Dynamic Modeling of Adverse Selection in Peer-to-Peer Insurance Pools: Learning, Exits, and Market Unraveling

Syed Taha Qadri

June 2026

Abstract

Experience-rated pricing is usually viewed as a remedy for adverse selection in insurance markets. As loss histories accumulate, prices become more accurate. This paper shows that in a voluntary peer-to-peer insurance pool, the same logic can become destabilizing before learning has converged. Members have persistent risk types, but the pool sees only coarse entry information and noisy claims histories. Renewal premiums therefore respond to imperfect estimates of risk. The key object is the classification wedge: the gap between what the pool thinks a member will cost and the member's true expected loss. When histories are noisy, some members are temporarily overpriced and leave, while members whose risks are underestimated remain. The surviving pool can then become underpriced, weakening reserves and tightening future participation margins. The model delivers a local speed-limit diagnostic for experience rating, and shows that repricing is safer when histories are informative, entry scores are accurate, buffers are large, or subsidies dampen the household-paid wedge. Simulations show the full dynamic loop, while USDA crop-insurance data provide suggestive component-level corroboration.

JEL classification: D82, D86, G22, L14. **Keywords:** peer-to-peer insurance, adverse selection, experience rating, Bayesian learning, reclassification risk, dynamic contracts.

1 Introduction

The paper studies a peer-to-peer (P2P) insurance pool that covers low-frequency risks such as crop loss or flood damage. At entry, the pool has no member-specific loss history and relies only on observable covariates such as location, coverage choices, demographics, and prior market records. These variables provide an initial but coarse risk score. As claims accumulate, the pool updates each member's estimated risk and adjusts renewal premiums accordingly. Two households may share the same entry score but differ in true expected losses, while households with identical true risk may realize different early claim histories. If early signals from either the entry score or short loss history lead to higher renewal premiums before true risk is precisely learned, some members may be overpriced and exit. The remaining pool becomes increasingly composed of members whose risks are initially underestimated. What begins as rational Bayesian updating in pricing can thus generate endogenous selection and a participation shock.

P2P pools are an ideal setting to study this mechanism because their balance sheets make the feedback loop transparent: pricing, participation, risk sharing, and reserves are all interconnected. We allocate retained losses and reserve pass-through using expected-loss weights, so underwriting gains and losses feed back into renewal premiums through a transparent balance-sheet rule. This is consistent with actuarial ratemaking practice, where expected costs, capital costs, and additional funding sources can enter prospective renewal rates (Actuarial Standards Board, 2017, 2011), and with insurance-pricing evidence linking prices to capitalization and financial quality (Cummins and Danzon, 1997). In our model, when mispriced members leave, the remaining members inherit both the selection effect and the reserve consequences through the pass-through rule. The mechanism is broader than just P2P insurance. Automobile bonus-malus systems use claim or violation histories (Lemaire, 1995), while usage-based auto programs and claim-history databases also feed renewal pricing and underwriting.[1]

The core issue is not whether insurers should use information, but how quickly a voluntary pool should translate coarse entry scores and noisy early histories into renewal prices. Better entry covariates reduce residual type uncertainty; noisy early histories can instead make unlucky members look expensive before their true risk is known.

Our analysis delivers three findings. First, the classification wedge enters the stay margin directly, so overpriced members are more likely to leave. Second, wedge dispersion gives a local diagnostic for how fast experience rating can safely move prices. Third, simulations show how participation and reserves can deteriorate before learning catches up. The U.S. Department of Agriculture Risk Management Agency (RMA) crop-insurance data provide suggestive evidence on selected components of the mechanism.

Section 2 positions the contribution. Sections 3-5 build the model; Section 6 shows how frequency, severity, and the Signal-to-Noise Ratio (SNR) govern learnability; and Section 7 discusses contract and policy remedies. Section 8 reports the RMA evidence, Section 9 the simulations, and the appendices collect secondary results and proofs.

2 Related Literature

The benchmark is the lemons logic of Akerlof (1970) and the insurance-screening logic of Rothschild and Stiglitz (1976) and Stiglitz (1977). Empirical tests for asymmetric information reach mixed verdicts (Cohen and Siegelman, 2010): some find the predicted risk-coverage relationship (Puelz and Snow, 1994; Finkelstein and Poterba, 2004), others little or none (Chiappori and Salanie, 2000; Cardon and Hendel, 2001; Cawley and Philipson, 1999), and Finkelstein and McGarry (2006) show that private information about tastes can offset private information about risk. Einav, Finkelstein, and Cullen (2010) and Einav and Finkelstein (2011) connect selection (both "adverse" and "advantageous") to welfare. Cohen (2005) is especially relevant because she shows that learning and asymmetric information shape the coverage-risk relationship in automobile insurance. We take the classical framework as a starting point but emphasize a different source of selection pressure: the pool's classification error rather than only the policyholder's informational advantage.

A second strand studies repeated insurance and experience rating. Dionne (1983), Dionne and Lasserre (1985), and Cooper and Hayes (1987) show that repeated contracting can mitigate adverse selection asymptotically, a logic operationalized at scale in bonus-malus systems (Lemaire, 1995), with related repeated-insurance analyses by Hosios and Peters (1989). Dionne and Doherty (1994) add limited commitment and renegotiation to the model. These papers point in the opposite direction from our result: experience *cures* selection when histories reveal fixed types accurately, and the community-rating death spiral (Cutler and Reber, 1998; Cutler and Zeckhauser, 1998; Buchmueller and DiNardo, 2002) is precisely the case experience rating is meant to remedy. In an early finite-sample precursor, Gal and Landsberger (1988) note that the experience-rated premium is itself a risk: after an unlucky record it can exceed the value of coverage and induce lapse. They rule out that lapse by assuming long-term commitment; this paper instead keeps voluntary renewal.

[1] For example North Carolina's Safe Driver Incentive Plan, Progressive Snapshot, and LexisNexis C.L.U.E. Auto.

Proposition 2 turns the same finite-sample noise into a dynamic selection force that can destabilize a voluntary pool.

A third strand concerns reclassification risk and dynamic contracts. Cochrane (1995) and Pauly, Kunreuther, and Hirth (1995) study reclassification risk in health insurance, and Handel, Hendel, and Whinston (2015) quantify its tradeoff against adverse selection. Hendel and Lizzeri (2003) provide evidence on front-loaded life insurance under one-sided commitment. Here the type μ_i is fixed: the reclassification risk comes from noise in the posterior estimate $\hat{\mu}_{i,t}$, not from a true change in risk. Cutler, Lincoln, and Zeckhauser (2010) document backward-looking selection on realized past spending and adverse retention. Our survivor pool is shaped by the same dynamic-composition forces, but the sorting runs on the pool's estimation error, not the member's private information.

Actuarial credibility theory supplies the Bayesian updating logic behind the posterior formula in Section 3 (Whitney, 1918; Buhlmann, 1967; Buhlmann and Gisler, 2005; Ohlsson and Johansson, 2010). Ludkovski and Young (2010) combine Bayesian learning with experience-rated premiums, but under mandatory insurance, so classification errors generate no exit, survivor selection, or reserve feedback. These two precedents bracket our setting from opposite sides: Gal and Landsberger (1988) have voluntary lapse without informational bite (commitment neutralizes it), whereas Ludkovski and Young (2010) have the finite-sample informational bite without voluntariness (mandation removes exit). The wedge becomes a selection force precisely where both assumptions are dropped at once: voluntary participation, no commitment, and noisy finite-sample learning. This is the cell this paper occupies. Eling et al. (2024) study how richer data reshapes classification. Our contribution begins after that estimation step: once the credibility estimate is used as a participation-relevant price in a voluntary pool, its finite-sample error becomes a composition force.

Our entry score also connects the paper to the economics of risk classification. Hoy (1982) and Crocker and Snow (1986) analyze the welfare and efficiency effects of imperfect categorical classification, Crocker and Zhu (2021) characterize the efficiency of voluntary risk classification, and Braun, Häusle, and Thistle (2023) study classification under on-demand, dynamically chosen coverage. However, this strand is largely static and efficiency-focused. We add a finite-sample learning layer and ask how classification error, coarse at entry and noisy thereafter, drives voluntary exit and reserve dynamics rather than the classification efficiency frontier.

Within mutual and P2P insurance (Denuit and Dhaene, 2012; Denuit, 2019; Abdikerimova and Feng, 2022; Abdikerimova et al., 2024), the closest paper is Chen et al. (2024). They show how coarse observable classification can unravel voluntary mutual aid. Our residual-wedge mechanism is a dynamic analogue in which the pool updates noisy individual histories after entry. Similarly, Klein et al. (2023) study risk rating as a remedy for selection in voluntary health insurance; we show that noisy risk rating can itself become a selection force.

3 Environment

Household i has a persistent expected-loss type $\mu_i > 0$. Time is discrete, with coverage periods $t = 1, \ldots, T$. Coverage period t is written at the renewal date $t-1$. At renewal, the pool observes the history $\mathcal{F}_{t-1}$, sets the premium paid at date $t-1$ and the sharing rule for period-t coverage, and households decide whether to participate. During period t, losses realize, indemnities are paid, and the end-of-period reserve becomes R_t. The period-t signal is added to $\mathcal{F}_t$ and is used at the next renewal.

The pool is an accounting and rule-setting entity that collects contributions, pays claims, holds reserves, and applies the announced pricing rule. It has capacity for at most N active members and prices according to the announced administered rule rather than for profit. Our object is the stability of this rule rather than a competitive or monopoly pricing equilibrium.

At entry, the pool observes covariates Z_i but no member-specific loss history. Renewal is one-sided, so conditional on eligibility, the pool cannot selectively deny renewal after adverse realizations; it can only update premiums through the announced rule. Members decide whether to remain according to their participation constraint. Vacancies created by exiting members may be filled by new applicants drawn from the entry distribution. Entrants arrive with fresh covariates and no member-specific history, and they join only if their own participation condition holds. Let

$$m_i \equiv m(Z_i) = \mathbb{E}[\mu_i \mid Z_i] \tag{1}$$

be the entry risk score. True expected loss decomposes as

$$\mu_i = m_i + \delta_i, \qquad \delta_i \mid Z_i \sim \mathcal{N}(0, \sigma_0^2). \tag{2}$$

Thus δ_i is the residual type unexplained by entry covariates, and σ_0^2 is residual type variance after entry classification. More predictive covariates lower σ_0^2; coarse covariates leave a wide residual distribution and can overprice some members even before claims histories accrue. The no-covariate benchmark is the special case $m_i = \bar{\mu}_0$ for all i. The Gaussian residual specification is used as a local analytic benchmark around positive expected losses.

3.1 Losses and Signals

Household i's period loss at t is

$$X_{i,t} = \mu_i + \varepsilon_{i,t}, \qquad \mathbb{E}[\varepsilon_{i,t}] = 0, \quad \mathrm{Var}(\varepsilon_{i,t}) = \sigma_X^2, \quad \mathrm{Cov}(\varepsilon_{i,t}, \varepsilon_{j,t}) = \rho_\varepsilon \sigma_X^2 \text{ for } i \neq j. \tag{3}$$

The loss variance σ_X^2 governs risk sharing. A high σ_X^2 means the pool creates large insurance value by smoothing idiosyncratic losses. After entry, the pool learns residual type through a signal

$$S_{i,t} = \mu_i + \zeta_{i,t}, \qquad \zeta_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \tag{4}$$

with $\{\zeta_{i,\tau}\}$ conditionally independent over time given μ_i and independent of the prior draw. The signal may equal the realized loss or a processed claims signal. Keeping σ_X and σ_ε distinct lets the model separate the value of pooling volatile losses from the precision with which claims histories reveal a member's true risk. Because entry covariates have already been used to form m_i, the relevant signal-to-noise ratio is formed by the residual type:

$$\mathsf{SNR} = \frac{\sigma_0^2}{\sigma_\varepsilon^2}. \tag{5}$$

It compares the remaining within-score type variance to signal noise. High SNR means residual types are learned quickly; low SNR means high residual risk is hard to distinguish from bad luck.

3.2 Bayesian Updating

At information date t, after $T_{i,t}$ observed signals through date t, the Gaussian posterior mean is the normal-normal credibility estimator in the tradition of Buhlmann (1967):

$$\hat{\mu}_{i,t} = \frac{\sigma_0^2 \sum_{\tau \le t} S_{i,\tau} + \sigma_\varepsilon^2 m_i}{T_{i,t}\sigma_0^2 + \sigma_\varepsilon^2}. \tag{6}$$

For $T_{i,t} > 0$, let $\bar{S}_{i,t} = T_{i,t}^{-1} \sum_{\tau \leq t} S_{i,\tau}$. Then

$$\hat{\mu}_{i,t} = (1 - z_{i,t})m_i + z_{i,t}\bar{S}_{i,t}, \qquad z_{i,t} = \frac{T_{i,t}\mathsf{SNR}}{1 + T_{i,t}\mathsf{SNR}}. \tag{7}$$

For $T_{i,t} = 0$, set $z_{i,t} = 0$ and $\hat{\mu}_{i,t} = m_i$. Thus Bayesian updating already shrinks sparse or noisy histories toward the entry score. The statistical credibility weight $z_{i,t}$ governs how much history enters the posterior. The experience-rating intensity χ (introduced in Section 5) governs how much the posterior enters the premium.

Definition 1 (Classification wedge)**.** For agent i at date t, the *classification wedge* is the deviation of the pool's current posterior mean from the true expected loss:

$$\xi_{i,t} \;\coloneqq\; \hat{\mu}_{i,t} - \mu_i. \tag{8}$$

At entry, before any member-specific loss history is observed, this is the coarse-entry wedge $\xi_{i,0} = m_i - \mu_i = -\delta_i$. For $T_{i,t} > 0$ it is the posterior wedge after learning from claims histories.

For $T_{i,t} > 0$, substituting the signal equation (4) into (7) gives

$$\xi_{i,t} = (1 - z_{i,t})(m_i - \mu_i) + z_{i,t}\bar{\zeta}_{i,t} = -(1 - z_{i,t})\delta_i + z_{i,t}\bar{\zeta}_{i,t}, \qquad \bar{\zeta}_{i,t} = T_{i,t}^{-1} \sum_{\tau \leq t} \zeta_{i,\tau}.$$

Shrinkage therefore reduces the entry wedge but does not eliminate it unless $T_{i,t}\mathsf{SNR}$ is large. The residual posterior variance is

$$\sigma^2_{T_{i,t}} = \frac{\sigma_0^2 \sigma_\varepsilon^2}{T_{i,t}\sigma_0^2 + \sigma_\varepsilon^2} = \frac{\sigma_0^2}{1 + T_{i,t}\mathsf{SNR}}. \tag{9}$$

This is the statistical engine of the paper. Better entry covariates lower the residual variance σ_0^2; longer loss histories lower the wedge dispersion through $T_{i,t}\mathsf{SNR}$. In low-frequency, high-severity lines, each time period adds little statistical information, so $T_{i,t}\mathsf{SNR}$ remains small even after a long observed history. This is the rationale behind repeated-insurance learning arguments: with sufficiently long histories, statistical tests can eventually separate risk types (Dionne and Lasserre, 1985), while finite-horizon experience-rated contracts might still not attain the first best (Cooper and Hayes, 1987). Appendix B.1 derives (6)-(9).

Lemma 1 (Wedge variance)**.** *Conditional on entry covariates and under the Gaussian benchmark,*

$$\mathrm{Var}(\xi_{i,t}) = \frac{\sigma_0^2}{1 + T_{i,t}\mathsf{SNR}}.$$

Thus wedge dispersion is increasing in residual type variance, increasing in signal noise, and decreasing in the length of the observed history.

Proof. See Appendix B.1.

Lemma 1 is a fixed-member learning result. In a closed cohort, each surviving member's history grows, so the wedge is eventually learned away if the pool survives long enough. Replenishing pool members changes this as new entrants arrive at renewal with $T_{i,t-1} = 0$, so the pool is again exposed to fresh classification error.

3.3 Preferences, Wealth, and Participation Value

Households are evaluated with a mean-variance certainty equivalent, used here as a tractable second-order representation of expected utility for the one-period renewal margin. At date $t-1$, household i compares coverage for period t with non-participation. If uninsured, it bears its period-t loss $X_{i,t}$ and pays the net exit cost $c_{i,t}$ (e.g. a cancellation fee). If insured, it pays premium $\pi_{i,t-1}$, receives full indemnity, and remains exposed to its rule-based share of retained aggregate pool risk. Let A_t be the active set chosen at renewal for coverage period t, and let $n_t = |A_t|$. The model uses expected-loss weights:

$$g_{i,t}(L_t) = \theta_{i,t-1} L_t, \qquad \theta_{i,t-1} = \frac{\hat{\mu}_{i,t-1}}{M_{t-1}}, \qquad M_{t-1} = \sum_{j \in A_t} \hat{\mu}_{j,t-1}. \tag{10}$$

We maintain $\hat{\mu}_{i,t-1} \geq 0$ and $M_{t-1} > 0$ on the relevant state space. Here $L_t = \sum_{j \in A_t} X_{j,t}$ is total realized loss in the coverage period. The weights are fixed at renewal, sum to one, and allocate retained losses as well as reserve pass-through below. Equal sharing is the homogeneous special case $\hat{\mu}_{i,t-1} = M_{t-1}/n_t$, where $\theta_{i,t-1} = 1/n_t$ for every active member.

The pool reserve evolves according to

$$R_t = (1 + r_f)\left(R_{t-1} + \sum_{i \in A_t} \pi_{i,t-1}\right) - \sum_{i \in A_t} X_{i,t}. \tag{11}$$

Here $r_f \geq 0$ is the per-period risk-free return earned on reserves and collected premiums. The reserve is a book-equity state. Members are not assessed between renewal dates. Instead, inherited reserve deficits or surpluses enter the next renewal premium below. With mean-variance preferences the certainty equivalent of any net position W is $\mathrm{CE} = \mathbb{E}[W] - \frac{1}{2}\gamma_i \mathrm{Var}(W)$. If household i does not participate, it bears its own loss in full and pays the net exit cost:

$$\mathrm{CE}^{\mathrm{out}}_{i,t-1} = -\mu_i - \tfrac{1}{2}\gamma_i \sigma_X^2 - c_{i,t}. \tag{12}$$

If it participates, the premium covers expected retained pool losses and full indemnity removes its own idiosyncratic loss. The remaining one-period risk inside the mutual contract is the member's fixed share of aggregate retained-loss risk.

$$\mathrm{CE}^{\mathrm{in}}_{i,t-1} = -\pi_{i,t-1} - \tfrac{1}{2}\gamma_i \, \mathrm{Var}_{t-1}(\theta_{i,t-1} L_t). \tag{13}$$

Using the Gaussian aggregate variance from (3),

$$\mathrm{Var}_{t-1}(L_t) = \sigma_X^2 \left[n_t + \rho_\varepsilon n_t (n_t - 1)\right],$$

the value of pooling is the difference in the two risk penalties,

$$V_{i,t-1} = \tfrac{1}{2}\gamma_i \left\{ \sigma_X^2 - \theta_{i,t-1}^2 \sigma_X^2 \left[n_t + \rho_\varepsilon n_t (n_t - 1)\right] \right\}. \tag{14}$$

When $\theta_{i,t-1} = 1/n_t$, this reduces to the symmetric-sharing benchmark

$$V_{i,t-1} = \tfrac{1}{2}\gamma_i \sigma_X^2 (1 - \rho_\varepsilon)(1 - 1/n_t).$$

This is the fundamental small-pool diversification effect that motivates P2P arrangements (Denuit and Dhaene, 2012; Denuit, 2019; Abdikerimova and Feng, 2022). It falls as correlation rises because common shocks cannot be diversified. Participation is voluntary unless the policy regime says otherwise, and household i stays for period t exactly when $\mathrm{CE}^{\mathrm{in}}_{i,t-1} \geq \mathrm{CE}^{\mathrm{out}}_{i,t-1}$. Substituting (12)-(13) and using (14) gives

$$\mu_i + V_{i,t-1} + c_{i,t} \geq \pi_{i,t-1}. \tag{15}$$

The left side is the value of staying, namely indemnity, pooling value, and avoided net exit cost. The right side is the premium. Rearranging,

$$V_{i,t-1} + c_{i,t} \geq \pi_{i,t-1} - \mu_i.$$

The object on the right is the household's price distortion relative to true actuarial cost.

Assumption 1 (Myopic participation)**.** Households evaluate the one-period stay condition in (15) at renewal, using the premium paid at $t-1$, the reserve state R_{t-1}, and the period-t net exit cost. They do not solve a dynamic optimal-stopping problem in the baseline model.

This is a renewal-margin simplification. Annual insurance contracts are repriced one period at a time, and the relevant decision at renewal is whether the current premium, reserve state, pooling value, and outside option make participation worthwhile. Any smooth continuation value of remaining in the pool can be folded into the reduced-form net exit cost $c_{i,t}$. A fully dynamic stopping model with forward-looking re-entry and learning incentives is left for future work.

The results that follow separate four claims. First, with true-type pricing, classification wedges are absent and voluntary stability depends only on loading and participation buffers. Second, with experience rating, the renewal-date wedge enters the stay margin as $-\beta\xi_{i,t-1}$, so overpriced members are more likely to exit. Third, this selection makes survivors negatively selected on wedges and creates an actuarial deficit when survivor wedges are negative on average. Fourth, the deficit becomes a spiral only when it exceeds the loading, investment, and reserve buffers.

4 Benchmarks and Entry-Score Pricing

Two benchmarks isolate the mechanism. Under symmetric information, the pool observes μ_i and charges $\pi^{FB}_{i,t-1} = (1+\ell)\mu_i$, the stay condition is

$$V_{i,t-1} + c_{i,t} \geq \ell\mu_i.$$

There is no classification wedge, so exits are not selected on pricing errors. Uniform pooling has the familiar opposite problem: a common price can induce low-risk members to leave and raise the average risk of the remaining pool. That classical composition feedback does not require Bayesian misclassification.

Entry risk classification sits between these two cases. With no member-specific loss history, the pool can only price on the entry score:

$$\pi^{Z}_{i,0} = (1-\chi)(1+\ell)\bar{m}_0 + \chi(1+\ell)m_i, \qquad \bar{m}_0 = \frac{1}{N}\sum_{j=1}^{N} m_j. \tag{16}$$

Good covariates lower σ_0^2 and make $m_i - \mu_i$ small. Coarse covariates leave $m_i - \mu_i = \xi_{i,0}$, so members can be overpriced before any loss history is observed. Bayesian experience rating adds the dynamic layer: as histories arrive, m_i is replaced at later renewals by the posterior $\hat{\mu}_{i,t-1}$.

5 Bayesian Experience Rating

We now formalize the design tradeoff introduced in the Introduction. At renewal date $t-1$, experience rating sets the premium paid at $t-1$ by blending a common pool component $\bar{\hat{\mu}}_{t-1}$ with an individual posterior risk estimate $\hat{\mu}_{i,t-1}$, which combines entry covariates m_i with member-specific signal history:

$$\pi^E_{i,t-1} = (1-\chi)(1+\ell)\bar{\hat{\mu}}_{t-1} + \chi(1+\ell)\hat{\mu}_{i,t-1}, \qquad \chi \in [0,1], \tag{17}$$

where

$$\bar{\hat{\mu}}_{t-1} = \frac{1}{n_t}\sum_{j\in A_t}\hat{\mu}_{j,t-1}.$$

The platform announces the pricing formula before participation; the displayed common component is the value in the renewal fixed point generated by the active set A_t. The parameter χ is the experience-rating intensity: $\chi = 0$ gives pure pooling and $\chi = 1$ gives full individual pricing. At entry, $\hat{\mu}_{i,0} = m_i$, so χ controls pass-through of entry covariates; after histories accrue, it controls pass-through of the Bayesian posterior. It is distinct from the statistical credibility weight $z_{i,t}$ in (7): $z_{i,t}$ governs how much history enters the posterior, while χ governs how much the posterior enters the premium. Let

$$\beta = \chi(1+\ell). \tag{18}$$

This is the slope of the premium with respect to the individual posterior risk estimate $\hat{\mu}_{i,t-1}$. The actual renewal premium also passes through a fraction of the inherited reserve state:

$$\pi_{i,t-1} = \pi^E_{i,t-1} - \lambda_R\theta_{i,t-1}R_{t-1}, \qquad \lambda_R \in [0,1]. \tag{19}$$

If $R_{t-1} < 0$, the term is a reserve-recovery surcharge; if $R_{t-1} > 0$, it is a premium credit. The case $\lambda_R = 1$ passes the inherited reserve state through immediately.

Substituting (17) and (19) into the participation condition and using $\hat{\mu}_{i,t-1} = \mu_i + \xi_{i,t-1}$ gives

$$(1-\beta)\mu_i - \beta\xi_{i,t-1} + V_{i,t-1} + c_{i,t} - (1-\chi)(1+\ell)\bar{\hat{\mu}}_{t-1} + \lambda_R\theta_{i,t-1}R_{t-1} \geq 0. \tag{20}$$

The common term $\bar{\hat{\mu}}_{t-1}$ contains household i's own posterior; for the individual participation geometry it is read in the price-taking (large-pool) sense, equivalently replaced by the leave-one-out average $\bar{\hat{\mu}}_{-i,t-1} = \frac{1}{n_t-1}\sum_{j\in A_t\setminus\{i\}}\hat{\mu}_{j,t-1}$. [2]

For $\beta \leq 1$, a positive wedge directly lowers the stay margin. Conditional on buffers, the marginal exiters are therefore the members whose risks the pool overestimates at renewal. The analysis below stays in this subcritical region.

5.1 Misclassification Selection

Relative to the entry-score and pooling benchmarks, the new term in (20) is $-\beta\xi_{i,t-1}$. A positive classification wedge ($\xi_{i,t-1} > 0$) means that the household is priced above its true expected cost, reducing its incentive to remain in the pool. With zero member-specific history this is coarse entry classification; for $T > 0$ it is also finite-history posterior error. This is close to Hirshleifer's point that pricing on individual information can undermine mutually beneficial risk sharing (Hirshleifer, 1971). In catastrophe markets, Mulder (2019) documents FEMA flood-map revisions as a learning shock

[2] The omitted own component is $O(1/n_t)$ and shares the sign of the direct wedge term $-\beta\xi_{i,t-1}$, so the affine-in-μ_i slope, the lower-cutoff structure, and the sign of the selection effect are unaffected; under this reading the buffer $H_{i,t-1}$ in Lemma 2 no longer depends on μ_i through the common term. For the small pools that motivate the analysis the $O(1/n_t)$ term is not negligible.

under which lower-risk homes lapse and higher-risk homes retain coverage. Our channel is distinct but related. The pool's own residual classification error is the driver, so *more* precise information stabilizes rather than erodes the pool.

Lemma 2 (Wedge selection)**.** *Suppose $\chi > 0$ and $\beta \in (0, 1]$. The renewal margin can be written as*

$$M_{i,t-1} = (1-\beta)\mu_i - \beta\xi_{i,t-1} + H_{i,t-1},$$

where

$$H_{i,t-1} = V_{i,t-1} + c_{i,t} - (1-\chi)(1+\ell)\bar{\hat{\mu}}_{t-1} + \lambda_R\theta_{i,t-1}R_{t-1}.$$

Household i stays for period t iff $M_{i,t-1} \geq 0$. Assume the Gaussian benchmark, a continuous positive conditional density around the renewal cutoff, and positive stay and exit mass. Also assume that, after accounting for the dependence of $V_{i,t-1}$ and reserve pass-through on posterior risk, the stay indicator is weakly decreasing in the wedge $\xi_{i,t-1}$, with strict decrease on a set of positive measure, and the stay probability is weakly increasing in residual type $\delta_i = \mu_i - m_i$ within entry-score cells. Then

$$\mathbb{E}[\xi_{i,t-1} \mid i \in A_t] < 0.$$

Remark 1 (Reserve pass-through and survivor selection). The selection result has two forces. First, holding true type fixed, a bad signal raises $\xi_{i,t-1}$, raises the renewal premium, and makes exit more likely. This removes positive-wedge members and pushes survivor wedges downward. Second, while learning is incomplete, high residual-risk members tend to have negative wedges because the pool has not fully learned their type, so they are more likely to remain. Reserve pass-through adds a qualification: because reserve credits or charges are allocated using $\theta_{i,t-1}$, a large inherited reserve state can affect relative premiums across estimated-risk groups. The lemma assumes these indirect effects do not overturn the direct wedge-selection force.

Proof. See Appendix B.2.

The pool's expected actuarial deficit before investment income is

$$\Delta_t = \sum_{i \in A_t} \left(\mu_i - \frac{\pi^E_{i,t-1}}{1+\ell} \right). \tag{21}$$

Under the uncapped raw experience-rating premium in (17), this decomposes as

$$\Delta_t = -\chi \sum_{i \in A_t} \xi_{i,t-1} + (1-\chi) \sum_{i \in A_t} (\mu_i - \bar{\hat{\mu}}_{t-1}). \tag{22}$$

Because the actually charged premium $\pi_{i,t-1}$ includes reserve pass-through, the financing deficit differs from the raw misclassification deficit. Under (19),

$$\sum_{i \in A_t} \left(\mu_i - \frac{\pi_{i,t-1}}{1+\ell} \right) = \Delta_t + \frac{\lambda_R R_{t-1}}{1+\ell}. \tag{23}$$

Thus reserve pass-through creates a negative feedback effect: a negative reserve reduces the current financing deficit by raising renewal premiums, while a positive reserve is gradually returned through premium credits.

The two terms in (22) are mechanically linked for a fixed active set. Since $\sum_{i \in A_t}(\mu_i - \bar{\hat{\mu}}_{t-1}) = -\sum_{i \in A_t} \xi_{i,t-1}$, the uncapped aggregate deficit simplifies to

$$\Delta_t = -\sum_{i \in A_t} \xi_{i,t-1} \tag{24}$$

Proposition 1 (The two roles of the experience-rating intensity). *For the uncapped raw premium and a* fixed *active set, the aggregate deficit* $\Delta_t = -\sum_{i\in A_t} \xi_{i,t-1}$ *is independent of the experience-rating intensity* χ*. The intensity therefore acts directly through the* selection *margin:* $\beta = \chi(1+\ell)$ *scales how far the wedge moves the participation cutoff and hence which members are in* A_t*. Reserve pass-through is governed separately by* λ_R *and* $\theta_{i,t-1}$*;* χ *affects it only indirectly by changing the active set. Raising* χ *does not, by itself, enlarge the misclassification deficit on a given pool; it sharpens who stays.*

This separation is easy to misread as a contradiction. The parameter χ is the key design lever yet drops out of the fixed-set deficit, so we state it explicitly. Its first-order effect is selection: higher χ makes classification wedges more participation-relevant and changes who remains. Negative average survivor wedges imply undercharging relative to true expected cost.

A positive deficit does not by itself lower expected reserves. The reserve law (11) carries loaded premiums and a gross investment return $1 + r_f$. Taking expectations in (11) (using $\mathbb{E}[X_{i,t}] = \mu_i$ and $\sum_{i\in A_t} \pi_{i,t-1} = (1+\ell)(\sum_{i\in A_t} \mu_i - \Delta_t) - \lambda_R R_{t-1}$) gives

$$\mathbb{E}[R_t \mid A_t, \mathcal{F}_{t-1}] - R_{t-1} = \big[r_f - (1+r_f)\lambda_R\big] R_{t-1} + \big[(1+r_f)(1+\ell) - 1\big] \sum_{i\in A_t} \mu_i - (1+r_f)(1+\ell)\, \Delta_t. \quad (25)$$

Expected reserves therefore deteriorate only when the misclassification deficit exceeds the loading-and-investment buffer

$$\Delta_t^* \equiv \frac{\big[r_f - (1+r_f)\lambda_R\big] R_{t-1} + \big((1+r_f)(1+\ell) - 1\big) \sum_{i\in A_t} \mu_i}{(1+r_f)(1+\ell)},$$

The threshold Δ_t^* is the absorbable deficit: the largest misclassification deficit that can be covered by loadings, investment income, and reserve recovery before expected reserves fall. If $\Delta_t < \Delta_t^*$, underpricing exists but is absorbed by the buffer. If $\Delta_t > \Delta_t^*$, the underpricing is large enough to deteriorate expected reserves. In the baseline case where reserve pass-through dominates the interest earned on reserves, an inherited deficit raises current premiums and increases this recovery buffer, whereas a positive reserve is partly returned through premium credits and therefore reduces the buffer. The next definition isolates the paths in which the selection force is large enough to affect reserves; it is not implied by wedge selection alone.

Definition 2 (Misclassification spiral). A dynamic path is a misclassification spiral on a set of dates if:

(i) stayers are negatively selected on classification wedges, $\mathbb{E}[\xi_{i,t-1} \mid i \in A_t] < 0$;

(ii) the aggregate deficit is positive:

$$\Delta_t = -\sum_{i\in A_t} \xi_{i,t-1} > 0.$$

(iii) the deficit exceeds the loading-and-investment buffer, $\Delta_t > \Delta_t^*$, so expected reserves deteriorate from R_{t-1} to R_t absent external transfers; and

(iv) the lower reserve state weakly tightens future participation margins through reserve pass-through at the next renewal.

The spiral may end in complete collapse, partial collapse, or a low-participation negatively selected steady state.

Proposition 2 (Classification noise can create selection pressure)**.** *Suppose* $\chi > 0$ *and the conditions of Lemma 2 hold. Whenever wedge-driven exits occur, the remaining pool is negatively selected on* $\xi_{i,t-1}$ *and the expected misclassification component of* (22) *is positive.*

Proof. See Appendix B.2.

5.2 When Experience Rating Works

Let the participation buffer excluding the wedge be

$$\begin{aligned} B_{i,t-1} = (1-\beta)\mu_i + V_{i,t-1} + c_{i,t} \\ - (1-\chi)(1+\ell)\bar{\hat{\mu}}_{t-1} + \lambda_R \theta_{i,t-1} R_{t-1}, \end{aligned} \tag{26}$$

so the household stays when $B_{i,t-1} \geq \beta \xi_{i,t-1}$. The relevant boundary is not experience rating itself, but the size of the classification pricing error after it has been translated into a household-paid premium and compared with the participation buffer. For $B_{i,t-1} > 0$, define the *reclassification-pressure ratio* as:

$$\Psi_{i,t-1} \equiv \frac{\beta\, \sigma_{\xi,i,t-1}}{B_{i,t-1}}, \qquad \sigma_{\xi,i,t-1} \equiv \frac{\sigma_0}{\sqrt{1 + T_{i,t-1}\mathsf{SNR}}}. \tag{27}$$

The numerator is the typical classification wedge after applying the premium slope; the denominator is the household's non-wedge reason to remain in the pool. When $\Psi_{i,t-1}$ is small, the usual classification error is too small to move the participation margin. When it is close to one, ordinary entry-score or posterior noise is economically capable of inducing exit.

Proposition 3 (A sufficient stability condition)**.** *Fix a renewal date* $t-1$*,* $\beta > 0$*, and* $\alpha < 1/2$*, and suppose* $B_{i,t-1} \geq \underline{B} > 0$ *for all renewing households. If each household's renewal-date wedge distribution satisfies* $\Pr\{\xi_{i,t-1} > \underline{B}/\beta\} \leq \alpha$*, the wedge-driven exit probability is at most* α*. Under the marginal Gaussian benchmark, a convenient ex-ante sufficient condition is*

$$\frac{\underline{B}}{\beta} \geq q_{1-\alpha} \frac{\sigma_0}{\sqrt{1 + T_{i,t-1}\mathsf{SNR}}}, \tag{28}$$

where $q_{1-\alpha}$ *is the standard normal quantile.*

Proof. See Appendix B.2.

Equivalently, (28) requires $\beta \sigma_{\xi,i,t-1}/\underline{B} \leq 1/q_{1-\alpha}$. This gives the paper's operational diagnostic: a sufficient local *speed limit* on experience-rating intensity. The diagnostic ratio is

$$\Psi_{i,t-1} = \frac{\beta \sigma_{\xi,i,t-1}}{B_{i,t-1}},$$

priced wedge dispersion relative to the participation buffer.

When $\beta = 0$, classification wedges do not affect individual premiums, so this exit channel is absent (though classical pooling selection may remain). A proportional subsidy generalizes this limit and is the relevant case for federally supported lines. If the household pays a fraction $1-\phi$ of the experience-rated premium, the household-paid wedge slope is

$$\beta_{\text{eff}} = (1-\phi)\,\beta = (1-\phi)\,\chi(1+\ell). \tag{29}$$

The stay condition has the same form,

$$\begin{aligned} B^{\phi}_{i,t-1} &\geq \beta_{\text{eff}} \xi_{i,t-1}, \\ B^{\phi}_{i,t-1} &= (1-\beta_{\text{eff}})\mu_i + V_{i,t-1} + c_{i,t} \\ &\quad - (1-\phi)(1-\chi)(1+\ell)\bar{\hat{\mu}}_{t-1} + (1-\phi)\lambda_R \theta_{i,t-1} R_{t-1}. \end{aligned}$$

Thus the diagnostic ratio becomes $\Psi^{\phi}_{i,t-1} = \beta_{\text{eff}}\sigma_{\xi,i,t-1}/B^{\phi}_{i,t-1}$. A larger subsidy lowers the household-paid wedge and raises the premium-level buffer, so the exit channel is attenuated. This is the lens for the heavily subsidized crop-insurance setting of Section 8.

Let $\underline{B}^{\phi}$ be a conservative lower bound on the subsidized buffer $B^{\phi}_{i,t-1}$ under the candidate rule. For a target tail probability α, the rule passes the local tail test only if

$$\beta_{\text{eff}}^{\max} = \frac{\underline{B}^{\phi}}{q_{1-\alpha}\sigma_{\xi,i,t-1}}, \qquad \beta_{\text{eff}} \le \beta_{\text{eff}}^{\max}.$$

Equivalently, if this buffer bound is held fixed when varying the experience-rating intensity,

$$\chi \le \chi^{\max} = \frac{\underline{B}^{\phi}}{(1-\phi)(1+\ell)q_{1-\alpha}\sigma_{\xi,i,t-1}} \quad (\phi < 1).$$

If the planned rule implies $\beta_{\text{eff}} > \beta_{\text{eff}}^{\max}$, the pool must slow repricing, reduce reserve pass-through, raise subsidies or reserves, or accept wedge-driven exit risk. Experience rating works when learning is precise, histories are long, the premium slope is moderate, subsidies are large, or participation buffers are large.

5.3 When Experience Rating Becomes Fragile

Experience rating becomes fragile only through a combination of margins. Noisy learning alone is not enough: wedges must be large enough to create exposed exit mass, exiters must be drawn from the positive-wedge tail, and the resulting undercharge must exceed the reserve or loading buffer. Let $\mathcal{E}_t$ denote the pre-renewal eligible set, and let

$$m_t \equiv \Pr_{i\in\mathcal{E}_t}\{0 \le B_{i,t-1} < \beta\xi_{i,t-1}\} \tag{30}$$

denote the wedge-exposed renewal mass: members who would stay absent the renewal-date wedge but leave once the wedge is priced. The condition $B_{i,t-1} \ge 0$ removes ordinary non-wedge exits from the diagnostic.

If $m_t > 0$, wedge selection is negative, and the resulting deficit exceeds Δ_t^*, expected reserves fall from R_{t-1} to R_t and future participation buffers tighten through (19). With sticky exit, the same conditions can generate a monotone decline until the pool reaches a minimum viable size, a solvency boundary, or a low-participation selected steady state. In a closed cohort this pressure fades only if the pool survives long enough for learning to catch up. With replenishment, endogenous exits can bring in zero-history entrants, so fresh classification errors can remain relevant without assuming any exogenous turnover.

5.4 The Three Links

The spiral can be decomposed into three links: classification wedges can induce exit, exit reshapes the survivor pool, and the resulting deficit feeds back into future prices or reserves (Figure 1). The empirical section organizes corroborative evidence around these links separately; the simulation illustrates the full loop.

Link (a) is Lemma 2: overpriced members leave faster. Link (b) is the resulting survivor composition: the remaining pool has a lower average wedge and is undercharged. Link (c) starts from the undercharge in (22) and feeds back through reserves via (19) and (25). Together $(a)\to(b)\to(c)$ is the closed cycle that makes Definition 2 dynamic. The partition lets us report what is and is not directly tested.

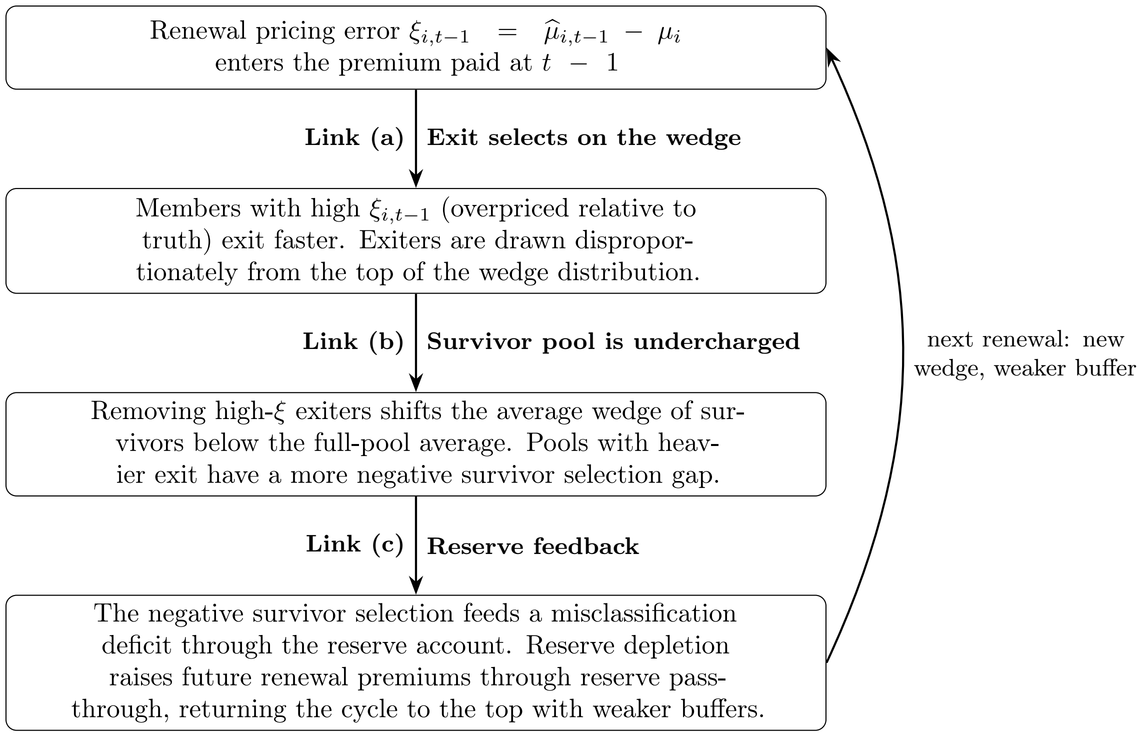

Figure 1: The misclassification spiral as three links.

6 Frequency, Severity, and Learnability

The Gaussian signal model is a reduced form for claim processes with frequency and severity. If annual losses are compound Poisson,

$$X_{i,t} = \sum_{k=1}^{N_{i,t}} Y_{i,k}, \qquad N_{i,t} \sim \text{Pois}(\lambda_i),$$

with severities iid, independent of the count, mean ν_i, and coefficient of variation cv, then $\mathbb{E}[X_{i,t}] = \lambda_i \nu_i = \mu_i$ and

$$\text{Var}(X_{i,t}) = \lambda_i \nu_i^2 (1 + cv^2). \tag{31}$$

Holding expected loss μ_i fixed while frequency varies ($\nu_i = \mu_i/\lambda_i$),

$$\text{Var}(X_{i,t}) = \frac{\mu_i^2(1 + cv^2)}{\lambda_i}. \tag{32}$$

Lower frequency therefore raises signal noise. The same expected loss has to be learned from fewer, larger, noisier events. The compound-Poisson reduction is standard in non-life pricing (Ohlsson and Johansson, 2010). When the annual loss is the signal, we map this into the Gaussian benchmark by setting the line-level signal variance σ_ε^2 to this conditional annual-loss variance, or a constant multiple of it.[3] The same increase in loss variance can also raise the renewal pooling value $V_{i,t-1}$

[3] This is a reduced-form normal approximation: the exact compound-Poisson distribution can be skewed and zero-heavy at low frequency. Because the closed-form benchmark uses a common σ_ε^2, the mapping is a homogeneous-line (representative-type) calculation; allowing type-specific variances preserves the frequency-severity comparative static with an i-specific SNR_i and turns the closed-form wedge distributions into mixtures over SNR_i rather than changing their direction.

in (14), so whether a low-frequency line destabilizes depends on whether the larger participation buffer offsets the noisier posterior price.

Holding residual type dispersion σ_0^2 fixed, $\mathsf{SNR} = \sigma_0^2/\sigma_\varepsilon^2$ is the *learnability index* of a line: with loss-derived signals it rises with frequency λ and falls with severity dispersion cv. The comparative static is stark. High-SNR lines (high-frequency, low-severity) are where experience rating learns quickly, misclassification is short-lived, and the instrument stabilizes the pool; low-SNR lines (low-frequency, high-severity, catastrophe-like) are where learning is slow, wedges stay large, and the same instrument can destabilize it. In the reclassification-pressure diagnostic, the loss-derived signal case gives

$$\Psi_{i,t-1} \propto \frac{\beta}{B_{i,t-1}} \frac{\sigma_0}{\sqrt{1 + T_{i,t-1}\sigma_0^2\lambda_i/[\mu_i^2(1+cv^2)]}},$$

up to the constant used to map compound-Poisson loss variance into σ_ε^2. Thus frequency and severity matter only through their effect on residual learnability and buffers: lower λ_i or higher cv raises classification-wedge dispersion, while a larger pooling value $V_{i,t-1}$ can offset it by raising $B_{i,t-1}$. This is the boundary the simulations of Section 9 trace and the empirics of Section 8 exploit through the subsidy-attenuated slope β_{eff}.

7 Contracts and Policy

7.1 Information Design: Entry Scores

Entry covariates are the first stabilizing instrument. Replacing Z_i by a richer signal vector Z_i' changes the residual variance from $\mathrm{Var}(\mu_i \mid Z_i)$ to $\mathrm{Var}(\mu_i \mid Z_i')$. Holding buffers fixed, the local tail test (28) improves whenever

$$\mathrm{Var}(\mu_i \mid Z_i') < \mathrm{Var}(\mu_i \mid Z_i),$$

because $\sigma_{\xi,i,t-1}$ falls at entry and throughout learning. But this is not a monotone argument for using every available variable. A covariate that is coarse, unstable, legally constrained, or weakly predictive may move prices more than it reduces residual wedge risk. The operational criterion is whether the new score lowers wedge-exposed mass,

$$\Pr\{0 \le B_{i,t-1} < \beta\xi_{i,t-1} \mid Z_i'\} < \Pr\{0 \le B_{i,t-1} < \beta\xi_{i,t-1} \mid Z_i\}.$$

Thus covariates are a remedy when they shrink residual classification error, and a source of fragility when coarse class prices are passed through in a voluntary pool before buffers or histories are strong enough.

7.2 Reserve Pass-Through

The parameter λ_R is the model's rate-stability lever. It treats inherited reserve deficits and surpluses as prospective pricing components rather than instant assessments, consistent with actuarial guidance on cost of capital, contingency provisions, and additional funding sources (Actuarial Standards Board, 2011, 2017). It also reflects the insurance-pricing view that prices respond to capitalization and capacity after loss shocks (Cummins and Danzon, 1997). Since $\sum_{i\in A_t}\theta_{i,t-1} = 1$, (19) implies

$$\sum_{i\in A_t}\pi_{i,t-1} = \sum_{i\in A_t}\pi_{i,t-1}^E - \lambda_R R_{t-1}.$$

A negative reserve is therefore recovered through future premiums rather than an immediate assessment, while a positive reserve is credited back only gradually. The extreme $\lambda_R = 1$ tries to clear the inherited reserve state at the next renewal. Smaller λ_R smooths deficits and surplus credits across cohorts: it lowers the immediate premium shock after bad years, but leaves more reserve debt on the balance sheet; it also prevents a good year from exhausting the buffer too quickly through premium credits. Subsidies and reserve pass-through act on different objects. A subsidy lowers the household-paid premium slope, while λ_R governs how aggressively accumulated reserve gains or losses are priced into renewal.

7.3 Loss Allocation and Commitment

The expected-loss weights allocate retained losses and reserve pass-through within the pool. They change the pooling value $V_{i,t-1}$ and the distribution of reserve credits or charges across members, but they do not remove the price wedge. The wedge $\xi_{i,t-1} = \hat{\mu}_{i,t-1} - \mu_i$ still enters the participation margin through the experience-rated premium.

Committed contracts can in principle shut down the wedge channel. If the pool fixes renewal terms before later posteriors are observed, $\xi_{i,t-1}$ no longer enters the participation margin. The difficulty is no-lapse rather than truth-telling: once a member is revealed to be low risk, a committed pooled premium may exceed the outside experience-rated price. Retaining such members requires continuation value or no-lapse rents. Commitment is therefore a genuine remedy, but a conditional one: it is most plausible when residual type dispersion is modest, histories become informative quickly, or the contract can use subsidies, reserves, or switching frictions to keep low-risk members inside the pool.

7.4 Mandation, Exit Costs, and Strategic Waiting

Mandatory participation sets $c_{i,t} = \infty$ or otherwise prevents exit, and sufficiently high transition costs can eliminate switching altogether (Cutler, Lincoln, and Zeckhauser, 2010). Such frictions can mechanically prevent a death spiral, but they introduce costs outside the model's reserve accounting, such as forced redistribution, enforcement, political resistance, and reduced responsiveness to private information. When some households' pooling value is below the loading cost and net exit cost, forcing participation lowers surplus, creating the classical over-insurance welfare loss described by Einav, Finkelstein, and Cullen (2010) and Einav and Finkelstein (2011). Similarly, Handel (2013) finds that inertia keeps relatively healthy enrollees in comprehensive plans, mitigating adverse selection, so lowering exit costs can intensify selection. Strategic waiting is a natural extension. If households observe reserves and expect premiums to normalize, they may exit temporarily and re-enter. The baseline keeps sticky exit because it matches enrollment-window and switching-cost settings, documented for life insurance by Hendel and Lizzeri (2003).

8 Data Requirements and RMA Corroboration

The ideal dataset to test our mechanism would be an individual/household level panel from a voluntary mutual-aid or P2P insurance platform. It would observe each member's entry covariates, renewal premium, claims history, posterior risk score or enough information to reconstruct it, stay/exit decision, replacement by new entrants, realized claims, and the pool's reserve or surplus position. Such data would allow a direct test of the three links in the model. Chinese mutual-aid (MA) platforms such as Xiang Hu Bao are a good fit for this kind of evidence because they combine digital enrollment, voluntary participation, and transparent pool-level accounting.

To the best of my knowledge, no public dataset cleanly identifies this full loop. We therefore use USDA RMA crop insurance as an imperfect but informative public analogue. RMA data contain persistent county-crop risks, experience-based repricing, subsidies, and observable participation changes. They do not observe individual membership histories or a literal P2P reserve account. The empirical exercise should therefore be read as component-level corroboration.

8.1 Data

We use USDA RMA crop insurance data (2000–2024) as a subsidized public-private analogue. County-crop risks are persistent, rates respond to loss experience, and participation can change through cancellations, nonrenewals, or coverage adjustments. We build a county-crop-year panel for corn, soybeans, and wheat from public Summary-of-Business files, yielding 130,050 observations after basic premium and liability screens.[4] Constructing the five-year trailing wedge and one-year-ahead policy growth trims the earliest and last years, leaving roughly 103,000 county-crop-year observations in the preferred participation specification; specification-specific sample sizes are reported in Table 2. Its main limitation is also informative: federal subsidies dampen the household-paid premium wedge and therefore attenuate voluntary survivor selection. The empirical exercise just asks whether the model's objects move in the predicted directions.

The empirical analogue of the wedge $\xi_{i,t} = \hat{\mu}_{i,t} - \mu_i$ is a forecast residual,

$$\text{wedge}_{i,t} = \text{priced\_lr}_{i,t} - \text{loss\_ratio}_{i,t}, \qquad \text{priced\_lr}_{i,t} = \frac{1}{k}\sum_{s=t-k}^{t-1} \text{loss\_ratio}_{i,s},$$

predetermined with $k = 5$.[5] Participation tests use the one-period-lagged wedge $\text{wedge}^{\text{pre}}_{i,t} = \text{wedge}_{i,t-1}$, formed before the renewal decision and free of the contemporaneous indemnity. We use within-county-crop loss-ratio volatility as a rough learnability proxy. Lower volatility is interpreted as a higher-SNR environment, but, in reality, can also reflect crop mix, common shocks, denominator volatility, and program rules.

8.2 Hypotheses and Strategy

Four predictions follow. First, wedge variance falls with history length and is lower in higher-SNR groups (learning premise). Second, holding predicted risk fixed, higher wedges predict exit (H_a). Third, pools with heavier exit have more negatively selected survivors (H_b). Fourth, cumulative deficits predict future premium-rate increases (H_c). Because the empirical wedge is a forecast residual, raw wedge-outcome correlations are uninformative. The design separates predicted risk levels from residual misclassification: ordinary adverse selection should operate through predicted risk itself, while the misclassification channel predicts that, holding predicted risk fixed, predetermined residual overpricing predicts exit. Link (a) therefore conditions on predicted risk and uses the predetermined lagged wedge. This is not a causal design: lagged loss shocks may still affect participation through wealth, indemnities, salience, crop conditions, or local shocks. Link (b) is estimated at the pool level, where any individual mechanical wedge-outcome relationship averages out. We write empirical coefficients as b to avoid confusion with the structural slope $\beta = \chi(1 + \ell)$. The learning premise is a persistence regression $\text{LR}_{i,t+1} = a + b\,\text{LR}_{i,t-1}$ plus the wedge-variance decay reported below. Here i indexes county-crop cells and t indexes years. The pool index p denotes state-crop

[4] Data is obtained from the USDA FPAC mirror.

[5] A short trailing window proxies experience rating on recent claims; the result is robust to $k \in \{3, 5, 7, 10\}$.

pools; Links (b) and (c) are estimated on pool-year observations (p,t). The three links map to

$$\begin{aligned}\Delta \log \text{policies}_{i,t+1} &= b_0 + b_1\,\text{wedge}^{\text{pre}}_{i,t} + b_2\,\text{priced\_lr}_{i,t} \\ &\quad + b_3\,\text{subsidy\_rate}_{i,t} + b_4\,\overline{\text{coverage\_level}}_{i,t} + e_{i,t+1}, \qquad (33)\end{aligned}$$

$$\text{gap}_{p,t} = b_0 + b_1\,\text{exit\_rate}_{p,t} + W'_{p,t}g + e_{p,t}, \qquad (34)$$

$$\Delta \log \text{premium\_rate}_{p,t+1} = b_0 + b_1\,\text{cum\_deficit\_scaled}_{p,t} + b_2\,\text{realized\_lr}_{p,t} + e_{p,t+1}. \qquad (35)$$

where $\text{gap}_{p,t} \equiv \overline{\text{wedge}}^{\text{surv}}_{p,t} - \overline{\text{wedge}}^{\text{full}}_{p,t}$ and $W_{p,t}$ collects pool-level controls such as baseline pool size, crop, state, and year effects where available. The Link (a) prediction is $b_1 < 0$: higher-wedge county-crops should have lower next-year policy-count growth. A $\text{wedge}^{\text{pre}} \times (1-\phi)$ interaction tests the $\beta_{\text{eff}} = (1-\phi)\beta$ channel of (29), with a predicted negative coefficient as producers bear a larger premium share, consistent with crop-insurance demand responding to the out-of-pocket share (Du et al., 2017). Link (b) predicts $b_1 < 0$. Link (c) predicts $b_1 > 0$: larger cumulative deficits should be followed by higher premium-rate growth. Because RMA rates mechanically incorporate past loss experience, Link (c) is an institutional consistency check rather than an independent identification test. Link (a) uses OLS with standard errors clustered by state; Link (c) clusters by state-crop. The pool-level Link (b) uses HC1 heteroskedasticity-robust errors (Cameron and Miller, 2015). Robustness is in Appendix A.1.

8.3 Results

Table 1 summarizes the verdict.

Table 1: Summary evidence for the empirical links in RMA crop insurance.

Test (prediction)	Empirical object	Estimate (SE)	Assessment
Learning premise: persistence regression (+)	Loss-ratio persistence	0.084*** (0.004)	Predicted sign, low R^2
Learning premise: wedge variance decay	Wedge variance by SNR group	–	Predicted ranking
H_a: wedge $\rightarrow$ exit ($-$)	Predetermined wedge $\rightarrow$ Δ log policies	-0.0039***	Predicted sign, suggestive
H_b: exit $\rightarrow$ survivor gap ($-$)	Exit rate $\rightarrow$ survivor-minus-full-pool wedge gap	-0.032 (0.047)	Predicted sign, imprecise
H_c: deficit $\rightarrow$ repricing (+)	Cumulative deficit $\rightarrow$ Δ log premium rate	0.236*** (0.077)	Predicted sign, institutional check

Notes: Entries report OLS slopes unless otherwise stated. Assessments are descriptive: *predicted sign, low* R^2 indicates a precise positive persistence coefficient, but weak explanatory power; *predicted sign, suggestive* indicates correct sign but remaining identification or specification concerns; *predicted sign, imprecise* indicates the correct sign but a confidence interval too wide to be informative; *predicted sign, institutional check* indicates consistency with RMA repricing practice rather than an independent behavioral test. $^{***}p < 0.01$.

Learning premise. A county-crop persistence regression gives $b = 0.084$ ($p < 0.0001$) with $R^2 = 0.007$. Types are learnable, yet the signal is weak, as the model predicts for low-frequency risks. The sharper test is wedge-variance decay. Across trailing-window lengths, high-SNR county-crops have the smallest wedge variance and low-SNR the largest (Figure 2), the ordering implied by Lemma 1. The empirical wedge includes realization noise, so magnitudes are upper bounds, but the ranking is the predicted one.

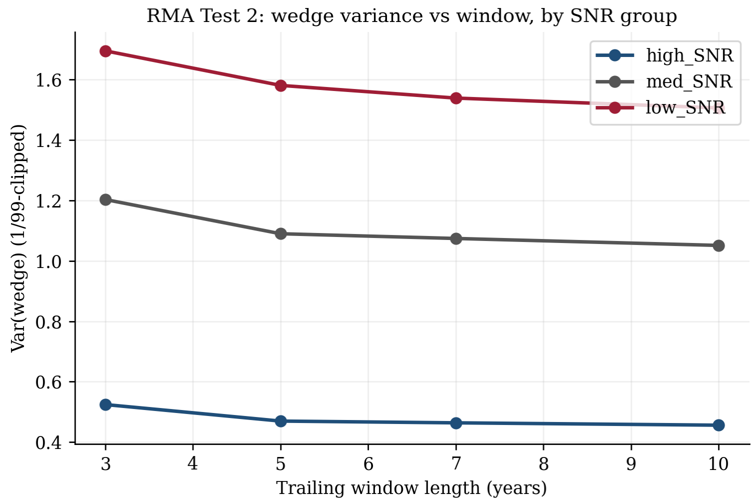

Figure 2: RMA wedge variance by trailing-window length and within-county-crop loss-ratio-SD tercile.

Link (a): exit selects on the wedge. The predetermined-wedge specification gives $\widehat{b}_1 = -0.0039$ (SE 0.0007, $p < 0.001, n = 102{,}737$). More overpriced county-crops have lower next-year policy-count growth (Figure 3), and the sign survives dropping the 2012 drought, the top loss-ratio decile, and restricting to low-SNR county-crops (Appendix A.1, Table 2). The sign is nonetheless specification-dependent. The appendix discrete-exit specification carries the opposite sign (+0.0019, $p = 0.10$), while the 2011-2018 continuous sub-period is null. Hence, we read Link (a) as suggestive. We cannot attribute the small magnitude to subsidy attenuation directly: the wedge $\times\,(1-\phi)$ interaction is imprecisely estimated (-0.0038, SE 0.0060, $p = 0.53$), so the RMA data do not separately identify this channel.

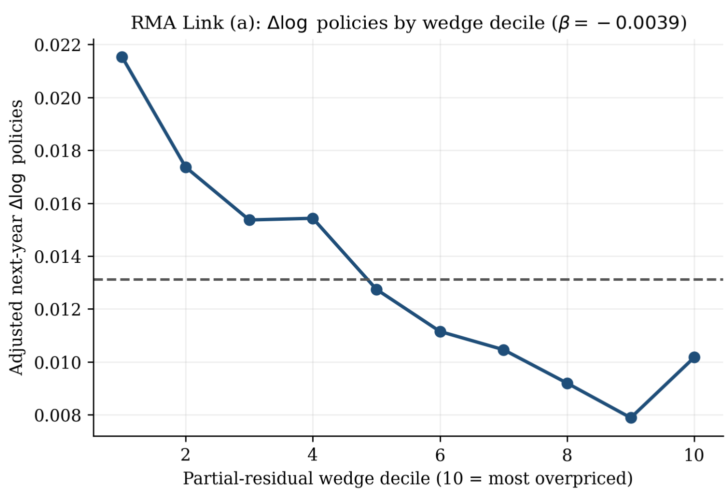

Figure 3: RMA link (a): next-year $\Delta\log$ policies against the partial-residual predetermined-wedge decile (Frisch–Waugh-Lovell residualization on priced LR, subsidy rate, and mean coverage level). Higher deciles have larger positive wedges; the downward gradient is consistent with the model's aggregate participation margin.

Link (b): exits change the survivor pool. At the state-crop-year pool level, regressing the survivor-minus-full-pool wedge gap on the exit rate with pool-size, state, crop, and year controls gives $\widehat{b}_1 = -0.032$ (SE 0.047, $p = 0.50$, $n = 2{,}279$ pool-years). The point estimate has the predicted sign, but the interval $[-0.12, +0.06]$ is too wide to determine the sign. It includes negative selection, a near-zero effect, and small positive effects. This is the closest RMA test of misclassification rather than generic adverse selection, because the outcome is a residual wedge gap rather than realized losses. Its imprecision is therefore a central empirical limitation, not evidence against the mechanism. The closed-loop composition result is assessed in simulation.

Link (c): reserve feedback. Regressing next-year log premium-rate changes on cumulative deficit-to-liability gives $\widehat{b}_1 = 0.236$ (SE 0.077, $p = 0.002, n = 2{,}856$): a 10-percentage-point higher deficit is associated with about a 2.4% rate increase (Figure 4). RMA rates mechanically use past loss experience, so this is an institutional consistency check; by itself it cannot distinguish misclassification from generic selection, common shocks, or statutory repricing.

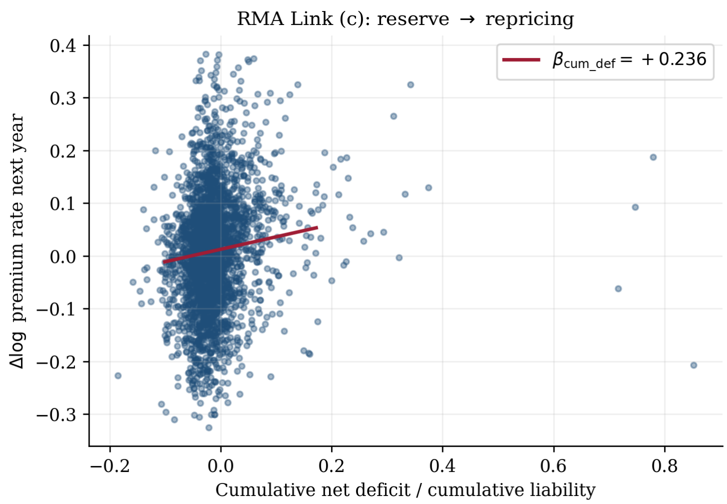

Figure 4: RMA link (c): next-year change in log premium rate against cumulative federal-program net deficit-to-liability at the state-crop level (fitted line includes realized LR).

In sum, the RMA evidence is consistent with selected directional components of the mechanism, but it does not identify the closed loop. The learning premise is strong, Link (a) is suggestive, Link (b) has the predicted sign but is imprecise under subsidies, and Link (c) matches RMA's institutional repricing logic.

9 Simulation Design and Results

The empirics test the links one at a time; the simulation shows how they operate together in one pool. Members enter with covariate scores but no individual loss history, the pool updates posteriors from claims signals, and renewal prices respond through experience rating. The exercise is a mechanism illustration under a transparent data-generating process, not a calibrated welfare estimate. Main figures average across deterministic seed offsets and report 10-90 percent seed bands. Calibration details are in Appendix A.3, Table 3.

9.1 Layered Mechanisms

Figure 5 illustrates the full cycle by layering the mechanisms. Full-information pricing removes the wedge and the pool is stable; hidden-type pooling shows the classical composition feedback; high-SNR experience rating learns quickly and stabilizes; low-SNR experience rating leaves large classification wedges that trigger selective exit and reserve deterioration. The panels are the simulation counterpart of the benchmarks in Sections 4–5.

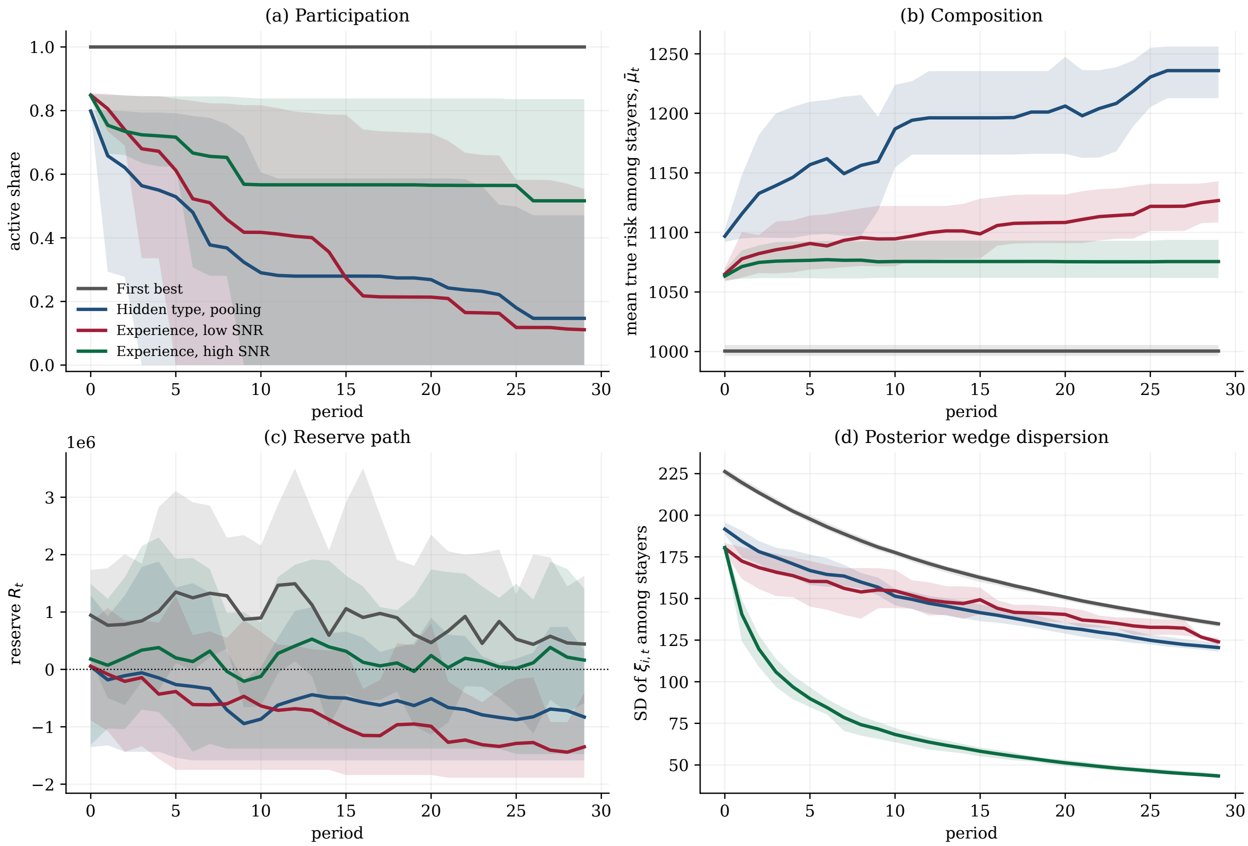

Figure 5: Layered mechanisms. Lines are means across seeds; shaded regions are 10–90 percent seed bands.

9.2 SNR Stability Region

Figure 6 varies σ_ε holding σ_X fixed, isolating learnability from risk-sharing value. Survival and reserve outcomes improve once residual SNR is high enough for wedges to shrink before they drive exits. Lower signal noise raises residual SNR and speeds learning. Better entry covariates stabilize through a distinct channel: they lower σ_0^2, shrinking wedge dispersion directly.

9.3 Frequency-Severity Environments

Figure 7 makes the catastrophe logic explicit: holding expected loss fixed, low frequency implies larger severity per event and noisier learning, so low-frequency, high-severity lines are hardest for voluntary experience rating. The same event that teaches the pool about risk also creates a large noisy premium revision.

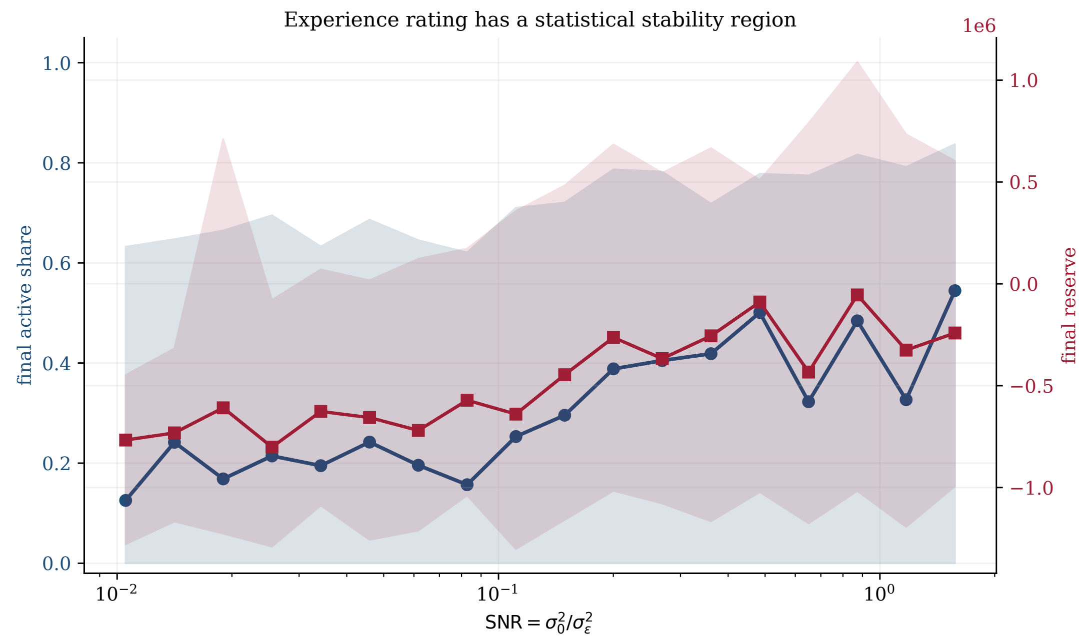

Figure 6: SNR stability map. The x-axis is residual $\mathsf{SNR} = \sigma_0^2/\sigma_\varepsilon^2$ after entry scoring; bands report 10-90 percent variation across seeds.

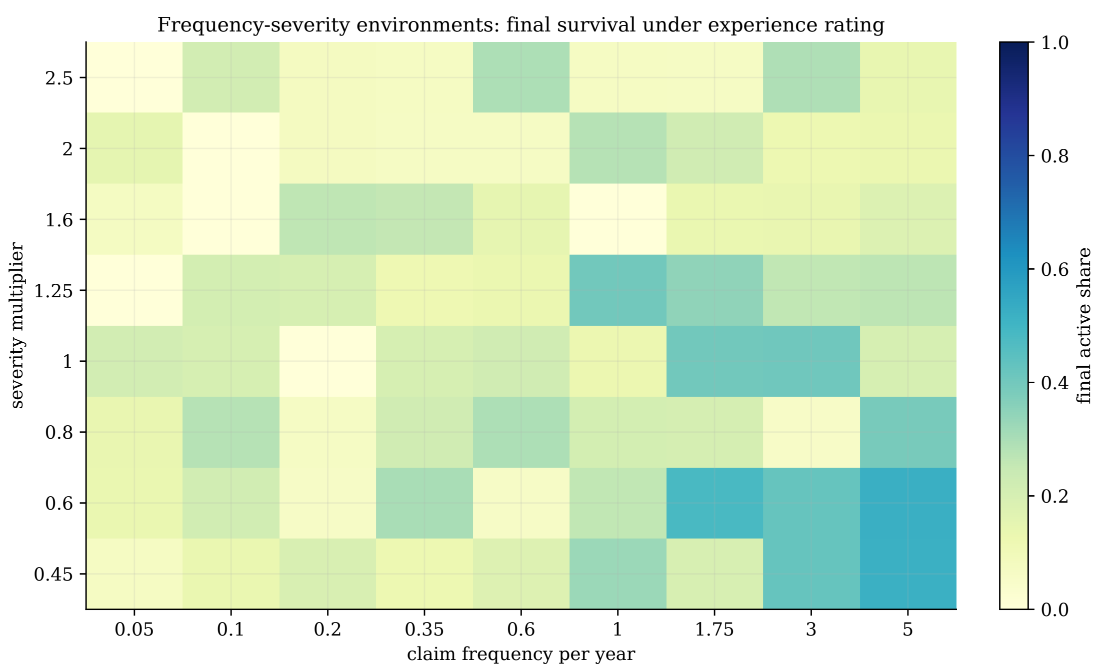

Figure 7: Frequency-severity grid. Cells report average final survival under experience rating across random seeds.

The regime map also explains the empirical pattern: low $T \cdot \mathsf{SNR}$ makes wedges large enough to move the participation margin, while subsidies lower β_{eff} and raise buffers. The model predicts that subsidies should attenuate the exit and survivor links by lowering the household-paid wedge. In the RMA data, this attenuation channel is plausible but not separately identified.

10 Conclusion

Risk classification does not uniformly solve adverse selection. In voluntary dynamic pools, it can create a selection force when coarse entry scores or noisy posteriors enter premiums before residual risk is learned precisely. With predictive covariates and informative histories, experience rating stabilizes the pool. With low residual SNR, the classification wedge becomes a participation shock: overpriced households exit, survivors are selected toward negative wedges, and reserves can deteriorate when buffers are insufficient. The pool may then enter a misclassification spiral.

The policy implication is conditional. High-frequency, low-severity lines can rely more heavily on experience rating and rich entry covariates. For low-frequency, high-severity lines, the model instead points toward slower adjustment, reserve buffers, commitment contracts, subsidies, or mandation as possible stabilizers. Because types are fixed and learning gradually reduces classification error, commitment-based remedies need only cover a finite learning window rather than last forever. The speed limit $\chi^{\max}$ provides a sufficient local diagnostic for the tradeoff between actuarial fairness and pool stability. The paper establishes a mechanism, a regime map, and a design rule, not a universal collapse theorem. The learning premise is established analytically, illustrated in simulation, and supported by RMA wedge-variance patterns. The selection mechanism is proved under $\beta \leq 1$ and receives suggestive corroboration in the data, while the full closed loop is demonstrated in simulation.

These results should not be read as an argument against experience rating. The model isolates a classification channel under fixed risk types, full loss reporting, full indemnity, and no moral hazard. In settings where policyholders can affect claim probabilities or severities, experience-rated premiums may encourage prevention and self-protection. Where small claims are not always reported they may also function like an endogenous deductible. Experience rating is therefore likely to be most valuable in high-SNR, controllable-risk lines such as automobile insurance, and less so in low-frequency catastrophe lines where short histories remain noisy. Extending the model to prevention effort, strategic claim reporting, partial coverage, and endogenous contract choice is left for future work.

A Additional Results and Robustness

A.1 RMA Link (a) Robustness

A.2 Finite Commitment Window

Corollary 1 (Finite commitment window)**.** *Fix a tail level $\alpha < 1/2$ and a conservative buffer bound $\underline{B} > 0$. Under the Gaussian benchmark the wedge-driven exit probability falls below α once*

$$T_{i,t-1} \;\geq\; T^*(\alpha) \;\equiv\; \frac{1}{\mathsf{SNR}}\left[\left(\frac{\beta\, q_{1-\alpha}\, \sigma_0}{\underline{B}}\right)^2 - 1\right]. \tag{36}$$

The result follows by inverting (28). Because the type is fixed, wedge dispersion and the tail probability of wedge-driven exit decline with $T_{i,t-1}\mathsf{SNR}$. In this benchmark, a temporary subsidy or price freeze can therefore cover the finite learning window rather than run forever.

Table 2: USDA RMA crop-insurance link (a) robustness across specifications. The wedge is the predetermined (one-period-lagged) participation wedge; state-clustered standard errors. The continuous-exit rows are the preferred test (predicted sign < 0) because the county-crop outcome aggregates many binary exit decisions and coverage adjustments, giving a higher-powered participation measure. The discrete specification is reported for completeness.

Specification	Wedge coef.	SE	p	n	R^2
Discrete exit, 2000–2024	$+0.0019$	(0.0012)	0.10	110,370	0.010
Continuous exit $\Delta\log(\text{policies})$, 2000–2024	-0.0039^{***}	(0.0007)	< 0.001	102,737	0.003
Continuous exit, 2011–2018	$+0.0004$	(0.0007)	0.57	39,857	0.001
Continuous, wedge $\times(1-\phi)$ interaction	-0.0038	(0.0060)	0.53	102,737	0.003
Continuous, long-run-mean wedge benchmark	-0.0011	(0.0034)	0.74	108,174	0.001
Continuous, drop 2012 drought	-0.0043^{***}	(0.0007)	< 0.001	97,762	0.004
Continuous, drop top-decile loss years	-0.0036^{***}	(0.0007)	< 0.001	93,471	0.003
Continuous, low-SNR county-crops only	-0.0028^{***}	(0.0008)	< 0.001	32,160	0.003

A.3 Additional Simulation Diagnostics

Figure 8 compares reserve pass-through rules and mandation in a low-SNR environment. Lower λ_R smooths inherited reserve gains and losses through renewal premiums, while mandation preserves membership mechanically but is not a costless welfare improvement, since the simulation does not price enforcement, redistribution, or political-economy costs.

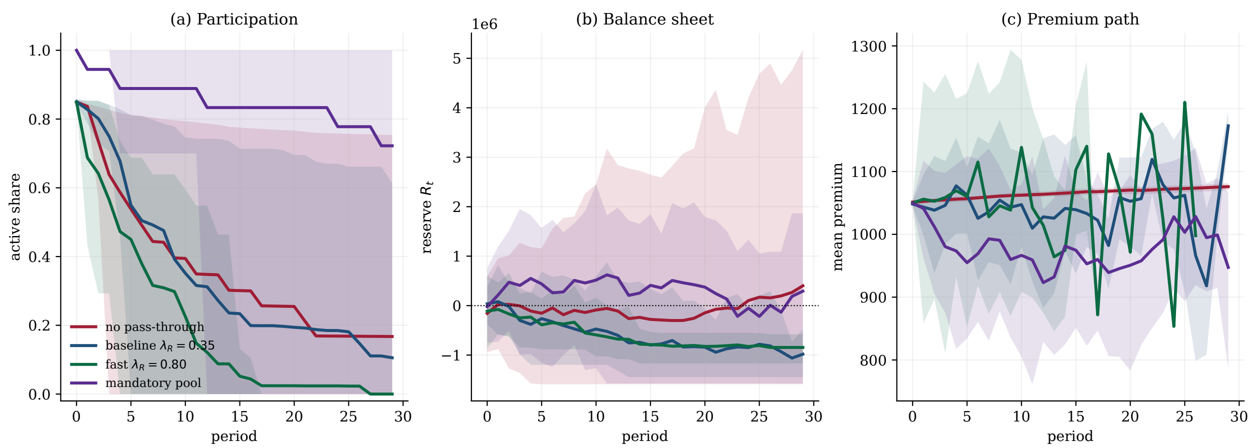

Figure 8: Policy comparison in a low-SNR environment. Reserve pass-through governs how quickly inherited reserve gains or losses enter renewal premiums; mandation keeps members in the pool by assumption.

Table 3: Baseline simulation calibration.

Object	Symbol/code	Baseline	Source / rationale
Population	N	4000	Illustrative baseline population size.
Horizon	T	30	Long enough to surface dynamic spirals, short enough to keep figures legible.
Initial signal history	T_0	0	Members enter with covariates but no member-specific loss history.
Mean expected loss	$\bar{\mu}_0$	1000	Currency-neutral normalization for $\mathbb{E}[m_i]$.
Total type SD	σ_μ	280	Cross-sectional expected-loss dispersion before entry scoring.
Entry covariate fit	R_Z^2	0.35	Share of type variance captured by entry covariates in the baseline.
Residual type SD	σ_0	226	Residual classification uncertainty after entry scoring.
Type floor	$\mu_{\min}$	250	Simulation-only nonnegativity floor for expected-loss types.
Loss SD	σ_X	700	Per-period loss CV $\approx$ 0.7, consistent with non-catastrophe lines.
Baseline signal SD	σ_ε	900	Baseline SNR $\approx$ 0.063; varied in Figure 6.
Correlation	ρ_ε	0.10	Mild common-shock exposure in the baseline.
Risk aversion	γ	0.00085	Gives pooling value $V \approx 187$, large but not exit-preventing.
Experience intensity	χ	0.85	High classification pass-through chosen to stress-test the mechanism.
Loading	ℓ	0.05	Standard small expense load, below the $\approx 8\%$ charged by some mutual-aid platforms (Chen et al., 2024).
Risk-free return	r_f	0.02	Baseline return on reserves and collected premiums.
Reserve pass-through	λ_R	0.35	Fraction of inherited reserve state passed into renewal premiums.
Premium slope	$\beta = \chi(1 + \ell)$	0.8925	Sub-critical ($\beta \leq 1$), so Lemma 2 applies.
Solvency floor	R_t/n_t	-450	Pool closes when reserve debt per active member falls below this threshold.

B Proofs

B.1 Gaussian Updating and Learning Results

Proof of the Gaussian updating formulas.

Proof. Fix an agent and suppress indices. The entry score is $m = m(Z)$ and the residual type is $\delta = \mu - m \sim \mathcal{N}(0, \sigma_0^2)$. Let T be the number of observed signals and, for $T > 0$, let $\bar{S} = T^{-1} \sum_{\tau=1}^{T} S_\tau$. Since $S_\tau - m = \delta + \zeta_\tau$, normal–normal conjugacy[6] for the residual type gives

$$\sigma_T^2 = \left(\frac{1}{\sigma_0^2} + \frac{T}{\sigma_\varepsilon^2} \right)^{-1} = \frac{\sigma_0^2 \sigma_\varepsilon^2}{T\sigma_0^2 + \sigma_\varepsilon^2} = \frac{\sigma_0^2}{1 + T\mathsf{SNR}},$$

which is (9). The posterior mean of $\mu = m + \delta$ is

$$\hat{\mu} = m + \sigma_T^2 \left(\frac{\sum_{\tau=1}^{T} (S_\tau - m)}{\sigma_\varepsilon^2} \right) = \frac{\sigma_0^2 \sum_{\tau=1}^{T} S_\tau + \sigma_\varepsilon^2 m}{T\sigma_0^2 + \sigma_\varepsilon^2}.$$

which is (6). Substituting $\sum_{\tau=1}^{T} S_\tau = T\bar{S}$ gives

$$\hat{\mu} = (1 - z_T)m + z_T \bar{S}, \qquad z_T = \frac{T\sigma_0^2}{T\sigma_0^2 + \sigma_\varepsilon^2} = \frac{T\mathsf{SNR}}{1 + T\mathsf{SNR}}.$$

This proves (7). For $T = 0$, set $z_T = 0$ and $\hat{\mu} = m$. □

Proof of Lemma 1.

Proof. Fix the entry covariates and the history length $T_{i,t}$. Let

$$\mathcal{F}_{i,t} = \sigma(Z_i, S_{i,1:T_{i,t}})$$

denote the information used to form the posterior. By Definition 1,

$$\xi_{i,t} = \hat{\mu}_{i,t} - \mu_i, \qquad \hat{\mu}_{i,t} = \mathbb{E}[\mu_i \mid \mathcal{F}_{i,t}].$$

Hence

$$\mathbb{E}[\xi_{i,t} \mid \mathcal{F}_{i,t}] = \hat{\mu}_{i,t} - \mathbb{E}[\mu_i \mid \mathcal{F}_{i,t}] = 0.$$

Therefore, conditional on Z_i and $T_{i,t}$,

$$\mathrm{Var}(\xi_{i,t} \mid Z_i, T_{i,t}) = \mathbb{E}[\mathrm{Var}(\mu_i \mid \mathcal{F}_{i,t}) \mid Z_i, T_{i,t}]\,.$$

Under the Gaussian benchmark, the posterior variance is data-independent and equals (9). Thus

$$\mathrm{Var}(\xi_{i,t} \mid Z_i, T_{i,t}) = \frac{\sigma_0^2}{1 + T_{i,t}\mathsf{SNR}}.$$

□

[6] This is the standard normal–normal Bayesian updating formula: if $\delta \sim \mathcal{N}(0, \sigma_0^2)$ and $Y_\tau = \delta + \zeta_\tau$ with $\zeta_\tau \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$, then $\delta \mid (Y_1, \ldots, Y_T)$ is Gaussian with posterior variance $\left(\sigma_0^{-2} + T\sigma_\varepsilon^{-2}\right)^{-1}$ and posterior mean $\left(\sigma_0^{-2} + T\sigma_\varepsilon^{-2}\right)^{-1} \sigma_\varepsilon^{-2} \sum_{\tau=1}^{T} Y_\tau$.

B.2 Experience-Rating Selection and Stability

Proof of Lemma 2.

Proof. Let $D_i = \mathbf{1}\{i \in A_t\}$ and write $T_i = T_{i,t-1}$. Work in the pre-renewal eligible population. Since the posterior is calibrated, $\mathbb{E}[\xi_{i,t-1} \mid m_i, T_i] = 0$, and hence $\mathbb{E}[\xi_{i,t-1}] = 0$. Therefore

$$\mathrm{Cov}(\xi_{i,t-1}, D_i) = \mathbb{E}[\xi_{i,t-1} D_i] = \mathbb{E}[\xi_{i,t-1} \mid D_i = 1] \Pr(D_i = 1).$$

Since there is positive stay mass, it suffices to show $\mathrm{Cov}(\xi_{i,t-1}, D_i) < 0$.

Conditional on (m_i, T_i, δ_i), Gaussian updating gives

$$\xi_{i,t-1} = -(1 - z_{T_i})\delta_i + z_{T_i}\bar{\zeta}_{i,t-1}, \qquad z_{T_i} = \frac{T_i \mathsf{SNR}}{1 + T_i \mathsf{SNR}}.$$

By assumption, holding (m_i, T_i, δ_i) fixed, the stay indicator is weakly decreasing in $\xi_{i,t-1}$ and strictly decreasing on a set of positive measure. Hence

$$\mathbb{E}[\mathrm{Cov}(\xi_{i,t-1}, D_i \mid m_i, T_i, \delta_i)] \le 0,$$

with strict inequality whenever there is positive wedge-driven exit mass.

Using the law of total covariance, nested by (m_i, T_i),

$$\begin{aligned}\mathrm{Cov}(\xi_{i,t-1}, D_i) = {} & \mathbb{E}[\mathrm{Cov}(\xi_{i,t-1}, D_i \mid m_i, T_i, \delta_i)] \\ & + \mathbb{E}[\mathrm{Cov}(\mathbb{E}[\xi_{i,t-1} \mid m_i, T_i, \delta_i], \Pr(D_i = 1 \mid m_i, T_i, \delta_i) \mid m_i, T_i)] \\ & + \mathrm{Cov}(\mathbb{E}[\xi_{i,t-1} \mid m_i, T_i], \Pr(D_i = 1 \mid m_i, T_i))\,.\end{aligned}$$

The last term is zero by calibration. The first term is nonpositive by the monotonicity of the stay indicator in the wedge. For the second term,

$$\mathbb{E}[\xi_{i,t-1} \mid m_i, T_i, \delta_i] = -(1 - z_{T_i})\delta_i,$$

which is decreasing in δ_i, while the stay probability is weakly increasing in δ_i by assumption. Its covariance is therefore nonpositive. Thus $\mathrm{Cov}(\xi_{i,t-1}, D_i) < 0$ whenever there is positive wedge-driven exit mass, and therefore

$$\mathbb{E}[\xi_{i,t-1} \mid i \in A_t] < 0.$$

□

Proof of Proposition 2.

Proof. Write the survivor sum over the fixed population with the random indicator: $\sum_{i \in A_t} \xi_{i,t-1} = \sum_{i \in \mathcal{I}} \xi_{i,t-1} \mathbf{1}\{i \in A_t\}$. By Lemma 2, $\mathbb{E}[\xi_{i,t-1}\mathbf{1}\{i \in A_t\}] = \mathbb{E}[\xi_{i,t-1} \mid i \in A_t] \Pr(i \in A_t) < 0$, so $\mathbb{E}[-\chi \sum_{i \in A_t} \xi_{i,t-1}] = -\chi \sum_{i \in \mathcal{I}} \mathbb{E}[\xi_{i,t-1}\mathbf{1}\{i \in A_t\}] > 0$. □

Proof of Proposition 3.

Proof. Wedge-driven exit requires $\xi_{i,t-1} > \underline{B}/\beta$. Under the Gaussian benchmark $\xi_{i,t-1}$ has standard deviation $\sigma_0 / \sqrt{1 + T_{i,t-1}\mathsf{SNR}}$, so the normal-tail bound gives (28). □

References

Abdikerimova, S. and R. Feng (2022). Peer-to-peer multi-risk insurance and mutual aid. *European Journal of Operational Research* 299(2), 735–749.

Abdikerimova, S., T. J. Boonen, and R. Feng (2024). Multiperiod peer-to-peer risk sharing. *Journal of Risk and Insurance* 91(4), 943–982.

Actuarial Standards Board (2011). *Actuarial Standard of Practice No. 30: Treatment of Profit and Contingency Provisions and the Cost of Capital in Property/Casualty Insurance Ratemaking.*

Actuarial Standards Board (2017). *Actuarial Standard of Practice No. 53: Estimating Future Costs for Prospective Property/Casualty Risk Transfer and Risk Retention.*

Akerlof, G. A. (1970). The market for lemons: quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84(3), 488–500.

Braun, A., N. Häusle, and P. D. Thistle (2023). Risk classification with on-demand insurance. *Journal of Risk and Insurance* 90(4), 975–990.

Buchmueller, T. and J. DiNardo (2002). Did community rating induce an adverse selection death spiral? Evidence from New York, Pennsylvania, and Connecticut. *American Economic Review* 92(1), 280–294.

Buhlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin* 4(3), 199–207.

Buhlmann, H. and A. Gisler (2005). *A Course in Credibility Theory and Its Applications*. Springer.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources* 50(2), 317–372.

Cardon, J. H. and I. Hendel (2001). Asymmetric information in health insurance: evidence from the National Medical Expenditure Survey. *RAND Journal of Economics* 32(3), 408–427.

Cawley, J. and T. Philipson (1999). An empirical examination of information barriers to trade in insurance. *American Economic Review* 89(4), 827–846.

Chen, Z., R. Feng, L. Wei, and J. Zhao (2024). Cost-effectiveness, fairness and adverse selection in mutual aid. *European Financial Management* 30(3), 1510–1544.

Chiappori, P.-A. and B. Salanie (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy* 108(1), 56–78.

Cochrane, J. H. (1995). Time-consistent health insurance. *Journal of Political Economy* 103(3), 445–473.

Cohen, A. (2005). Asymmetric information and learning: evidence from the automobile insurance market. *Review of Economics and Statistics* 87(2), 197–207.

Cohen, A. and P. Siegelman (2010). Testing for adverse selection in insurance markets. *Journal of Risk and Insurance* 77(1), 39–84.

Cooper, R. and B. Hayes (1987). Multi-period insurance contracts. *International Journal of Industrial Organization* 5(2), 211–231.

Crocker, K. J. and A. Snow (1986). The efficiency effects of categorical discrimination in the insurance industry. *Journal of Political Economy* 94(2), 321–344.

Cummins, J. D. and P. M. Danzon (1997). Price, financial quality, and capital flows in insurance markets. *Journal of Financial Intermediation* 6(1), 3–38.

Crocker, K. J. and N. Zhu (2021). The efficiency of voluntary risk classification in insurance markets. *Journal of Risk and Insurance* 88(2), 325–350.

Cutler, D. M., B. Lincoln, and R. Zeckhauser (2010). Selection stories: understanding movement across health plans. *Journal of Health Economics* 29(6), 821–838.

Cutler, D. M. and S. J. Reber (1998). Paying for health insurance: the trade-off between competition and adverse selection. *Quarterly Journal of Economics* 113(2), 433–466.

Cutler, D. M. and R. J. Zeckhauser (1998). Adverse selection in health insurance. *Forum for Health Economics & Policy* 1(1).

Denuit, M. (2019). Size-biased transform and conditional mean risk sharing, with application to P2P insurance and tontines. *ASTIN Bulletin* 49(3), 591–617.

Denuit, M. and J. Dhaene (2012). Convex order and comonotonic conditional mean risk sharing. *Insurance: Mathematics and Economics* 51(2), 265–270.

Dionne, G. (1983). Adverse selection and repeated insurance contracts. *Geneva Papers on Risk and Insurance* 8(29), 316–333.

Dionne, G. and N. A. Doherty (1994). Adverse selection, commitment, and renegotiation: extension to and evidence from insurance markets. *Journal of Political Economy* 102(2), 209–235.

Dionne, G. and P. Lasserre (1985). Adverse selection, repeated insurance contracts and announcement strategy. *Review of Economic Studies* 52(4), 719–723.

Du, X., H. Feng, and D. A. Hennessy (2017). Rationality of choices in subsidized crop insurance markets. *American Journal of Agricultural Economics* 99(3), 732–756.

Einav, L. and A. Finkelstein (2011). Selection in insurance markets: theory and empirics in pictures. *Journal of Economic Perspectives* 25(1), 115–138.

Einav, L., A. Finkelstein, and M. R. Cullen (2010). Estimating welfare in insurance markets using variation in prices. *Quarterly Journal of Economics* 125(3), 877–921.

Eling, M., I. Gemmo, D. Guxha, and H. Schmeiser (2024). Big data, risk classification, and privacy in insurance markets. *The Geneva Risk and Insurance Review* 49(1), 75–126.

Finkelstein, A. and K. McGarry (2006). Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economic Review* 96(4), 938–958.

Finkelstein, A. and J. Poterba (2004). Adverse selection in insurance markets: policyholder evidence from the U.K. annuity market. *Journal of Political Economy* 112(1), 183–208.

Gal, S. and M. Landsberger (1988). On "small sample" properties of experience rating insurance contracts. *Quarterly Journal of Economics* 103(1), 233–243.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: when nudging hurts. *American Economic Review* 103(7), 2643–2682.

Handel, B., I. Hendel, and M. D. Whinston (2015). Equilibria in health exchanges: adverse selection versus reclassification risk. *Econometrica* 83(4), 1261–1313.

Hendel, I. and A. Lizzeri (2003). The role of commitment in dynamic contracts: evidence from life insurance. *Quarterly Journal of Economics* 118(1), 299–328.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review* 61(4), 561–574.

Hosios, A. J. and M. Peters (1989). Repeated insurance contracts with adverse selection and limited commitment. *Quarterly Journal of Economics* 104(2), 229–253.

Hoy, M. (1982). Categorizing risks in the insurance industry. *Quarterly Journal of Economics* 97(2), 321–336.

Klein, P. P., R. van Kleef, J. Henriquez, and F. Paolucci (2023). The interplay between risk adjustment and risk rating in voluntary health insurance. *Journal of Risk and Insurance* 90(1), 59–91.

Kousky, C. (2018). Financing flood losses: a discussion of the National Flood Insurance Program. *Risk Management and Insurance Review* 21(1), 11–32.

Kunreuther, H., M. Pauly, and S. McMorrow (2013). *Insurance and Behavioral Economics: Improving Decisions in the Most Misunderstood Industry*. Cambridge University Press.

Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers.

Ludkovski, M. and V. R. Young (2010). Ex post moral hazard and Bayesian learning in insurance. *Journal of Risk and Insurance* 77(4), 829–856.

Mulder, P. (2019). Dynamic adverse selection in flood insurance. Working paper.

Ohlsson, E. and B. Johansson (2010). *Non-Life Insurance Pricing with Generalized Linear Models*. Springer.

Pauly, M. V., H. Kunreuther, and R. Hirth (1995). Guaranteed renewability in insurance. *Journal of Risk and Uncertainty* 10(2), 143–156.

Puelz, R. and A. Snow (1994). Evidence on adverse selection: equilibrium signaling and cross-subsidization in the insurance market. *Journal of Political Economy* 102(2), 236–257.

Rothschild, M. and J. Stiglitz (1976). Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *Quarterly Journal of Economics* 90(4), 629–649.

Stiglitz, J. E. (1977). Monopoly, non-linear pricing and imperfect information: the insurance market. *Review of Economic Studies* 44(3), 407–430.

Whitney, A. W. (1918). The theory of experience rating. *Proceedings of the Casualty Actuarial Society* 4, 274–292.