(0.703)	(1.172)	(0.692)
Plan age 12	−7.894***	−8.057***	−6.881***
	(0.675)	(1.314)	(0.967)
Entry exposure (E^{class})	0.534	0.128	0.459
	(0.860)	(0.834)	(1.023)
Year FE	Yes	Yes	Yes
Insurer FE	Yes	Yes	Yes
Coverage-class FE	Yes	Yes	Yes
Birth characteristics	Yes	Yes	Yes
Observations	64,446	64,446	64,446

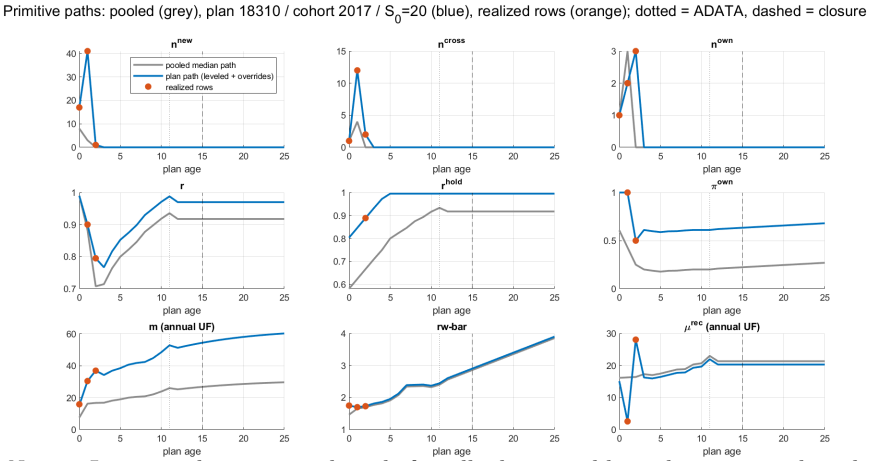
Notes: Poisson regressions at the plan-year level, 2017 to 2019; one observation per plan and year. The dependent variable counts arrivals by origin: new to the private system (n^{new}), from another insurer (n^{rival}), from another plan of the same insurer (n^{own}). Exposures are specific to each origin: the log of the year's entrants to the system for n^{new}, the log of the enrollment outside the insurer for n^{rival}, and the log of the enrollment in the insurer's other plans for n^{own}. Plan-age effects are relative to age 0, the year of introduction. Ages 13 to 21 are estimated but not displayed; ages 22 and above are pooled. Birth characteristics: base price at introduction, mean coverage, and density of the menu at introduction. Entry exposure is the entry share of the segment and year. Standard errors in parentheses, clustered by insurer (6 clusters). * p<0.10, ** p<0.05, *** p<0.01.

Table A17: Stay probability, own-insurer share of leavers, average cost and mean rating weight: plan-age effects

	r	r^{hold}	π^{own}	$\ln \bar{c}$	$\overline{rw}$
Plan age 1				0.777**	0.082
				(0.198)	(0.051)
Plan age 2	−0.134***		−0.115***	0.806**	0.134
	(0.018)		(0.017)	(0.242)	(0.068)
Plan age 3	−0.154***	0.041***	−0.158***	0.852**	0.196
	(0.022)	(0.005)	(0.016)	(0.269)	(0.099)
Plan age 4	−0.108***	0.112***	−0.135***	0.911**	0.292*
	(0.020)	(0.012)	(0.023)	(0.297)	(0.130)
Plan age 5	−0.077***	0.140***	−0.136***	0.932**	0.392**
	(0.018)	(0.015)	(0.018)	(0.286)	(0.147)
Plan age 6	−0.054*	0.169***	−0.132***	1.018**	0.471**
	(0.023)	(0.018)	(0.018)	(0.296)	(0.160)
Plan age 7	−0.023	0.198***	−0.136***	1.158***	0.634***
	(0.024)	(0.011)	(0.017)	(0.286)	(0.141)
Plan age 8	−0.004	0.210***	−0.128***	1.178***	0.721***
	(0.028)	(0.015)	(0.017)	(0.288)	(0.159)
Plan age 9	0.027	0.231***	−0.088***	1.220***	0.786***
	(0.020)	(0.012)	(0.019)	(0.263)	(0.166)
Plan age 10	0.048**	0.263***	−0.092***	1.313***	0.851***
	(0.016)	(0.015)	(0.014)	(0.255)	(0.163)
Plan age 11	0.057**	0.268***	−0.067*	1.367***	0.953***
	(0.021)	(0.012)	(0.031)	(0.240)	(0.163)
Plan age 12	0.070***	0.287***	−0.102***	1.202***	1.177***
	(0.017)	(0.017)	(0.004)	(0.279)	(0.162)
Plan age 15	0.071**	0.284***	−0.027	1.178***	1.566***
	(0.018)	(0.014)	(0.042)	(0.284)	(0.259)
Plan age 18	0.084***	0.294***	0.028	1.405***	1.947***
	(0.020)	(0.017)	(0.079)	(0.264)	(0.400)
Plan age 21	0.100***	0.306***	−0.050	1.442***	1.653**
	(0.008)	(0.031)	(0.038)	(0.284)	(0.470)
Entry exposure (E^{class})	0.059	0.110	−0.182**	−0.166*	0.010
	(0.048)	(0.083)	(0.065)	(0.081)	(0.119)
Year FE	Yes	Yes	Yes	Yes	Yes
Insurer FE	Yes	Yes	Yes	Yes	Yes
Coverage-class FE	Yes	Yes	Yes	Yes	Yes
Birth characteristics	Yes	Yes	Yes	Yes	Yes
Observations	60,688	38,811	31,629	62,593	64,446

Notes: OLS at the plan-year level, 2017 to 2019; r^{hold} uses 2018 and 2019 and plan ages 2 and above, since tenure requires two prior panel years. r is the retention of the full pool of the year before, weighted by lagged enrollment; r^{hold} is the retention of tenured members, in the plan in both previous years, weighted by their count; π^{own} is the share of leavers landing in plans of the same insurer, weighted by leavers; $\ln \bar{c}$ is the log average cost of the pool; $\overline{rw}$ is the mean rating weight of the pool. Plan-age effects are relative to the youngest age in each sample (0 for $\ln \bar{c}$ and $\overline{rw}$, 1 for r and π^{own}, 2 for r^{hold}); blank cells are ages at or below the omitted base. Ages 13, 14, 16, 17, 19 and 20 are estimated but not displayed; ages 22 and above are pooled. Controls as in Table A16. Standard errors in parentheses, clustered by insurer (6 clusters; analytic $t(5)$ inference). * p<0.10, ** p<0.05, *** p<0.01.

Figure 13: Estimated transition objects paths



Notes: In gray, the estimated path for all plans, in blue, the estimated path correctly shifter by this plans' level, and in orange the observed data points for this plan.

F.5 The first two years of the exact condition

Because the 2018 and 2019 menus are observed, we can compute the exact condition for the first two years of the plans introduced in 2018. This section uses these two years to illustrate why the future matters for the price. The figures are illustrative, but the approximated condition of Section 5.3.1 uses the exact objects plan by plan.

We measure the response of a plan to its price with its elasticity. In the year of introduction it is the static elasticity, $\eta_j(0) = p_j\,\Omega_j(0)/S_j(0)$. In the second year, using (68), it splits into two terms,

$$\eta_j(1) = \frac{p_j\,\Omega_j(1)}{S_j(1)} = \underbrace{\frac{p_j\,d_j(1)}{S_j(1)}}_{\text{direct}} + \underbrace{\frac{p_j\,T_j(1)\,\Omega_j(0)}{S_j(1)}}_{\text{persistence}}. \tag{74}$$

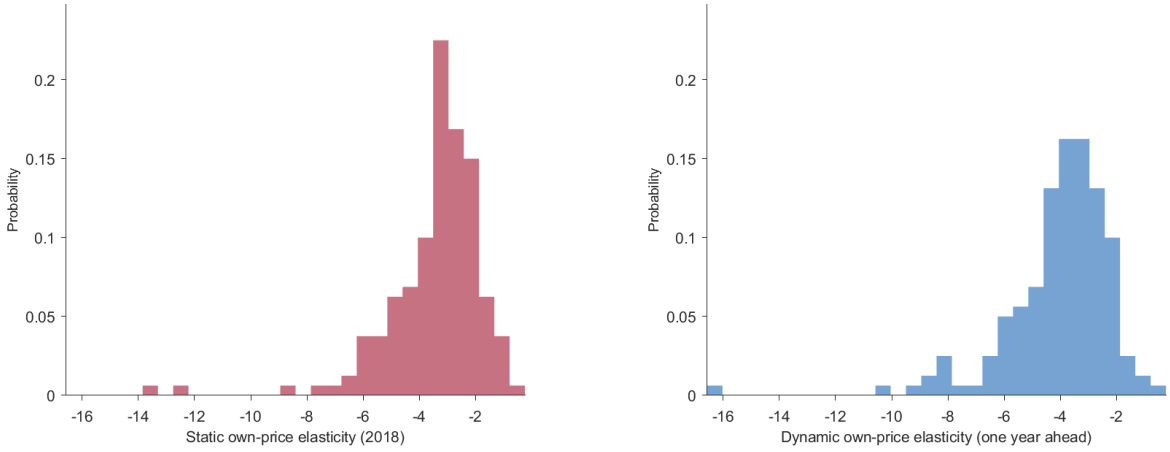
The direct term is the response of this year's choices: the new enrollees of the plan and last year's enrollees close to leaving. The persistence term is last year's response carried forward: the enrollees the price attracted in the first year are still in the plan, and those it turned away are still out of it, net of those who come back.

In Figure 14, the histogram of the static elasticities $\eta_j(0)$ and the dynamic elasticities $\eta_j(1)$ of the plans introduced in 2018 shows that the dynamic elasticities tend to be larger than the static ones. When deciding the price of the plan for its full lifetime, insurers take into account how the composition of the enrollment will evolve. According to Figure 15, insurers would underestimate how elastic consumers are in a given year if they did not take into account the persistence term. We decompose the elasticity of the plan into the direct term of (74), given by the price sensitivity in each year, and the persistence term, which captures the previous response carried forward. Figure 15 compares different elasticity definitions. Under 2018 static, the dots represent the elasticity of each plan taking into account only the direct term. If the insurer took into account the persistence term but only had information on the elasticities of 2018, the result would be the naive dynamic elasticity, which is larger. This reflects the compounding over years that comes from the repeated

interaction with the same consumers. However, the naive elasticity does not take into account that the direct term also changes, because part of the plan's enrollment in 2019 consists of enrollees who already chose the plan in 2018, and inertia gives them a much lower individual elasticity. The plan elasticity depends on the composition of the plan in 2019, which combines the new enrollees of the plan and last year's enrollees. When we take into account the direct term of 2019, the elasticity shrinks back. Figure 16 illustrates the decomposition of enrollees: last year's enrollees have a much smaller individual elasticity, while the new enrollees of the plan are quite elastic. Depending on how they are combined, the direct term in 2019 can be larger or smaller than in 2018.

Figure 17 uses the two years to decompose the price implied by the exact condition, by insurer, into the markup, the own cost term, the cost and margins of the insurer's other plans, both those existing at introduction and those introduced later, and compares it with the observed price.

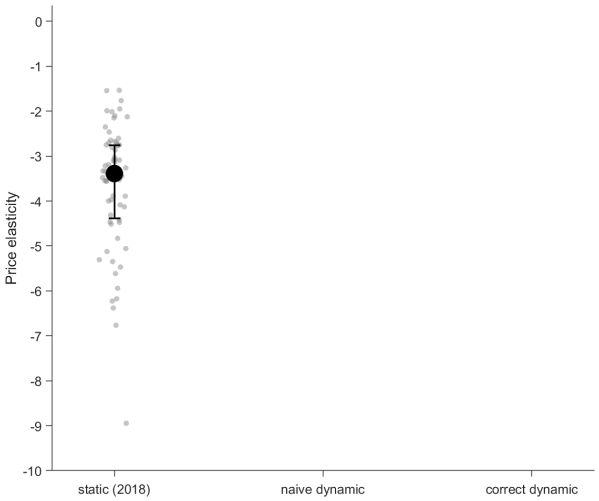
Figure 14: Elasticities



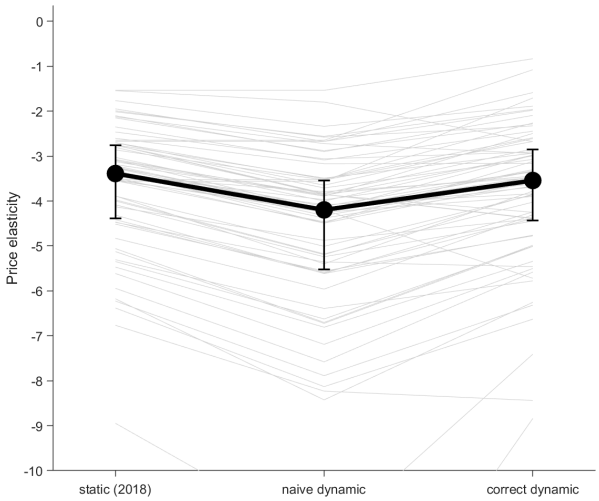
(a) Static elasticities

(b) Dynamic elasticities

Figure 15: Comparing elasticity definitions



Notes: Static elasticities in 2018. Points represent elasticities for different plans, summarized by the mean big black point.



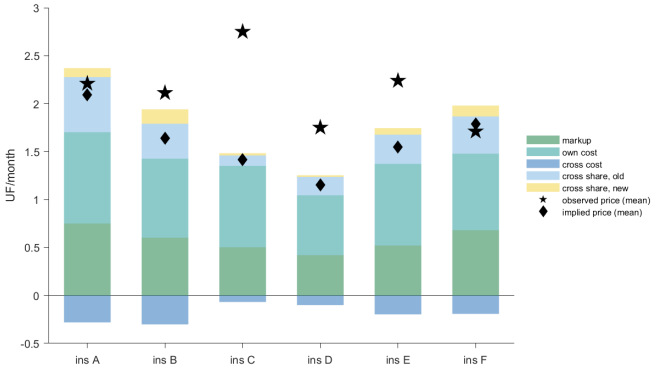
Notes: Naive dynamic accounts for the persistence term in the 2-year elasticity, but using the static one in 2018. Correct dynamic accounts for persistence and direct static elasticity in 2019.

Figure 16: Into the direct term of the elasticity: captive versus non.



Notes: Individual elasticities, aggregated to plan 1 in the diamond, and in 2019 enrollees divided by captive: had the same plan last period or did not have the plan last period.

Figure 17: Structural FOC elements - 2 years



Notes: Decomposition of the FOC terms for plans introduced in 2018.

Figure 18: Approximated FOC residuals, expressed as percentage of price



Notes: Residuals of the Approximated FOC.

F.6 How much of the recaptured margin do insurers internalize?

Section 5.3.1 evaluates the condition with the recaptured margin fully internalized. Here we introduce a parameter χ that scales the recaptured margin, replacing $\hat{X}_j(\tau)$ with $\chi\hat{X}_j(\tau)$ in (17). A value of $\chi = 1$ means the insurer takes the recaptured margin fully into account, and $\chi = 0$ that it prices each plan on its own.

We estimate χ by GMM. For each plan, the residual of (17) divided by the revenue lost at the margin is

$$e_j(\chi) = \frac{p_j^*(\chi) - p_j}{p_j}, \tag{75}$$

the price gap shown in Figure 18. Under rational expectations at introduction, this gap is uncorrelated with any variable z_j that the insurer knew when it set the price, so $\mathbb{E}[z_j\, e_j(\chi)] = 0$. The instruments are insurer dummies, dummies for the tercile of the observed price, entry exposure at introduction, and cost and balance-sheet shifters by insurer or cluster and year of introduction.

We also add moments that do not contain parameters. These require the forecast errors of the profiles, defined as observed minus predicted enrollment, rating weight and average cost at plan ages 1 to 3, to be uncorrelated with the insurer dummies. Standard errors are clustered by insurer and coverage class.

Table A18 reports the estimates. A single value gives $\hat{\chi} = 0.43$. By tercile of the observed price, χ rises from 0.03 for the cheapest plans to 0.41 and 0.84, and equal values are rejected. The forecast-error moments, which test the estimated transitions behind the future terms, do not reject ($J = 52.1$, $p = 0.84$).

A higher price affects the margins an insurer recaptures in its other plans in two opposite ways. It pushes some enrollees out of the plan today, and those who move to the insurer's other plans are not lost. But it also leaves the plan with fewer enrollees, and the plan's enrollees are the people who move to the insurer's newer plans in later years. A higher price therefore raises recapture today and lowers it in the future. The approximated condition counts only the first effect. The omitted effect is largest for cheap plans. Their enrollees are the most price sensitive, so a price increase shrinks the pool the most. They also lose enrollees to newer plans at the highest rate, so the smaller pool would have supplied many future recaptures. And when a plan is introduced, no newer plan exists yet, so the effect the approximation does count starts only a year later. The approximation therefore overstates how much a cheap plan gains from raising its price, and the estimated χ falls toward zero to compensate.

F.7 Description of the industry using the approximated condition

This section uses the approximated condition to describe how much of the price depends on the future and on recapture, and what happens when one insurer stops introducing plans.

Because (18) is a sum over the plan's life, the implied price splits into the price that the year of introduction alone would set and a future part[29]. Let $D_{j,0} = -\hat{\rho}_j^{rw}(0)\Omega_j(0)$ be the first term of D_j and $P_{j,0} = \left[\widehat{rw}_j(0)S_j(0) - \lambda_f G_j(0) + \hat{X}_j(0)\right]/D_{j,0}$ the price that the year of introduction alone would set. Then $p_j^* = P_{j,0} + P_j^{cont}$, with

$$P_j^{cont} = \underbrace{\frac{1}{D_j}\sum_{\tau\geq 1}\beta^\tau\, \widehat{rw}_j(\tau)S_j(\tau)}_{\text{level (harvest)}} - \underbrace{\frac{1}{D_j}\sum_{\tau\geq 1}\beta^\tau \lambda_f G_j(\tau)}_{\text{cost at the margin}} + \underbrace{\frac{1}{D_j}\sum_{\tau\geq 1}\beta^\tau \hat{X}_j(\tau)}_{\text{recapture}} - \underbrace{P_{j,0}\,\frac{D_j - D_{j,0}}{D_j}}_{\text{attrition penalty}}. \quad (76)$$

The last term is the price reduction that comes from future enrollment responding to today's price. The future part represents about 58% of the base price of the median plan, and 61% weighted by enrollment. Recapture accounts for about 31% of it, so the future part of a multiproduct insurer is about 45% larger than that of a single-plan insurer: part of the enrollment a higher price pushes out lands in the insurer's other plans[30]. The recapture term can be split into plans that existed when the plan was introduced and plans introduced later. About 80% of recapture comes from

[29] The year of introduction is not a static condition, because inertia still operates in the demand model.

[30] These numbers correspond to the recaptured margin fully internalized, $\chi = 1$. At $\chi = 0.43$ the two shares are 45% and 16%.

Table A18: Internalization of recapture: GMM on the composed introduction-price FOC

	(1) Single χ	(2) χ by price tercile
χ	0.425	
	(0.057)	
χ_{T1} (lowest paid price)		0.030
		(0.106)
χ_{T2}		0.407
		(0.095)
χ_{T3} (highest paid price)		0.835
		(0.133)
Hansen J (p-value)	21.0 (0.10)	22.5 (0.03)
Plans	2,635	2,635
Clusters (insurer × class)	64	64

Notes: Linear GMM on the composed set-once FOC in bracket units, $\big(A_j + p_j\rho_{rw}\Omega_j - \lambda_f G_j + \chi X_j\big)/|p_j\rho_{rw}\Omega_j|$, at horizon $\bar{\tau} = 20$; G prices the deterred enrollee at the marginal leaver's cost; the recapture term X uses the measured margins on the firm's own recaptured leavers, with deterred own-switchers kept at their observed firm-retention rate ($w_{im} = \pi^{own}$). The loading λ_f is fixed at its accounting estimate per insurer. Instruments (15): insurer dummies, paid-price tercile dummies, and cluster-by-introduction-year cost and balance-sheet shifters. Standard errors in parentheses, cluster-robust at insurer × class, conditional on the composed FOC blocks (pipeline-bootstrap standard errors pending). $\chi = 1$ (full internalization) is rejected in both columns. A Wald test of $\chi_{T1} = \chi_{T2} = \chi_{T3}$ rejects the single-χ restriction ($W = 14.5$, $p = 0.001$). The parameter-free forecast-error moments on the primitive paths do not reject the expectations model behind the continuation ($J = 52.1$, $p = 0.84$).

plans introduced later, from plan age 2 onward.

We also study a one-shot deviation in which one insurer stops introducing plans while its rivals continue, holding rivals' prices fixed. This exercise uses the approximated condition at the estimated $\hat{\chi} = 0.43$. The recapture terms for new plans disappear, and the result depends on where the enrollees who were going to move to new plans go instead. If all of them stay in the plan, the gain in margins on the plan corresponds to a price 0.38 UF per month higher (19% of the base price, weighted average), and the lost recapture lowers the net change to 0.32 UF. If they still move away, the gain in margins is the same but the lost recapture is larger, and the price falls by 0.11 UF. At the measured share of these enrollees who would stay, the two effects offset and the price does not change. One insurer that stops introducing plans on its own therefore gains almost nothing on the price. Section 6.1.2 solves for prices when all insurers cannot introduce new plans.

G Simulated Costs: Markov Chain Model

Individual costs are estimated from panel data covering 2016 to 2021. Each person-year observation is assigned to one of five risk states based on cost quintiles within age group, gender, and year. We estimate the following objects from the data: transition matrices across risk states by age group and gender, a cost schedule for each risk state consisting of the share of zero-cost observations and the mean and standard deviation of log costs for positive spells, and the prevalence of diabetes and cardiology conditions together with their added log-cost effects.

Starting from each individual last observed risk state, we simulate 100 forward paths of up to 50 years. In each year of a simulated path, we draw a new risk state from the relevant transition matrix and then draw costs from a lognormal distribution, adjusting for chronic conditions. Conditions evolve with fixed annual onset probabilities and are permanent once acquired. We then average simulated costs across the 100 paths at horizons of 1, 2, 3, 5, 10, 20, and 50 years. These average simulated costs serve as inputs to a Gradient boosting model that predicts expected out-of-pocket costs per plan, accounting for plan-specific cost-sharing rules. The resulting predictions are used as c_i in the counterfactuals (Section 6.1.2 and 6.2) and as input to $EOP_{ij\tau}$ in the simulated demand model.

H Counterfactual No Entry

To construct a no entry counterfactual that affects the prices of plans in 2018, we build a structural FOC with one new-plan option for each insurer and each period. The new-plan option represents all newly introduced plans in that period. When we remove this plan for all insurers, new joiners must choose between the existing plans and the outside option, while movers can no longer switch to newly introduced plans. We first estimate the parameters of the new-plan options using the demand model. We simulate the model over a lifetime, for individuals from age 26 to age 56, and require it to match the flows observed in the data. These flows include individuals entering and leaving the system and individuals switching within the system. We construct the corresponding moments using auxiliary regressions and estimate the parameters by minimum distance so that the moments from the simulated model match the moments observed in the data. Once we estimate these parameters, we recompute prices using the FOC. The FOC now includes a new-plan option that enters directly into the recapture term for plans priced in 2018.

H.1 Estimating new-plan option parameters

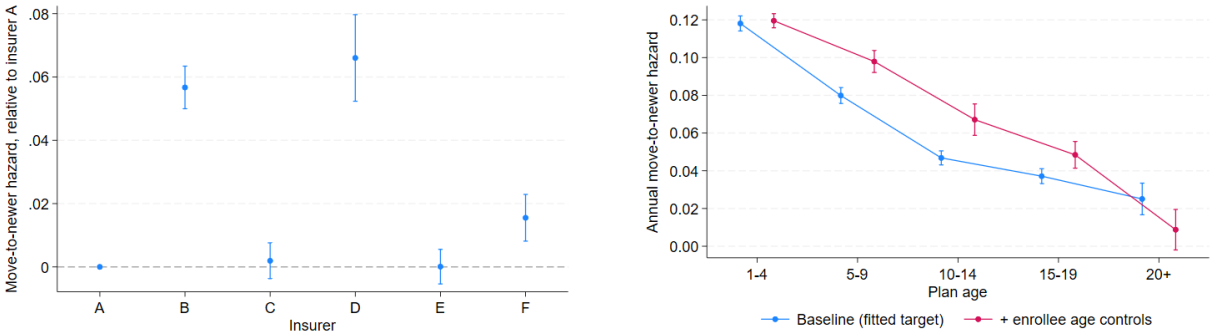
In this section we explain the auxiliary regressions estimated in the data. Matching these coefficients between data and model is equivalent to imposing $\mathbb{E}[x_i(y_i - P_i(\theta^{new}))] = 0$, with y_i outcome variable of the auxiliary regression and P_i the choice probability of the demand model, the standard micro-moment condition, with the regressors as instruments.

Auxiliary regression 1: Switching into newer plans. This moment represents the flow of individuals into plans that were introduced after the plan under analysis, which is being priced. The hazard rate is smaller as the plan under analysis is older and insurers B and D have a higher rate of movers into newer plans. In the model, the newer plans are represented by a unique per-period new-plan option, and therefore the new-plan options need to absorb the switching rates of Figure (19). The outcome f_{ijt}^{new} equals one if enrollee i, holding plan j in year t, is enrolled in year $t+1$ in a plan introduced more recently than j. The estimated equation is the linear probability model

$$f_{ijt}^{new} \;=\; \alpha \;+\; \sum_{a} \beta_a \, \mathbf{1}\{\text{age}(j,t) \in a\} \;+\; \sum_{f} \gamma_f \, \mathbf{1}\{j \in f\} \;+\; \delta_t \;+\; \epsilon_{ijt}, \tag{77}$$

where a indexes plan-age bands (base 1-4; plan age 0 is excluded, since a more recently introduced plan barely exists), f indexes insurers (base insurer A), and δ_t are year effects. Standard errors are clustered at the similar-plan level. The moments taken to the estimation are α, the β_a, and the γ_f. This regression should pin down the insurer-specific mean taste utility of the new-plan option δ_{0f} and the mean taste preference parameter specific to the age of the incumbent plan g_a of equation (19) (10 parameters).

Figure 19: Moment 1: Hazard rate movers to newer plans



(a) Insurer rate relative to A.

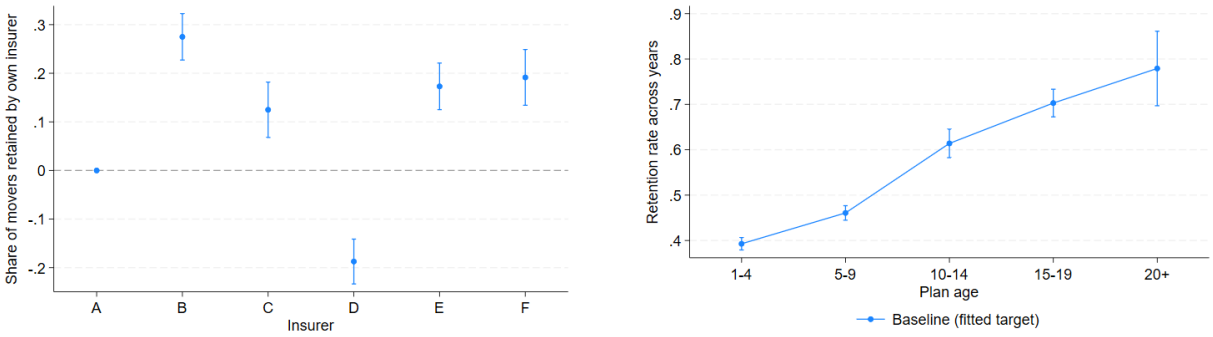
(b) Rates across plan lifetime. Red-line controls for enrollee's age.

Notes: Hazard rates of movers into plans that were introduced after the priced plan. The regression is a linear probability model of moving to newer plan on insurer and plan age fixed effects. In the left, coefficients represent pure insurer effects, in the right, the marginal effects are reported. Clustered standard errors at the similar-plans level.

Auxiliary regression 2: Movers to newer plans retained within-insurer. This regression represents the share of movers to newer plans that stay within the same insurer. The regression is a linear probability model of moving to newer plan within same insurer on insurer and plan age fixed effects. In the model, the new-plan option's own-insurer term has one retention parameter per insurer, $g_{O,f}$, that does not depend on the age of the plan the enrollee leaves. The model can only match the retention gradient across time through the changing composition of the enrollees. The data shows the own-insurer share among movers rises from 0.39 in young plans to 0.78 in old plans, while the model matches retention for old plans (0.8) but overstates it for early years (0.6 retention share). Insurers reclassify their members into their own new plans more intensely

the longer those members have been held. Matching the moments of this regression identifies the per-insurer (new-plan option) retention parameter g_{Of} from equation (19) , also using the already estimated insurer-inertia parameter. The plan's age dummies work as overidentifying restrictions.

Figure 20: Auxiliary regression 2: Retention rate of movers to newer plans



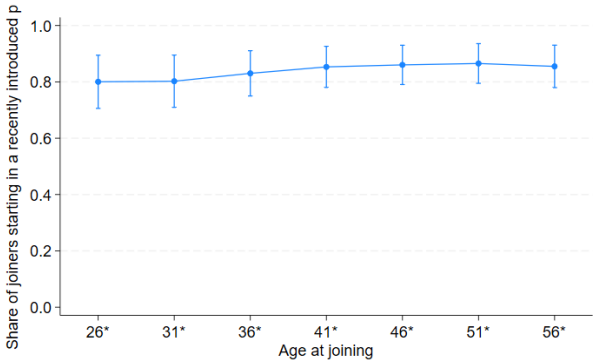
(a) Insurer retention rate relative to A.

(b) Rates across plan lifetime.

Notes: Among movers to newer plans, share retained by same insurer. The regression is a linear probability model of moving to newer plan within same insurer on insurer and plan age fixed effects. In the left, coefficients represent pure insurer effects, in the right, the marginal effects across plan's age are reported. Clustered standard errors at the similar-plans level.

Auxiliary regression 3: Share of newcomers choosing new plans. This regression represents the share of newcomers that are joining a recently introduced plan —introduced the same year or one year before joining the private sector—. This moment disciplines the new-plan option's appeal to system entrants. In the model, individuals joining at band age a choose between the observed plans and the new-plan options, and the parameters h_a are fitted so the share choosing the new-plan option reproduces these landing shares, at every age, not just for the young. New plans absorb joiners of all ages at the same high rate, so in the no-entry counterfactual the whole age range of joiners, not only the young, faces a smaller choice-set.

Figure 21: Share of newcomers in new plans.



Notes: Share of newcomers starting in a recently introduced plan, by age at joining. The sample is individuals entering the system between 2016 and 2019. Each point is the share whose first plan had been introduced recently, from a linear probability model on age-at-joining bands with year effects (base 26–30), marginal effects are plotted, bars are 95% confidence intervals, standard errors clustered at the similar-plan level. The share is high and nearly flat: at every joining age, more than 80% of new members start in a recently introduced plan.

Table A19: Auxiliary regression 4: Insurer's cost of movers.

	(1) Cost (UF/yr)	(2) Cost (UF/yr)	(3) Cost (pctile)
Mover to newer plan	-4.848***	-5.118***	-1.941***
	(0.485)	(0.489)	(0.220)
Mean cost (base)	25.02		
Age× year FE	Yes	Yes	No
Plan FE	Yes	No	No
Plan× year FE	No	Yes	Yes
N	2,183,767	2,183,746	2,183,746

Notes: Standard errors clustered at the similar-plan level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Auxiliary regression 4: It measures how much cheaper the enrollees who move to a newer plan are, relative to the enrollees who stay, comparing individuals within the same pool and the same age-by-year cell. Movers carry about 20 percent of a mean member's cost out of the pool. This moment identifies the health term of the new-plan option, φ: switching to a new plan is easier when healthy, because new plans can screen applicants. The demand model has an expected-out-of-pocket term, which mainly works as preferring higher coverage rates if sick, but it cannot reproduce the switching rate of the healthy. So without this moment the counterfactual would misstate how much cost leaves a plan when its enrollees move away. Table A19 shows the regression of insurer's cost on individual switching to new plans under 3 specifications. The fit the structural model uses is the second column. Column 3 constructs the rank of the individual within demographic cell, so movers to newer plans sit about 2 percentile points lower in the cost distribution of their demographic cell than stayers in the same pool-year.

Fit. Table A20 shows the fit between the auxiliary regressions and the model. The model uses the sample of individuals aged 26 in 2018, and are simulated through their lifecycle until age 56, with their costs constructed with the Markov chain model. The model reproduces the erosion side of the data well. The hazard of moving to a newer plan falls from 0.10 per year in young plans to 0.03 in old plans in the data, and the model matches both the level and the decline in the first panel.

The joiners' side is matched exactly. At every joining age, the model sends slightly above 80 percent of new members to the new-plan option, as in the data in the third panel. This match is mechanical rather than a test. The model has one joiner parameter per age group, so the shares shown in the panel are reproduced exactly by construction.

The cost composition is also matched almost exactly. Movers are 20.3 percent cheaper than stayers in the model, compared with 20.5 percent in the data in the fourth panel. The moments with a worse fit are concentrated in the second panel, and they are informative. First, the model retains too many movers within their own insurer. The base own-insurer share is 0.47 in the model compared with 0.27 in the data (age 1 to 4 for insurer A). Second, the data show mover

retention rising strongly with plan age, from 0.39 in young plans to 0.78 in old plans. The retention parameters are constant over the life of the plan, so the model can change this share only through changes in the composition of movers. This generates about half of the increase. The model matches retention in the oldest plans but overstates it in young plans. The data therefore reject constant retention over the plan lifecycle, with the gap concentrated in young plans. Insurers retain fewer early leavers than the model implies. Third, part of the cross-insurer drain pattern remains unmatched. Insurers B and D lose members faster in the data than in the model, while insurers C and E lose members more slowly. These differences in lock-in cannot be generated by the destination-side parameters. The main consequence of the first mismatch is that the model somewhat overstates how much of each insurer's drain is recaptured internally. We therefore carry this as a caveat when interpreting the recapture credit in the pricing condition.

Four objects enter the model as constraints or measured inputs rather than fitted parameters. First, we fix the exit rate to the outside option, mainly because we use a simulated path for the age 26 cohort. Each enrollee's probability of leaving the system is rescaled by a common factor for each gender and age band. We choose this factor so that, at observed prices, the aggregate exit share matches the exit rate measured in the data. The remaining choice probabilities are rescaled proportionally, so the relative choices among plans remain unchanged. Second, we fix the joining rate in the same way. Joining probabilities are rescaled by a common factor for each gender and age band, calibrated so that, at observed prices, the aggregate joining share matches the joining rate measured in the data. We then hold these two factor fixed, so exit and joining rates can respond to prices and choice-set in both the entry and no entry scenarios. This is mainly needed to model the newcomers at other ages in the future lifecycle path. Third, the price of the new-plan option is fixed using the data. b_f^{new} is the measured mean introduction price of each insurer's recently introduced plans. Fourth, we construct the value of a recaptured enrollee, V_f^{rec}, using accounting data. It is based on the measured margin the insurer earns on members who move to its own newer plans, which we convert into a present value using the measured stay rate of settled enrollees. All demand parameters, including the price coefficients and inertia terms, come from the demand estimation and are not re-estimated.

Model Validation Table A21 show the insurer's market share for 26-year-old cohort, for the data and the model as a validation exercise as these are not targeted moments. Table A22 shows the implied exit rate from origin insurer and price tercile. Even if the market exit rate is targeted, we can still compare the exit rates by insurer.

H.2 Prices under Entry scenario

After estimating the new-plan options parameters, we recompute prices using the structural FOC for plans introduced in 2018, which simulates the future behavior of the plans following the cohort of 26 years-old in 2018. When an enrollee moves to a newer plan from the same insurer, the insurer keeps earning a margin on that enrollee in the new plan. The pricing condition assigns each recaptured enrollee a value of $V_f^{rec} = \mu_f^{rec}/(1 - \beta r)$. This value is based on two measured

Table A20: Moment fit of the new-plan options parameters

	Data	(s.e.)	Model
A. Annual rate of moving to a newer plan			
Base (insurer A, plan age 1–4)	0.096	(0.003)	0.097
Plan age 5–9	−0.038	(0.003)	−0.034
Plan age 10–14	−0.071	(0.003)	−0.071
Plan age 15–19	−0.081	(0.003)	−0.080
Plan age 20+	−0.093	(0.005)	−0.094
Insurer B	0.057	(0.003)	0.038
Insurer C	0.002	(0.003)	0.044
Insurer D	0.066	(0.007)	0.047
Insurer E	0.000	(0.003)	0.022
Insurer F	0.016	(0.004)	0.020
B. Own-insurer share among movers to newer plans			
Base (insurer A, plan age 1–4)	0.270	(0.024)	0.475
Insurer B	0.275	(0.024)	0.301
Insurer C	0.125	(0.029)	0.317
Insurer D	−0.187	(0.023)	−0.166
Insurer E	0.173	(0.024)	0.116
Insurer F	0.191	(0.029)	0.118
Plan age 5–9	0.068	(0.010)	0.036
Plan age 10–14	0.221	(0.017)	0.105
Plan age 15–19	0.310	(0.017)	0.139
Plan age 20+	0.386	(0.042)	0.184
C. Share of joiners starting in a recently introduced plan			
Joining at age 31–35	0.802	(0.006)	0.802
Joining at age 36–40	0.830	(0.013)	0.830
Joining at age 41–45	0.853	(0.018)	0.853
Joining at age 46–50	0.860	(0.020)	0.860
Joining at age 51–55	0.865	(0.022)	0.865
D. Health of movers			
Cost gap, movers vs. stayers	−0.205	(0.020)	−0.203

Notes: Each row is a regression coefficient computed with the same formula on the data and on the flows generated by the model at the estimated parameters. Panel A is a linear probability model of moving to a newer plan within the year on plan-age and insurer dummies (plan age 0 excluded). Panel B is the same regression on the subsample of movers to newer plans, with the own-insurer destination as the outcome. Panel C reports the share of new joiners whose first plan was recently introduced (introduced in the joining year or the year before), by age at joining (levels, base age 26–30), the five levels are matched exactly by construction. Panel D is the yearly cost difference between enrollees who move to a newer plan and those who stay, as a share of the mean cost. Data from the 2016–2019 enrollment panels. The estimation weights each moment by the inverse of its standard error; 26 moments against 22 parameters.

Table A21: Allocation of 26-year-old joiners across insurers: model vs. data

	Female		Male	
	Model	Data	Model	Data
Insurer A	0.321	0.338	0.211	0.219
Insurer B	0.280	0.280	0.308	0.233
Insurer C	0.024	0.024	0.027	0.024
Insurer D	0.023	0.043	0.079	0.087
Insurer E	0.150	0.180	0.178	0.237
Insurer F	0.203	0.134	0.197	0.200

Notes: Shares of enrolled 26-year-olds across insurers (conditional on choosing a plan), at observed prices. The model column is the implied choice. There is no estimated new-plan option parameters and no moment targets this allocation. Data: 2018 age-26 enrollment by insurer.

Table A22: Exit incidence across origins: model vs. data (relative to mean)

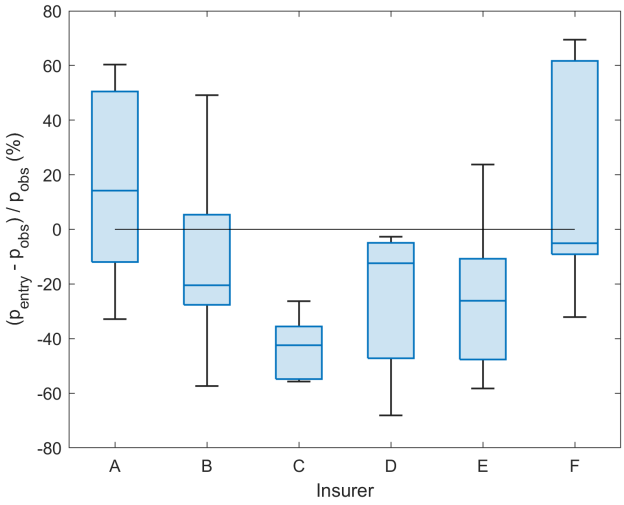
	Insurer						Price tercile		
	A	B	C	D	E	F	1	2	3
Data	0.95	0.92	0.71	1.33	0.98	1.10	1.33	0.97	0.70
Model	0.92	1.06	0.91	1.09	1.02	1.01	1.06	1.00	0.94

Notes: Exit rates by origin insurer and origin price tercile, each normalized by its own mean. The exit level is calibrated by construction; the incidence pattern across origins is a prediction. The model reproduces the ordering but understates the dispersion.

objects. μ_f^{rec} is the average annual margin that insurer f earns on enrollees who move into one of its newer plans. We compute this margin as the price received minus the claim cost paid by the insurer. Across insurers, the annual margin ranges from 7 to 12 UF. The term r is the annual stay rate of settled enrollees, which we take from the observed retention path of mature plans. Together, these objects imply values of V_f^{rec} between 45 and 80 UF across insurers. We measure V_f^{rec} directly from the data rather than estimate it. It is not used to fit the new-plan option parameters. It enters only the pricing condition, where it multiplies the model's derivative of the flow into the insurer's new-plan option. The individual's cost that enter the FOC are multiplied by the estimated loading parameter by insurer that matches into the accounting loss ratio (Table A15) .

Before running the counterfactual, we compare the prices solved by the model in the entry scenario with observed prices. The model predicts prices that are too low. The mass-weighted entry scenario price is about half of the observed price on average, while the non-weighted average ratio is 0.84. The ratio of model to observed prices average is 1.13 in the cheapest tercile, 0.78 in the middle tercile, and 0.60 in the most expensive tercile. Two parts of the model help explain this pattern. First, the model includes claim costs but not administrative costs, which lowers predicted prices across plans. Second, the logit demand model does not generate enough market power for the most expensive plans. In the model, demand for these plans is more elastic than what we observe for their more stable pools in the data. We report this limitation rather than adjust the model to match these prices. These differences do not affect the main counterfactual comparison because we compare two solutions from the same model. Any common difference in price levels appears in both cases. For this reason, we report the counterfactual results using the ratio of no-entry prices to entry scenario prices. Figure (22) presents the distribution of the difference between model and observed prices as a percentage of the observed price. The fit is worst for insurer C. For this insurer, the model implies an elasticity that is too high relative to the data. Enrollees in the data are much less likely to move away from insurer C than other insurers but the model does not include insurer-specific stickiness based on the enrollee's current insurer.

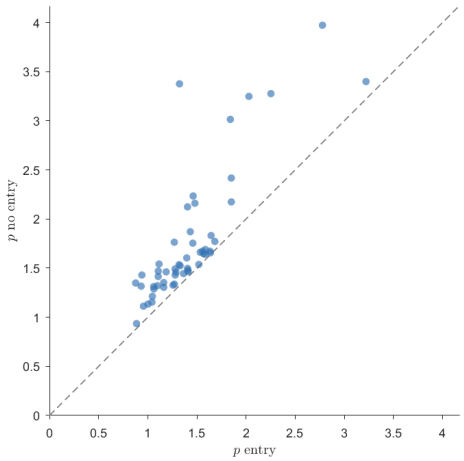
Figure 22: Price differences between (Entry) Model and Data.



*Notes:*Differences between observed prices and model-implied prices (entry) by insurer.

H.3 No Entry results

Figure 23: Prices under no entry are larger



Notes: Model implied prices from Entry scenario on the x-axis and prices from No-entry scenario in the y-axis.

I Comparison with perfect competition

This appendix compares the condition (12) of Section 4.3 with the long-term contracts of Hendel and Lizzeri (2003) in a two-period version in which the insurer chooses prices.

An individual joins in period 1 with expected cost c_1. In period 2 the individual is healthy (H) with probability Pr(H) and cost c_{H}, or sick (S) with cost $c_{\mathrm{S}} > c_{\mathrm{H}}$. Information and learning are symmetric: at joining, the insurer and the individual know the same about the individual's future health, and both learn period-2 health at the same time. Commitment is one-sided: the insurer cannot reprice an enrollee after joining, and the enrollee can leave. The plan offers one contract to all its enrollees, and a sick enrollee cannot switch (health declaration). Individuals

differ continuously in the probability of joining and in their value of staying (inertia, coverage). The probability of staying healthy is the same for all. Selection comes from health revealed in period 2: only the healthy can switch, so those who leave are healthy when they leave. The plan sets both payments (P_1, P_2) at introduction.

The insurer takes into account two constraints when choosing prices: participation constraint and lapsation constraint (probability of leaving when healthy). *Participation:* individual i joins with probability $\Pr_i(\text{join})$, decreasing in $\tilde{P} = P_1 + \beta\nu P_2$, where $\nu \in [0,1]$, common to all, is the weight joiners give the period-2 payment, relative to the first, when deciding whether to join. $\nu = 1$ when joiners anticipate the period-2 payment, expect to pay it, and face no borrowing constraints. The optimal-contract literature assumes instead that consumers cannot borrow, so $\nu < 1$: ν is the ratio of the marginal utilities of money in the two periods, which falls as the contract collects more in period 1, even with flat income, and is lower when income rises over time. This is what limits front-loading in Hendel and Lizzeri (2003) and Ghili, Handel, Hendel, and Whinston (2024). Two further motives lower ν in our setting: joiners who expect to leave before paying P_2, such as the young and mobile, which arises because one contract serves enrollees with different values of staying and myopic joiners who do not fully anticipate P_2. *Lapsation:* a healthy enrollee can leave the contract, and stays unless P_2 exceeds P^{new}, the price of the newest plan open to a healthy person, by more than their value of staying, so $\Pr_i(\text{stay} \mid \text{H})$ falls with P_2. On a healthy enrollee who leaves, the insurer keeps a margin $m^{rec} \geq 0$: zero if they leave the insurer, positive if they move to one of its new plans. The period-2 margin on joiner i is

$$M_{2,i} = \Pr(\text{H}) \Pr_i(\text{stay} \mid \text{H}) \, (P_2 - c_{\text{H}}) + \Pr(\text{H}) \big(1 - \Pr_i(\text{stay} \mid \text{H})\big) \, m^{rec} + (1 - \Pr(\text{H})) \, (P_2 - c_{\text{S}}) \quad (78)$$

and the profit per potential enrollee is $\Pi = \int \Pr_i(\text{join}) \left[(P_1 - c_1) + \beta M_{2,i}\right] di$. Among the joiners, $\Pr(\text{stay} \mid \text{H}, \text{join}) = \int \Pr_i(\text{join}) \Pr_i(\text{stay} \mid \text{H}) \, di \,/\, \Pr(\text{join})$ is the share of healthy joiners who stay. It falls with P_2 and depends on $\tilde{P}$ through who joined.

The period-2 payment follows the same rule in both markets. Under market power, maximizing Π, and under perfect competition, where competition selects the zero-profit contract with the lowest $\tilde{P}$, the period-2 payment satisfies

$$P_2 - c_{\text{H}} - m^{rec} = \frac{\Pr(\text{H}) \Pr(\text{stay} \mid \text{H}, \text{join}) + (1 - \Pr(\text{H})) - \nu}{\Pr(\text{H}) \left(- \partial \Pr(\text{stay} \mid \text{H}, \text{join}) / \partial P_2 \right)} \quad (79)$$

where the response of staying to P_2 holds fixed who joined, and where under perfect competition $m^{rec} = 0$ and $P^{new} = c_{\text{H}}$. The left side is the margin at stake at the lapsation constraint: what the insurer earns on a retained healthy enrollee, net of what it keeps if that enrollee leaves. The numerator is the part of P_2 not taken into account at joining, the mass that pays P_2 minus the weight joiners give it. The denominator is the response of the healthy to P_2. Both are taken over the joiners, so the price that attracts joiners chooses the pool from which P_2 is collected.

The introduction price differs. Under perfect competition, zero profit on the pool attracted

gives

$$P_1 - c_1 = -\beta\, M_2^{\text{comp}}, \qquad M_2^{\text{comp}} = \frac{\int \Pr_i(\text{join})\, M_{2,i}\, di}{\Pr(\text{join})} \tag{80}$$

the period-2 margin of the average joiner. Under market power, the condition for P_1 gives

$$P_1 - c_1 = \frac{\Pr(\text{join})}{-\partial \Pr(\text{join})/\partial P_1} - \beta\, M_2^{\text{marg}}, \qquad M_2^{\text{marg}} = \frac{\int \frac{\partial \Pr_i(\text{join})}{\partial P_1}\, M_{2,i}\, di}{\int \frac{\partial \Pr_i(\text{join})}{\partial P_1}\, di} \tag{81}$$

the joining markup minus the period-2 margin of the joiners the price moves. This is the two-period version of $p_j = AC_j^{PV}$ against p_j = marginal cost + markup in (12) and (13): perfect competition returns to joiners the later margin of the average joiner, market power the later margin of the marginal joiner. When the joiners the price moves are those most likely to leave when healthy, and the margin on a retained healthy enrollee exceeds m^{rec}, M_2^{marg} is below the period-2 margin of the average joiner: market power passes back less of the later collection into the introduction price, what represents the selection wedge across ages.

If $\nu = 1$, (79) holds at any $P_2 \leq c_{\text{H}}$: no healthy enrollee leaves, and the timing of payments below c_{H} does not matter. At $P_2 = c_{\text{H}}$, the smallest collection in period 1, (80) gives

$$P_1 - c_1 = \beta\,(1 - \Pr(\text{H}))\,(c_{\text{S}} - c_{\text{H}}) > 0 : \tag{82}$$

the period-2 losses on the sick are collected in period 1, before health is known, because afterwards the healthy will not pay more than their alternative. This is full front-loading. With $\nu < 1$, for example because borrowing constraints make paying early costly for the young, competitive contracts collect less in period 1 and more in period 2, at the cost of some healthy enrollees leaving: the partial front-loading of Hendel and Lizzeri (2003). Under perfect competition the lapsation constraint is a feasibility condition: it decides which payment paths break even on the pool they attract.

Comparing front-loading. Let $m_1 = P_1 - c_1$ be the margin collected before health is known and M_2, the average of $M_{2,i}$ over joiners as in (80), the margin collected after, including the margin recaptured on the healthy who leave. The superscripts comp and ncomp denote perfect competition and market power, and M_2^{marg}, defined in (81), is the period-2 margin of the joiners the price moves under market power. Market power front-loads less than perfect competition if it collects less before health is known and more after, and more if the reverse holds. If it collects more in both periods, the two are not ranked.

By (80) and (81), with $\text{mkup} = \Pr(\text{join})/(-\partial \Pr(\text{join})/\partial P_1)$ the joining markup,

$$m_1^{\text{ncomp}} - m_1^{\text{comp}} = \text{mkup} - \beta\big(M_2^{\text{marg}} - M_2^{\text{comp}}\big). \tag{83}$$

Market power earns a non-negative margin on its joiners, so $m_1^{\text{ncomp}} + \beta M_2^{\text{ncomp}} \geq 0 = m_1^{\text{comp}} + \beta M_2^{\text{comp}}$, and collecting less before health is known implies collecting more after. Market power

therefore front-loads less exactly when

$$\beta\big(M_2^{\text{marg}} - M_2^{\text{comp}}\big) \geq \text{mkup}, \tag{84}$$

and more exactly when $M_2^{\text{ncomp}} \leq M_2^{\text{comp}}$. The joining markup is small when joining responds strongly to the price, as when an outside option contests every entrant. The period-2 margins follow from (79): joining depends only on $\tilde{P}$, so for a given $\tilde{P}$ the joiners are fixed and both market structures choose P_2 to maximize $M_2 - \nu P_2$. A channel that raises $M_2 - \nu P_2$ at every P_2 and does not lower P_2 therefore raises M_2^{ncomp} above M_2^{comp}. When all joiners behave alike, $M_2^{\text{marg}} = M_2^{\text{ncomp}}$. With $\nu < 1$, the ranking depends on the channel.

(o) *No channel.* If all joiners behave alike, alternatives are priced at cost and there is no recapture, $M_2 - \nu P_2$ is the same in both markets, so P_2 and M_2 are the same and m_1^{ncomp} exceeds m_1^{comp} by mkup: market power front-loads more.

(i) *Recapture.* Inertia toward the insurer lets it keep a margin $m^{rec} > 0$ on its healthy leavers, even when rivals price at cost. This adds $\Pr(\text{H})\big(1-\Pr(\text{stay} \mid \text{H}, \text{join})\big)m^{rec}$ to $M_2 - \nu P_2$, an amount that rises with P_2, so P_2 rises and M_2^{ncomp} exceeds M_2^{comp} by at least this amount evaluated at the competitive P_2. With joiners who behave alike, a sufficient condition for market power to front-load less is

$$\beta\ \Pr(\text{H})\ \big(1 - \Pr(\text{stay} \mid \text{H}, \text{join})^{\text{comp}}\big)\ m^{rec}\ \geq\ \text{mkup} \tag{85}$$

the discounted margin recaptured on the healthy who leave at the competitive price exceeds the joining markup. Otherwise market power collects more in both periods. It never front-loads more.

(ii) *Alternatives priced above cost.* When rival new plans carry a margin, $P^{new} > c_{\text{H}}$, more healthy enrollees stay at every P_2. If the share who stay falls linearly in the gap $P_2 - P^{new}$, P_2 rises by half the alternatives' margin, $(P^{new} - c_{\text{H}})/2$, and $M_2^{\text{ncomp}} > M_2^{\text{comp}}$. With joiners who behave alike, market power front-loads less exactly when $\beta(M_2^{\text{ncomp}} - M_2^{\text{comp}}) \geq \text{mkup}$, and otherwise collects more in both periods. It never front-loads more.

(iii) *Composition.* Assume the individuals most responsive to the price at joining are also the most likely to switch if they turn out healthy in period 2, with the same probability of being healthy in period 2 as everyone else. The markup raises $\tilde{P}$ and keeps them out, so the joiners switch less and respond less to P_2: P_2 rises and $M_2^{\text{ncomp}} > M_2^{\text{comp}}$. But P_1 is set for the joiners the price moves, who are the mobile ones, so $M_2^{\text{marg}} < M_2^{\text{ncomp}}$, which makes the condition harder to meet. Recapture adds $\Pr(\text{H})\Pr_i(\text{leave} \mid \text{H})\, m^{rec}$ to each joiner's period-2 margin, the most for the mobile joiners at the margin, and so raises M_2^{marg}. Composition combined with recapture is the configuration of our market: the joiners at the margin are mobile, and recapture is what keeps their later margin high.

In the cases where market power front-loads less, more of the losses on the sick are paid after health is known than under perfect competition, partly through margins a competitive market cannot sustain: margins recaptured in the insurer's other plans, and the higher later payments that rivals' margins make possible. Under perfect competition the lapsation constraint decides which paths are feasible. Under market power, it prices a positive margin the insurer wants to

keep, net of recapture, on a pool its own price has chosen.

In the market we study, $P_2 = rw \cdot P_1$ with rw set by regulation, so the insurer chooses one base price and the two conditions combine into the lifetime condition (12). Rating weights are not an instrument in our setting, but we still need to know how front-loading would look like if weights were and instrument. Own new plans can play the role of the lower continuation price of Hendel and Lizzeri (2003): they offer the healthy a lower price and turn their departure into recapture, raising m^{rec}. Removing new plans eliminates this margin, and the healthy who used to move now stay or leave to the outside option, so the enrollees at the lapsation margin are no longer the cheapest and the selection wedge closes.

The analysis above keeps the main assumption of Hendel and Lizzeri (2003) that individuals have the same expected cost when they join, and differences in health appear only in period 2, so at introduction a mobile joiner costs the same as an inert one. Suppose instead that the individuals most responsive to the price at joining are also healthier when they join, with a lower period-1 cost $c_{1,i}$, and the same probability of switching when healthy as everyone else. If health persists, they are also more likely to be healthy in period 2, $\Pr_i(\mathrm{H})$ higher, which raises their later margin and reinforces the effect below. The conditions for the introduction price become

$$P_1 - \bar{c}_1 = -\beta M_2^{\mathrm{comp}} \quad \text{(perfect competition)}, \qquad P_1 - \bar{c}_1^{\,\mathrm{marg}} = \mathrm{mkup} - \beta M_2^{\mathrm{marg}} \quad \text{(market power)}, \tag{86}$$

where $\bar{c}_1$ is the period-1 cost of the average joiner, $\bar{c}_1^{\,\mathrm{marg}}$ that of the joiners the price moves, and each joiner's period-2 margin uses their own $\Pr_i(\mathrm{H})$. With $m_1 = P_1 - \bar{c}_1$, market power front-loads less when

$$\big(\bar{c}_1 - \bar{c}_1^{\,\mathrm{marg}}\big) + \beta\big(M_2^{\mathrm{marg}} - M_2^{\mathrm{comp}}\big) \;\geq\; \mathrm{mkup}. \tag{87}$$

Healthier marginal joiners raise both terms on the left. They are cheaper now, so under market power P_1 follows a lower cost, the selection wedge of (13) at entry. They are less likely to fall sick, so their later margin is higher, through smaller expected losses $(1 - \Pr_i(\mathrm{H}))(c_{\mathrm{S}} - P_2)$. A lower introduction price then pays without recapture, because it attracts joiners who are cheap now and less likely to become costly later.