Endogenous Belief Hazards

July 22, 2026

Abstract

When an agent's belief is a held model, revised by hypothesis testing—kept until its record rejects it, then re-selected—the current belief is not a sufficient statistic for its own revision: today's forecast does not pin down the law of the forecasts to come. The primitive is Ortoleva's hypothesis-testing model, run in continuous time. We show the abandonment rate is an *endogenous belief hazard*, governed by the record under which the model has survived. Consequently, the irreducible state *enlarges*, from the belief to the belief with its surviving record—and, across revisions, the carried ranking—finite-dimensional and Markov on exponential-Lévy menus. Two forecasters holding the same model and sharing every public posterior can nevertheless face different revision hazards. Across a population watching one path, differing only in tolerance for surprise, belief heterogeneity is endogenous: the cross-section of belief states is the exact pushforward of the tolerance distribution.

Keywords: belief revision, hypothesis testing, non-Bayesian updating, adjustment hazard, sufficient statistic, heterogeneous beliefs.
JEL classification: D81, D83, D84, C12.

1 Introduction

Macroeconomics and finance infer beliefs from summaries: where they sit (the consensus forecast) and how spread out they are (the dispersion). Two forecasters who report the same forecast are, by those summaries, in the same state.[1] This paper is about a coordinate they leave out. Ada and Bea both report three percent inflation and are very differently exposed to abandoning it. Ada reached the number only recently; when she adopted the model, its test record reset, so the record she now monitors is short and well inside its ordinary range. She is *robust*. Bea has held three percent through a long stretch in which the data crept toward the edge of her model's ordinary range; her record, aged and near its limit, leaves her one surprise from revising. She is *fragile*.[2] They forecast alike and revise unalike.[3] What the held forecast misses is a belief's fragility, the hazard of revision, carried by the *since-adoption* record under which the model has survived, through the age-dependent band the record is tested against.

The behavior that makes the record matter is the way a cautious statistician, or an investor who acts like one, treats a model she is unsure of. Rather than average the candidate models into a posterior and act on the mixture, she holds the single model she finds most plausible and acts on it: she keeps it while the data look ordinary under it, abandons it the first time they look too unlikely to be a fluke, and then adopts whatever model now looks best. This is the hypothesis-testing rule Ortoleva (2012) axiomatizes, cast here in continuous time.[4] We model agents who carry a finite menu of candidate models and a ranking over it, hold their current front-runner, and monitor the one statistic the menu disputes: an average growth rate when the models disagree about a trend, a count of disasters when they disagree about a crash rate. Each agent retains her front-runner while her since-adoption record lies in the *ordinary range* of a size-α test under that model and re-selects when the record exits it. Across agents, the sole heterogeneous behavioral type is tolerance for surprise. The object we

[1] Treating the belief as the state has a decision-theoretic warrant: Epstein and Le Breton (1993) show that dynamic consistency forces Bayesian updating, under which the belief is its own sufficient statistic. The hypothesis-testing agent here is not Bayesian, so that warrant lapses, and the coordinate the summaries omit is the one this paper recovers (§3.1).

[2] Fragility and robustness are used here in their revision sense, distinct from the robustness literature: the *fragile beliefs* of **?** are beliefs tilted by concern for model misspecification, and robustness there is the preference generating the tilt. Here a belief's fragility is its exposure to the rule's own revision trigger, the hazard defined below.

[3] The polarity is the trend channel's, where a just-reset record is shielded through its probation interval and tenure ages it toward the edge. The channels invert the exposed party: just revised means immune through probation under a disputed trend, and one disaster from rejection under a disputed crash rate (Corollary 2).

[4] **?** find laboratory forecasters doing just that, revising in infrequent discrete jumps though no adjustment cost is present.

study is the rate at which the held model is abandoned: the *endogenous belief hazard*. It is history-dependent, the dynamic manifestation of a state the current belief does not contain.

The central result is that the record under which a belief has survived, its content and its age, is an irreducible state variable. The record carries the hazard, so the held belief is not a sufficient statistic for its own revision, and neither is any posterior a Bayesian forms by updating a fixed prior on the public data. Two forecasters who have watched the same public history share every posterior formed from a common fixed prior, yet differ in fragility, because they adopted their models on different dates. The state is thereby *enlarged*, from the held belief to the belief together with the record under which it has survived, and the revision hazard is determined by that enlarged state: on exponential-Lévy menus this enlargement—the carried ranking adjoined across revisions—yields a finite-dimensional Markov representation, so the dynamics of revision are tractable. With heterogeneous tolerances on one common history, the population endogenously staggers adoption dates, generating recurrent periods in which two forecasters hold the same model and share every fixed-prior public posterior but face different conditional revision laws. The closed state permits three further characterizations: a correct model is revised forever; holding times are heavy-tailed, with a common exponent across the two canonical channels; and the population state is the exact pushforward of the tolerance distribution under the history-dependent flow.

Across a population, the individual rule poses a macroeconomic question: when forecasters share the data but not their tolerance, how does a common shock move the cross-section of their beliefs? A more tolerant forecaster sits through news that tips a less tolerant one, so a common shock crosses the thresholds only of whoever was already close, and the consensus moves through them alone. The disagreement is not about information, since every forecaster has seen the same history, but about who is still willing to live with it: the spread of beliefs is the spread of tolerance carried forward along the one path the economy has taken, with nothing idiosyncratic to average away. Even when held-model heterogeneity eventually disappears in a closed population containing the truth, heterogeneity in the full belief state does not: ages and surviving records remain dispersed behind a potentially unanimous forecast. The cross-section is the fixed distribution of thresholds seen through the common history, and every aggregate is a functional of that history, with no market average to solve for; aggregate revision is set at the boundary, by the mass that crosses and where its re-selection sends it.

Related literature

The mechanism sits among several literatures without being contained by any. Its primitive, hold one model and abandon it when its own record rejects it, is Ortoleva's hypothesis-testing rule (surveyed in Ortoleva, 2024), which arose independently in macroeconomics as the *inferential expectations* of Menzies and Zizzo (2009), rational expectations its zero-conservatism limit; where that literature switches a representative agent's false null once to the rational expectation, the fixed-size test here rejects even a correct model recurrently and makes the cross-section a pushforward, neither available from a one-way switch. Full dynamic consistency would force Bayesian updating (Epstein and Le Breton, 1993), and Weinstein (2017) shows that an α-spending rule bounding the *lifetime* probability of such shifts keeps the agent approximately consistent, a guarantee the per-test reading forgoes, retaining the multiple-testing cost he flags.[5]

The nearest dynamic relatives switch models by comparison. The validation learner of Cho and Kasa (2015) runs its test against a time-varying threshold, and in the vanishing-gain limit a correct model is spared; our test size is fixed, so even a correct model is rejected recurrently, the classical *sampling to a foregone conclusion* (Anscombe, 1954; Robbins, 1970) that Proposition 2 imports, and it is that recurrence, not the presence of a better rival, that drives both the dynamics and the enlargement. The misspecified-learning literature updates over a model set by likelihood, as in Berk–Nash equilibrium (Esponda and Pouzo, 2016; Fudenberg et al., 2021; Frick et al., 2023); Ba (2026), nearest to us, switches models on a threshold in relative fit and asks which misspecifications persist. Our trigger is absolute: each model is judged against its own band rather than a rival's fit, so an incumbent can fail with no better rival on offer; and our question is timing rather than persistence, the hazard with which the held model is abandoned as its location settles, the recurrence itself the continuous-time analogue of the escape dynamics of Cho et al. (2002).

[5] The statistical ancestor is quickest change detection (Page, 1954; **?**), whose false-alarm and detection-lag trade-off Theorem 2(b) prices; the per-test rule is behaviorally axiomatized rather than statistically optimal, its game-theoretic ancestor (**?**) selecting equilibrium conjectures rather than revision times, and Lemma A.7 covers the anytime-valid flank.

The belief-driven-dynamics literature nearest to us keeps the belief itself as the state. In the smooth-filtering account of Dew-Becker et al. (2025) a constant true intensity is filtered into a continuously varying perceived one that delivers leverage, skewness, and volatility dynamics; we share the constant-intensity premise and generate perceived risk instead as a saw-tooth, an upward jump at each disaster that finds a complacent mass poised at the band edge and a scheduled decline between, the jump-and-timing surface smooth filtering does not produce, with the higher moments conceded to it. Across this literature (Wachter and Zhu, 2025; Kozlowski et al., 2020; Collin-Dufresne et al., 2016; Weitzman, 2007) the posterior is the state, so two agents who share it share their futures; in extrapolative and diagnostic expectations (Bordalo et al., 2018, 2020) the state is one systematic distortion, which makes predictable forecast errors and predictable outcomes the same object. Our agent carries neither: the held model consolidates on the Kullback–Leibler closest menu model (Proposition 2), any bias past consolidation is a fixed projection gap, and revision carries no forecast-error predictability of its own on the transition path and the turnover-sustained cross-section where observed revision lives. What we add is a setting in which the belief is not sufficient for its own motion (Theorem 1). Where Baley and Turén (2025) read reported forecasts as infrequent adjustments of continuously moving beliefs, the wedge runs the other way: the belief sits still, and observed changes thin an already sparse internal trigger.

The working-paper version (**?**) develops what this submission only points to: the hybrid nesting of cost-based hazards, stakes-invariance and its identification, the observed-count thinning, and the applied patterns—a complacent market that moves sharply when boundary-poised mass is surprised, lumpy individual forecasts against a sluggish consensus, disagreement widening in crises, experience-type effects (Malmendier and Nagel, 2011), a slow-decaying belief overhang that puts an informational trend on a stationary process, and the disaster economy's countercyclical price of jump risk. The companion equilibrium paper closes the loop where records are made by the beliefs that test them.

2 The rule in continuous time

A belief's fragility lives not in where it sits but in how its record stands against the model holding it up. This section states the rule that maps a record to a revision, in continuous time, and shows that the band it tests against is itself selected by the primitive. The technical groundwork is collected in Appendix A; the one interpretive choice continuous time forces, the reading of the tolerance, is faced at the section's end.

Fix a probability space carrying the data, with $\mathbb{P}^\star$ the physical (true) law and $(\mathcal{F}_t)_{t\geq 0}$ the filtration generated by the observed path. Consider an investor who watches the data continuously. We state her rule in three parts.

Setup. The data are a process $X = (X_t)_{t\geq 0}$. The investor entertains a finite menu of models $\Theta = \{\theta_1, \dots, \theta_K\}$, each a candidate law for X on $(\mathcal{F}_t)$, and a second-order belief $\rho \in \Delta(\Theta)$ ranking them. She holds the top-ranked model

$$\hat{\theta}_t = \arg\max_{\theta\in\Theta} \rho_t(\theta). \tag{1}$$

Ties are broken by a fixed priority order on Θ (retaining the incumbent when it is among the maximizers), so $\hat{\theta}_t$ is single-valued and $(\mathcal{F}_t)$-adapted. Under the continuous record law, *data-induced* likelihood ties at fixed times are $\mathbb{P}^\star$-null; a carried tie (the allowed tied initial ranking, frozen through a spell) persists by construction and is resolved by the order; and no almost-sure statement below turns on tie-nullness at the stopped re-selection times, where the order, not nullness, keeps the selection single-valued. In the counting channel ties have positive probability outright and the priority order is then substantive, part of the rule, keeping the selection deterministic and the held-model process adapted; the results below hold for any fixed choice of that order. Let σ_t be the last time she revised (or 0), so that

$$a_t := t - \sigma_t \tag{2}$$

is the age of her current model and $X_{(\sigma_t,t]}$ the data observed since she adopted it, her track record for the held model. We require the menu laws pairwise mutually locally equivalent on $(\mathcal{F}_t)$, so the record has a density under each model and the cross-menu likelihood ratios in (8) are well-defined (Appendix A).

(i) What she monitors. She judges her model on the one feature of the record the candidate models disagree about: the average growth rate, if they dispute the trend; the number of disasters, if they dispute the crash rate. Call this summary T_a, the minimal sufficient statistic of the record for the menu (Appendix A), and write its held-model density at age a as $f_a^{\hat\theta}$. The record accrues at a fixed *recording resolution* $\delta > 0$; the continuous-time objects below are its $\delta \downarrow 0$ limit (Appendix A).

Assumption 1 (Exponential-Lévy menu)**.** *The menu is an exponential family of Lévy processes: there is an additive sufficient statistic S_a—a Lévy functional of the record, with (T_a, a) a one-to-one function of (S_a, a)—the summary recovering S_a at each fixed age—such that for any two menu laws the record's likelihood ratio through age a has the* Wald form

$$\left.\frac{dP^\theta}{dP^{\theta'}}\right|_{\mathcal{F}_a} = \exp\!\big((\eta_\theta - \eta_{\theta'})\, S_a - a\,(\psi_\theta - \psi_{\theta'})\big), \tag{3}$$

a pathwise *function of (S_a, a) holding for every age a, $\mathbb{P}$-a.s., with η_θ the canonical parameter and ψ_θ the cumulant normalizer of model θ—only the differences $\psi_\theta - \psi_{\theta'}$ enter, so the normalizers are fixed up to one common shift, and the channel instances below take the natural gauge.*

Both channels are instances: the disputed-trend menu on a Brownian-with-drift record has $S_a = X_a$, $\eta_\theta = \mu_\theta/\sigma^2$, $\psi_\theta = \mu_\theta^2/(2\sigma^2)$, and $T_a = \bar X_a = S_a/a$ (Girsanov); the disputed-crash menu has $S_a = N_a$, $\eta_\theta = \log \lambda_\theta$, $\psi_\theta = \lambda_\theta$, and $T_a = N_a$. Two consequences are used below: (T_a, a) is minimal sufficient and Markov under each menu law, so the state $(\hat\theta, T_a, a)$ is *finite*-dimensional; and at any revision the stopped-record likelihood ratio is again (3), now at $a = \tau$. Exponential-Lévy is exactly the class in which a scalar statistic stays sufficient at stopping times, not merely fixed ages (Küchler and Sørensen, 1989; Küchler and Lauritzen, 1989; Küchler and Sørensen, 1998); outside it, the insufficiency (Theorem 1(i)) still holds with the raw since-adoption record, which carries its own length, in place of (T_a, a), while the coordinate-wise irreducibility and the finite-dimensional closure (Theorem 1(ii)–(iii)) are statements about the summary state and are stated at exponential-Lévy scope.

(ii) When she abandons it. She keeps her model as long as the observed summary looks ordinary under it, and abandons it the first time the summary looks too unlikely. Fix a tolerance $\alpha \in (0, 1)$, the size of the retention test. The *ordinary range* is the set of outcomes her model finds most likely, carved out to carry probability $1 - \alpha$:

$$A_\alpha(a) = \big\{x : f_a^{\hat\theta}(x) \ge c_\alpha(a)\big\}, \qquad c_\alpha(a) \text{ chosen so that } \mathbb{P}^{\hat\theta}\big(T_a \in A_\alpha(a)\big) \ge 1 - \alpha. \tag{4}$$

The region carries mass exactly $1 - \alpha$ in the continuous channel and is the smallest probability-ordered integer interval of at least that mass (the Sterne region) in the discrete one, cutoff ties resolved toward the larger count, the right-continuous convention of Lemma D.1, so the displayed cutoff determines the region. It is the highest-density region: interval acceptance is the general form of an exact test (Lemma A.2), and among size-α regions the primitive pins this one, the continuous-time limit of Ortoleva's one-shot test, uniquely so absent a flat plateau of the held density at the cutoff level (Lemma A.3; Appendix A), taken closed, abandonment strictly outside (the continuous boundary is null, Lemma A.3). Proposition 1 is that chain assembled; Lemma D.7 (Appendix D.5) records which downstream objects survive any selection with the same asymptotic endpoints. A larger α tightens the band; "tolerance α" is the type label throughout, decreasing in α.

She abandons her model at the first age, after a probation onset $a_0 \geq 0$, at which the summary leaves the ordinary range:

$$\tau = \inf\{a > a_0 : \; T_a \notin A_\alpha(a)\}. \tag{5}$$

The onset a_0 is a primitive of the recording environment, not a behavioral parameter: positive on a diffusive record, where monitoring from $a = 0$ would reject even the truth at arbitrarily small ages (the law of the iterated logarithm; Lemma A.6), and zero on a counting record, where the floor at zero rules out an early lower exit though a disaster can still trigger an upper one. The holding-time results of Section 3.3 rest on it. In the two channels we use throughout, the range is an explicit interval around what the model expects:

$$\text{disputed trend:} \qquad A_\alpha(a) = \Big\{ \big|\bar{X}_a - \hat{\mu}\big| \leq z_\alpha \, \sigma/\sqrt{a} \Big\}, \qquad z_\alpha = \Phi^{-1}\Big(1 - \tfrac{\alpha}{2}\Big); \tag{6}$$

$$\text{disputed crash rate:} \qquad A_\alpha(a) = \big\{ \, l_\alpha(a) \leq N_a \leq u_\alpha(a) \, \big\}, \tag{7}$$

with $\bar{X}_a$ the average growth observed, $\hat{\mu}$ the held drift, N_a the disaster count, and $\hat{\lambda}$ the held crash rate. The crash-rate endpoints have no elementary closed form, $A_\alpha(a)$ is the integer *Sterne region* (Sterne, 1954; Crow and Gardner, 1959), the smallest probability-ordered interval of held-model mass ($N_a \sim \text{Poisson}(\hat{\lambda}a)$) at least $1 - \alpha$, with explicit leading shape and step construction in Section 3.3 (Lemma D.1).

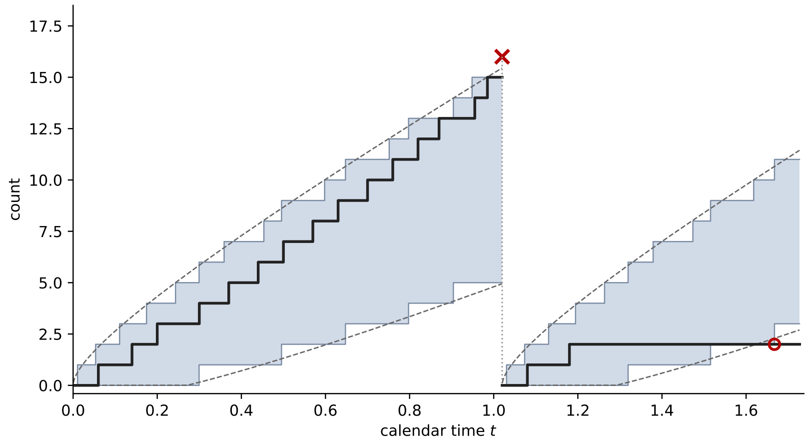

Figure 1: One realized record against its band, calendar time on the horizontal axis: counts since adoption at the held rate $\hat{\lambda} = 10$ (solid), the probability-ordered Sterne region of (7) at $\alpha = 0.10$ (shaded staircase), and the Gaussian guides $\hat{\lambda}a \pm z_\alpha\sqrt{\hat{\lambda}a}$ (dashed), each per regime since its adoption. The first regime ends in an upper exit at a disaster ($\times$); the record restarts at zero on a fresh staircase, whose acceptance region at adoption is $\{0\}$; the second ends in a lower exit at a predictable floor step. The smaller test size carries the wider band.

Figure 1 illustrates one realized spell in the crash-rate channel through both deaths: over the ceiling at a disaster, restarted at the singleton region $\{0\}$, and later overtaken by the rising floor. The staircase is the ordinary range, the solid path the accumulating record, and the restart after rejection begins a new spell under the newly adopted model. Three features of the dynamics are immediately visible. First, the acceptance region evolves with age rather than remaining fixed. Second, the record resets after each abandonment, so every spell begins from a fresh origin. Third, upward and downward exits have different geometry in the crash channel, foreshadowing the directional asymmetry of Section 3.3.

The figure explains how the rule operates once an acceptance region has been specified. It does not explain why this particular region is the correct one. Exactness alone leaves many size-α acceptance regions. The next proposition shows that the hypothesis-testing primitive itself selects the probability-ordered region.

Proposition 1 (The band from the primitive)**.** *Fix an age a and a held model whose predictive law for T_a is unimodal—with, in the continuous channel, no level set of its* *density carrying mass; in the counting channel the atomic law of* (A.2)*. Among Borel acceptance regions of held-model mass at least* $1-\alpha$*:* (i) *exactness fixes the mass floor;* (ii) *every exact test is weakly dominated by an interval test (Lemma A.2);* (iii) *the fine-resolution limit of Ortoleva's cell rule is the density-cutoff region (Lemma A.3(i)), which, calibrated to the mass floor, is the unique minimal region up to null sets (Lemma A.3(ii))—an interval, by unimodality—and, in the counting channel, the probability-ordered Sterne region (Lemma A.3(iii)). The band of Definition 1 is thus pinned by the primitive under these regularity conditions, not by exactness alone: alternative exact-size selections—equal-tailed, unbiased—exist and are not generated by the primitive.*

Proposition 1 separates two distinct ingredients. Exactness fixes the probability mass of the acceptance region, but it does not determine its geometry. The hypothesis-testing primitive selects the probability-ordered region, which in the continuous channel is unique up to null sets and in the discrete channel is Sterne's conservative region, exact size on a lattice being available only by randomizing at the two endpoints. The interval form of these regions is not a modeling choice; it is the general form of an exact test of the held model: in either canonical channel the interval-acceptance tests form an essentially complete class, every admissible test almost everywhere an interval test (Lemma A.2, Appendix A; Birnbaum, 1955; Matthes and Truax, 1967). Either way one scalar remains, and the band's shape in age is inherited from the held model's own fluctuation law, not chosen: the $a^{-1/2}$ contraction of (6) (a $\sqrt{a}$ widening of the cumulative record's band) and the integer staircase of (7). Running that per-age test at every age is the per-test rule itself, and the dynamic objects (the first-exit time (5), its probation and right-continuity conventions, the hazard and its atoms) are its disciplined continuous-time embedding, the classical *repeated significance test* run continually (Armitage et al., 1969; Siegmund, 1985), a convention adopted here rather than a limit theorem. The form is thus general; the dynamics are not: what a record can do, drift to its edge or jump across it, is channel content, and the results below are sharp because the two records are canonical.

(iii) What she does next. On abandoning her model she re-ranks the menu in light of the track record and adopts the new front-runner:

$$\rho'(\theta) \propto L_\theta\big(X_{(\sigma,\sigma+\tau]}\big)\,\rho(\theta), \qquad \hat{\theta}' = \arg\max_\theta \rho'(\theta), \tag{8}$$

where $L_\theta(X_{(\sigma,\sigma+\tau]})$ is the likelihood under model θ of the just-ended regime's record, the segment from adoption σ across the holding time τ; we write X_σ for this segment where the regime is clear—the record, not the time-σ value. The new model is judged only on data from its adoption onward, the *test record* restarts, while the *ranking* ρ carries forward and moves only at surprises. Re-ranking on each regime's record in turn equals re-ranking on the entire history at once (Appendix A).

Definition 1 (Surprise-governed belief revision)**.** *Given a menu* Θ *with pairwise locally equivalent laws, an initial ranking* ρ_0*, a tolerance* $\alpha \in (0,1)$*, a recording resolution* δ*—entering the continuous-time objects only through the fine-resolution band selection of Proposition 1—and a fixed probation interval* $a_0 \geq 0$*, the investor holds* $\hat{\theta}_t = \arg\max \rho_t$*, initialized at* $\arg\max \rho_0$*; she monitors the held model through the summary* T_{a_t} *of its track record and abandons it at the first exit* (5) *from the ordinary range* (4)*; and at each revision she re-ranks and re-adopts via* (8)*, restarting the track record. The resulting process* $(\hat{\theta}_t, T_{a_t}, a_t, \rho_t)$*—held model, monitored record, age, and ranking—is the* surprise-governed belief process*.*

Two features matter. The investor carries two memories: the track record she tests against restarts at each revision, a new model is judged on its own tenure, while the ranking ρ remembers everything, since learning which model is right is the work of a lifetime. The non-Bayesian feature is thus not a reset second-order prior (ρ accumulates full-history likelihood by (8)) but that the held model is the maximum-a-posteriori selection of ρ, not a posterior mean, revised only at exits of (5). Every clause of Definition 1 is Ortoleva's one-shot rule carried into continuous time or forced by a degeneracy the alternative would create (Appendix A), except one, which reflects a genuine choice.

That choice is how the tolerance α is interpreted once the model is tested again and again: as a *per-test* tolerance, the same size-α test at every instant, consulting only the held model, the literal continuous-time reading of Ortoleva's rule, under which a model is abandoned when *its own* record looks too unlikely, not when a rival fits better. On that interpretation, the rule applies one *fixed* evidentiary standard that never rises with tenure, and the recurrence of Section 3.3 is its *signature*, not a defect of myopia.[6]

[6] The alternative reads α as a *lifetime* tolerance, bounding the probability of *ever* wrongly abandoning a correct model. Weinstein (2017) shows this bound is attainable by an α-spending rule

The rule is now fully specified: a menu and a ranking, a band the primitive itself selects, one behavioral scalar, and a restart at every revision. What the investor does is settled. The law of when she will revise is not, and pinning that law down takes more than her current belief.

3 The enlarged state and the endogenous hazard

We now ask what information is required to predict when the rule revises. The answer enlarges the state, from the held belief to the surviving record under which it has endured: once that enlarged state is identified, the endogenous belief hazard follows as an implication rather than an assumption, and the law of the holding time follows from the hazard.

3.1 The record is a state variable

Once revision is governed by a test on the since-adoption record, the state a forecaster carries is her belief together with the record under which it has survived: the revision law, and the hazard derived from it in §3.2, is a coordinate of that enlarged state, not recoverable from any *fixed-prior* Bayesian posterior on the public data. Its testable face is concrete: two forecasters reporting the same forecast today can revise differently tomorrow, their forecasts having survived different records.

Theorem 1 (The surviving record is a state variable)**.** *Consider any rule of Definition 1 with a menu satisfying Assumption 1.*

(i) Insufficiency. *The conditional law of the revision time* (5) *is not a function of $\hat{\theta}_t$ alone, nor measurable with respect to $\sigma(\beta_t)$ for any posterior β_t formed by updating a fixed prior on the public filtration $\mathcal{F}_t$—a reset-free functional of the public path, common across holders.*

and keeps the agent approximately dynamically consistent, the arbitrage against her is of order the lifetime rejection probability. But attaining it with power against the menu draws the test to the rival models' likelihoods (Lemma A.7; by the admissibility characterization cited in its proof, any time-uniform guarantee rests on a nonnegative supermartingale), the model-switching rule of the paradigm-shift learner (Cho and Kasa, 2015; Ba, 2026), and forfeits surprise-governance. The per-test reading instead holds the size fixed at every instant, so the lifetime rejection probability is one: this is the regime the α-spending rule avoids, where that arbitrage bound is vacuous (Lemma A.7).

(ii) Sufficiency and irreducibility. *It is a measurable function of* $(\hat{\theta}, T_a, a)$*, finite-dimensional by Assumption 1, and—when the band is nondegenerate in age and the monitored statistic discriminates the menu, as in both canonical channels—of no two of* $\hat{\theta}, T_a, a$*.*

(iii) Closure. *Whenever the since-adoption statistic is* autonomous*—its law of motion determined by* (T_a, a) *alone, as in both canonical channels, whose stationary independent increments make the post-adoption record a fresh copy of the environment—*$(\hat{\theta}, T_a, a, \rho)$ *is a Markov process—piecewise-deterministic (Davis, 1984) in the crash channel, a boundary-reset diffusion in the trend channel—whose revisions are the successive first exits of* T_a *from* (4)*, each post-revision ranking* ρ' *a function of* (ρ, T_τ, τ) *by the stopped sufficiency of Assumption 1 (cf. Küchler and Sørensen, 1997).*

Each part has a settled lineage; the rule of Definition 1 is what joins them. Carrying the age since a reset to recover a Markov state is the method of supplementary variables (Cox, 1955) (parts i, iii); the collapse of the record to a finite summary (T_a, a) is the recursive sufficiency of an exponential family (Küchler and Sørensen, 1989) (part ii); the joined state, revision its first exit, is, in the crash channel, a piecewise-deterministic process (Davis, 1984), deterministic between disasters, and in the trend channel a boundary-reset diffusion, both closed under the menu (Küchler and Sørensen, 1998) (part iii). What is new is their composition: the age is the time since the last *model abandonment*, the sufficient statistic is the *test record* the held model is judged on, and the first exit is its *rejection*. The one place this leaves classical ground is part (i): the belief is not a sufficient statistic for its own revision, inverting the Bayesian benchmark in which the posterior is its own state (Epstein and Le Breton, 1993).

Part (i) is where the two forecasters of the introduction part: sharing every public-history posterior, they still face different hazards, because the hazard turns on the private adoption date σ (equivalently the age $a = t - \sigma$) that no public-filtration posterior carries. The enlargement to $(\hat{\theta}, T_a, a)$ is a property of the record-resetting rule itself, not of the Sterne-band implementation below.

Part (iii) licenses the dynamics that follow: because the enlarged state $(\hat{\theta}, T_a, a, \rho)$ is a bona fide Markov process, its evolution between and at revisions, its long-run behavior, and its cross-sectional law are all well-posed objects rather than artifacts of an ill-specified state.

The economic content is that a belief carries a hidden coordinate. Two investors agreeing on the held model need not agree on how fragile that agreement is: the one whose record sits at the edge of its band revises on the next surprise, the one in the interior does not. Consensus can thus mask fragility, a population agreeing on a forecast may be uniformly robust or one surprise from fracture, and the level alone cannot tell them apart. The prediction is cross-sectional and testable on forecaster panels (the Survey of Professional Forecasters, the Michigan survey): conditional on the same current forecast, the probability of revising should depend on regime age and run length, how long the forecast has been held and the cumulative surprise it has absorbed, not on the level. By Lemma H.1 of the working-paper version this is an implication for the *observed* hazard conditional on the full inference state (record, age, and ranking): the self-re-adoption thinning that makes reported revisions sparse is itself stakes-free, but it also lets a reported spell differ from the internal regime age, so the panel test conditions on reported duration and cumulative surprise as *proxies*, exact when self-re-adoptions are absent, a structural restriction rather than a regression prediction otherwise.

What the theorem excludes, exactly: the held model alone, the reported forecast, the pair of model and spell age without the record's position, and every posterior formed from a *fixed* prior on the public filtration. What it does not exclude is the state itself. The carried ranking ρ_t, likelihood-based but frozen between revisions, enters the *re-selection* and not the pre-exit hazard, so consulting it is not a hidden-state relabeling of the belief. Nor is the excluded configuration one a modeler must initialize: the population realizes it along its own path, permanently in the trend channel and recurrently in the crash channel (Corollary 2, §3.3).

3.2 The revision hazard

The enlarged state of Theorem 1 governs a rate: the rate at which the investor abandons her model, the endogenous belief hazard the title names. Its absolutely continuous part is

$$h(a) = \lim_{da \downarrow 0} \frac{1}{da}\, \mathbb{P}\big(\tau \in (a, a+da] \mid \tau > a\big), \tag{9}$$

the age hazard of the holding time τ of (5). In the crash channel the band endpoints (7) step at discrete ages, so revision also carries point masses and the full object is a *measure*, the continuous hazard (9) together with atoms at the scheduled band-step dates (Appendix C).

The hazard is not a primitive: it is taken from the first-exit law of the rule (4)–(5), not posited. One distinction the measure hides drives everything below. The rate $h(a)$ is that of the *marginal* holding-time law, the law of τ averaged over records, so it is a function of the reported coordinates $(\hat{\theta}, \alpha, a)$ and closes there, per tolerance fibre. The *conditional* exit propensity given the full state, how close the record T_a sits to the band edge, is a different object: two age-a holders of the same $(\hat{\theta}, \alpha)$ share $h(\hat{\theta}, \alpha, a)$ yet, sitting at different T_a, stand at different distances from revision. That gap is the insufficiency of Theorem 1(i) expressed on the hazard itself. The split takes a different form in each channel, a state-dependent jump rate at the inaccessible crash exits and a predictable boundary flux in the trend channel; Appendix C records it together with the measure above.

Displayed, the gap is the theorem's sharpest small statement:

Corollary 1 (The hazard closes; the fragility does not)**.** *Fix a rule of Definition 1 in either canonical channel and a held model $\hat{\theta}$.*

(i) *The holding-time law of τ from adoption—hence the marginal hazard h of* (9) *and, in the crash channel, the atom dates and masses of the hazard measure—is determined by $(\hat{\theta}, \alpha)$ and the recording environment: at every age it is a function of the reported coordinates $(\hat{\theta}, \alpha, a)$.*

(ii) *The $\mathcal{F}_t$-conditional law of the next exit admits no $(\hat{\theta}, \alpha, a)$-measurable version. In the crash channel, at any age at which the band has width at least two and does not step, the ceiling and the slack-two record positions carry positive survivor mass, and their conditional upper-exit probabilities over $(t, t+h]$ are $\lambda^\star h + o(h)$ and $o(h)$. In the trend channel, at any age past probation and for every $\eta \in (0, \frac{1}{2})$ there is a window length at which the conditional exit probability exceeds $1 - \eta$ on a band-edge neighborhood of the record position and falls below η on a center neighborhood, both of positive survivor mass; the conditional window-exit probability is therefore nondegenerate given $(\hat{\theta}, \alpha, a)$.*

Part (i) is the closure an econometrician uses: the hazard estimated from reported coordinates is a well-posed function of them. Part (ii) is why that function is not the fragility any one investor carries, and why the failure is not an artifact of null states: the crash witnesses are atoms of positive survivor mass, and the trend statement is a nondegeneracy of the conditional probability under the survivor law itself. The proof is in Appendix C.

Within the class the rule adds no primitive beyond the scalar α: a cost-based hazard is priced off a cost distribution that is itself a free object, calibrated to fit, while the testing hazard inherits its shape from the null's own fluctuation law, a prediction rather than a calibration target. Stakes-invariance is then the separating restriction between the two microfoundations: the trigger consults the record, never the payoff, so an exogenous shift in what the forecaster stands to lose moves every cost-based member of the class and leaves the testing member fixed (Proposition G.1 of the working-paper version); the identification problem this raises, and the hybrid nesting that spans the class, are developed there. Every non-representability statement below is relative to a named class, a fixed-prior Bayesian's held belief, the conditional expectation given $(\mathcal{F}_t)$ of a single prior over the menu, and is established over a positive time interval, never by conditioning on an instantaneous event.

3.3 Location consolidates, activity does not

Two forces act on a held model over a long career. The ranking ρ accumulates evidence and so its argmax settles; the test, applied afresh at every instant, keeps finding the held model surprising. The first force is consolidation, the second recurrence, and they do not cancel: the model a believer holds stops changing in *identity* while never ceasing to change in *tenure*.

Proposition 2 (Location consolidates, activity does not)**.** *Run the rule of Definition 1 at a fixed tolerance $\alpha \in (0,1)$ in either canonical channel—the disputed-trend record (a Brownian running mean) or the disputed-crash record (a Poisson count)—and suppose the menu has a unique closest model in relative-entropy* rate, *$\theta^\circ = \arg\min_\theta \kappa(\theta)$ with $\kappa(\theta) := \lim_{t\to\infty} t^{-1}\mathrm{KL}(\mathbb{P}^\star_t \| \mathbb{P}^\theta_t)$—the finite object here, the full-path divergence being infinite—with $\rho_0(\theta^\circ) > 0$.*

(a) Location. *Under the chained update (Lemma A.4) there is an a.s. finite calendar time T° such that $\sigma_k \geq T^\circ \Rightarrow \arg\max_\theta \rho_{\sigma_k}(\theta) = \theta^\circ$ almost surely.*

(b) Activity. *For any held model the holding time τ of* (5) *is $\mathbb{P}^\star$-almost surely finite, and, re-adopting the model after each exit, the agent abandons it infinitely often,*

$\mathbb{P}^\star$-almost surely. When the menu contains the truth, every regime begun after T° starts with θ° and is abandoned again.

Both halves are established facts; the rule of Definition 1 is what holds them together. Part (a), proved by the strong law for the log-likelihood ratio, makes no reference to whether or when a revision occurs, that is part (b). It places the held-model *location* where the misspecified-learning literature places the *limit* of belief: the Kullback–Leibler projection of the truth (Berk, 1966; Esponda and Pouzo, 2016; Fudenberg et al., 2021; Frick et al., 2023). When the truth is on the menu, θ° is the truth and the ranking *merges* with it, the condition $\rho_0(\theta^\circ) > 0$ is the grain of truth (Kalai and Lehrer, 1993) under which opinions merge (Blackwell and Dubins, 1962), which is why positive weight on the closest model, not full support, suffices for part (a). Any residual gap under a misspecified menu is the menu's, not a dynamic error of the testing, and at a tie the front-runner consolidates to the closest set and cycles within it.

The economic content is that a correct model is right on average yet locally surprising: tested at a fixed size and monitored without limit, its own ordinary fluctuations eventually carry the record out of band. This is "sampling to a foregone conclusion", the fact that a fixed-level test on accumulating data rejects a true null with certainty, by the boundary-crossing of the law of the iterated logarithm (Anscombe, 1954; Armitage et al., 1969; Robbins, 1970; Robbins and Siegmund, 1970); here it is the engine of the dynamics rather than a defect. In its modern form this is the object that safe, anytime-valid inference is built to correct: holding the test size fixed at every instant is exactly what test-martingales and e-processes avoid (Ramdas et al., 2023; Grünwald et al., 2024), and it is what sends the lifetime rejection probability to one. Our agent forgoes that correction: the canonical rules attaining a lifetime bound mix over the rival models (Lemma A.7), the one thing surprise-governance gives up, so the recurrence is the signature of that choice. It rests on the fluctuation law of each record, not on the Markov closure of Theorem 1 alone; we do not assert recurrence for an arbitrary time-homogeneous data-generating process. So a correct-on-average belief is revised forever, correctness buying no permanence, and a correct-on-average population keeps revising, its members agreeing on the model yet never on how close each stands to abandoning it (Section 4). We claim almost-sure finiteness and infinite recurrence, not a finite expected holding time, which in the trend channel can fail (§3.4).

Consolidation and recurrence combine into the paper's answer to a natural objection: that the two-holder comparison behind Theorem 1(i) is initialized rather than generated. It is generated, in both canonical channels, by opposite routes. Heterogeneity in tolerance alone staggers the population's adoption dates along one path; in the trend channel the staggering, once begun, never heals, while in the crash channel the same common disasters that synchronize aggregate crossings can re-align a pair, so the staggering recurs rather than absorbs:

Corollary 2 (The witness is realized)**.** *Let two investors run the rule of Definition 1 on the common path (Assumption 2) with distinct tolerances, under the hypotheses of Proposition 2 with the truth on the menu. Then $\mathbb{P}^\star$-a.s.:*

(i) Disputed trend *(probation $a_0 > 0$): there is a finite time S such that for every $t > S$ both investors hold $\hat{\theta}_t = \theta^\star$, every posterior formed from any fixed prior on $\mathcal{F}_t$ is one random variable shared by the two, and their regime ages differ. On each window extending from an exit date $s > S$ of either investor to the earlier of $s + a_0$ and the other investor's next exit—nonempty open windows recurring with infinite total measure, $\mathbb{P}^\star$-a.s.—the $\mathcal{F}_t$-conditional laws of the two next-exit times differ at every t: the just-revised investor's law assigns no mass to $(t,\, s + a_0)$, the other's assigns positive mass there; the identity of the sooner-supported law changes hands infinitely often along the path, each investor's exits opening windows on which the other holds it.*

(ii) Disputed crash rate*: there are unboundedly many nonempty open intervals, each within a stretch of differing adoption dates, on which both investors hold $\hat{\theta}_t = \theta^\star$, every fixed-prior posterior on $\mathcal{F}_t$ is shared, their regime ages differ, and the $\mathcal{F}_t$-conditional laws of the two next-exit times differ at every t: one investor's count sits at her ceiling and she exits at the first disaster, $\mathbb{P}\big(\tau \in (t, t+h] \mid \mathcal{F}_t\big) = 1 - e^{-\lambda^\star h}$, while the other's is strictly interior and requires at least two, $\mathbb{P}\big(\tau \in (t, t+h] \mid \mathcal{F}_t\big) \le 1 - e^{-\lambda^\star h} - \lambda^\star h\, e^{-\lambda^\star h}$, for every h in the bands' step-free gap. At each interval's opening the exposed, at-ceiling investor is the just-revised one.*

In both channels the tolerance pair is fixed throughout, so the difference in the conditional laws is carried by the ages and records; and the channels invert the exposed party—just revised means immune through probation under a disputed trend, and one disaster from rejection under a disputed crash rate.

The corollary discharges what the theorem's proof initializes. From a common start, one prior, one date, one path, the model manufactures its own witness: two investors holding the same model, sharing every fixed-prior public posterior, carrying different conditional revision laws, forever after a finite time in the trend channel and recurrently without bound in the crash channel. Ada and Bea of the introduction are realized, not assumed. The engines are two desynchronization lemmas (Lemmas B.1 and B.2, Appendix B). In the trend channel the narrower band exits strictly first unless the record breaches the *wider* band at the probation onset, an event of menu-uniform probability below one per regime, and once the adoption dates part no two later exits ever coincide: the staggering is absorbing. In the crash channel synchronized phases end in a lone exit just as surely, but a common disaster that finds both counts at their ceilings exits both at once and, past consolidation, re-synchronizes the pair, the pairwise face of the boundary-mass synchronization behind the aggregate-jump condition of Section 4; the staggering is recurrent instead. Either way the witness windows recur with the tolerance pair fixed, so the law difference is carried by the record coordinate the belief omits. The law difference is claimed where it is explicit, on the probation windows and the ceiling-versus-interior intervals; at other desynchronized dates the enlarged states differ while their conditional laws are not compared. The channels' inversion of the exposed party is the disaster asymmetry of §3.4 at the level of a pair: a fresh model cannot be rejected during a trend probation, and can be rejected by the very next disaster under a disputed crash rate.

3.4 The holding-time law

Recurrence says a model is abandoned; the holding-time law says when. In the crash channel the survival has an exact finite matrix-product form, an age-inhomogeneous phase-type construction, the band dimension changing across step ages (Lemma D.3, Appendix D.4): it is the ex-ante single-spell law, distinct from the pathwise push-forward that moves the realized cross-section of Section 4. What the population dynamics inherit is its *tail*, and that tail is heavy in both channels, heavy enough that the mean quiet spell can be infinite, the continuous-monitoring descendant of Blackwell and Freedman (1964)'s observation that the two-sided $c\sqrt{n}$ exit of a coin-tossing record has finite mean only for $c < 1$. The recurrence itself is the continuous-monitoring form of the classical fact that repeated significance testing at fixed size rejects a true hypothesis with probability one (Armitage et al., 1969; Siegmund, 1985).

Theorem 2 (Holding-time law)**.** *In the setting of Proposition 2, where the holding time τ is a.s. finite:* (a) *holding the correct model, one exponent governs the tail of τ in both channels—the principal Dirichlet eigenvalue $\theta(\alpha) > 0$ of the Ornstein–Uhlenbeck generator $\frac{1}{2}\partial_{vv} - \frac{1}{2}v\,\partial_v$ on $(-z_\alpha, z_\alpha)$: in the trend channel the tail is an exact power law,*

$$\mathbb{P}(\tau > t) \sim C\,t^{-\theta(\alpha)} \qquad (t \to \infty), \tag{10}$$

and in the crash channel

$$\mathbb{P}(\tau > t) = t^{-\theta(\alpha)\,(1+o(1))} \qquad (t \to \infty), \tag{11}$$

so the mean quiet spell is finite when $\theta(\alpha) > 1$ and infinite for sufficiently tolerant investors (wide bands, $\theta(\alpha) < 1$); (b) *holding an incorrect model, τ is finite, with, conditional on no earlier false alarm—no exit of the noise component across its fixed band before the deterministic detection horizon, the event each channel's proof makes precise—detection scale*

$$a^\star_{trend} \approx \frac{z_\alpha^2\,\sigma^2}{(\mu^\star - \hat{\mu})^2}, \qquad a^\star_{crash} \approx \frac{z_\alpha^2\,\hat{\lambda}}{(\lambda^\star - \hat{\lambda})^2}, \tag{12}$$

the scale growing as the inverse square of the misfit, so larger misfits are detected sooner—in the local-misspecification regime an order-one estimate of the scale, $\log\tau/\log a^\star \to 1$ *conditional on no earlier false alarm, not a sharp limit—so that for two held models equidistant in Kullback–Leibler* rate *from the truth the conditional detection scales agree to leading order, $a^\star \sim z_\alpha^2/(2K)$ in both channels—K the common Kullback–Leibler rate—the narrower band detecting faster only at the next order.*

Part (a) is exact and unconditional for the correct model; part (b) is a conditional, detection-scale statement under local misspecification—the two are read at different scales.

Which of these conclusions depend on the Sterne selection? Only the calendar. Any exact-size nested family of intervals with the same asymptotic standardized endpoints $\mp z_\alpha$, and in the symmetric limits of both channels the natural selections, highest-density, equal-tailed, central, share them, has the same tail exponent, the same finite/infinite-mean threshold, and the same recurrence: the trend problem by a domain sandwich and the continuity of λ_0, the crash tail because the exit lemma's hypotheses consume only the asymptotes. What moves with the selection is the schedule, the step dates, the predictable atoms, the finite-age band geometry (Lemma D.7, Appendix D.5).

One killed principal eigenvalue $\theta(\alpha)$ sets the tempo of the recurrence of Proposition 2(b) through the heavy tail (the trend statistic is the Ornstein–Uhlenbeck process exactly; the standardized crash count converges to it), with the crash channel carrying its survival through the matrix-product law of Lemma D.3. The two parts are the rule's *error profile*: a correct model held long, a wrong one caught fast, and widening the band lengthens both: tolerance trades false alarms against detection lag.

The individual law is now complete: the state, the hazard it governs, and the two deaths that end a spell. What one investor does over one spell, and how likely she is to do it at every moment, is settled. What a population of such investors does to the cross-section of beliefs is not, and the question changes scale: the object is no longer one hazard but the flow of many records, watching one history through different tolerances.

4 The cross-section is a pushforward

Because investors watch the same data and differ only in their tolerance for surprise, the cross-section of beliefs is the *exact* image of the tolerance law under the common flow, with no law of large numbers and no equilibrium average to solve for. Smaller α is the more tolerant investor, the wider band. The aggregate dynamics therefore carry no behavioral parameter beyond the test size α; the heterogeneity they carry enters through the tolerance law G, a primitive that quantitative work must specify or estimate. Three kinds of statement follow, kept apart throughout. *Support-structure* predictions, which dates can carry upward jumps in consensus fear and which downward, hold for every admissible G, fixed by the rule alone. *Magnitudes*, the boundary mass, the revision rate, the atom sizes, are G-functionals: identities that identify G from observed revision mass rather than free predictions. *Shape-conditional* results carry their conditions by name, consensus continuity under atomless G the leading case. This section states the pushforward and derives its immediate dynamics; the cross-sectional and asset-pricing patterns it implies are developed in the working-paper version.

Index an investor by a type $\xi \in \Xi$, a tolerance $\alpha(\xi)$, and a fixed menu and prior, drawn from a cross-sectional law $G \in \mathcal{P}(\Xi)$. All investors observe the same path of the economy's process X and run the rule of Definition 1 on it; types differ only in the band width $z_{\alpha(\xi)}$, so they exit and re-select at different times along the one path.

Assumption 2 (Common path)**.** *Investors observe the same public record and differ only in tolerance, holding a fixed menu and prior; the cross-section carries no idiosyncratic agent-level noise. The tolerance law G has support bounded away from the degenerate band—$\alpha(\xi) \leq \bar{\alpha} < 1$ for G-a.e. ξ—so every type's standardized critical value is bounded below, $z_{\alpha(\xi)} \geq z_{\bar{\alpha}} > 0$; the raw band width still varies with age, contracting like $a^{-1/2}$ in the trend statistic, with single-integer Sterne regions possible at early ages. In the trend channel the fixed probation onset of Definition 1—not the support bound on α—gives each type the deterministic floor $\tau_k \geq a_0 > 0$; in the crash channel holding times are a.s. positive but* not *deterministically bounded below ($a_0 = 0$, since a disaster can carry an upper exit arbitrarily close to adoption), and local finiteness of the aggregate flux comes channel-specifically: finitely many disasters on compacts, and a uniform bound on the G-average number of scheduled floor-step exits per compact—the flux* measure *locally finite, not the population-level step dates finitely many (proof of Corollary 3).*

The model's one structural commitment for aggregation is that all cross-sectional heterogeneity sits in the test threshold, not in information or private beliefs (the overlapping-generations extension adds birth date as the only other axis). This is what makes the pushforward exact, the cross-section is the image of a primitive, not a calibrated object, so which patterns appear follows from G and the environment, not from a tuned parameter; G itself is the object a quantitative application must specify or estimate. We maintain Assumption 2 throughout; two further *environment-specific* assumptions, local misspecification (Assumption D.1) and demographic turnover, are invoked only where needed. The results below divide by what they condition on. Proposition 3 is an identity in G: given the common path it holds for every tolerance distribution and at every date, a statement about the *map* from primitive to cross-section, valid along any transition path, presuming no stationarity. The statements that follow are conditional exactly as their hypotheses say: turnover and the weighting license the terms of Corollary 3; atomless G governs which parts of the flux are continuous; the local-misspecification regime (Assumption D.1) scopes every re-selection-direction claim; and quantitative patterns are G-conditional, they follow once G is specified and change with it. Statements about stationary cross-sections are separate results with their own existence hypotheses, not consequences of the push-forward identity. The aggregate-jump claims are additionally conditional on (17) together with a positive-measure model-changing crossing set, verified per application. Two revision counters run through the paper: the *internal* trigger, every band exit, self-re-adoptions included, whose recurrence Proposition 2 establishes; and the *observed* model-changing count, its (U)-thinning (Lemma H.1 of the working-paper version). Past consolidation with the truth on the menu, exits recur while observed changes can vanish; every aggregate statement that turns on observable revision is stated in the thinned counter.

Write $s_t(\xi) = (\hat{\theta}_t, \alpha(\xi), a_t) \in \mathcal{S}$ for the investor's *held-model state* at t, the held model, the type's tolerance, and the age of its test record, $\mathcal{S} = \Theta \times (0,1) \times \mathbb{R}_+$; the tolerance is a coordinate of the state because it sets the band the record is tested against, and it is constant along each path ($\dot{\alpha} = 0$), so it adds no transport term and the flow acts fibrewise in α. Write $\bar{s}_t(\xi) = (\hat{\theta}_t, \alpha(\xi), a_t, R_t, \rho_t) \in \bar{\mathcal{S}}$ for the *full* belief state, which adjoins the since-adoption record summary $R_t = T_{a_t}$ that drives the boundary and the carried-forward ranking ρ_t, with $\bar{\mathcal{S}} = \mathcal{S} \times \mathcal{R} \times \mathcal{P}(\Theta)$. Let $\bar{\Psi}_t : \Xi \to \bar{\mathcal{S}}$ be the flow the common path induces on full states and $\Psi_t = \pi \circ \bar{\Psi}_t$ its held-model projection, $\pi : \bar{\mathcal{S}} \to \mathcal{S}$. The cross-section of beliefs we report is the held-model marginal

$$m_t := \Psi_t \# G = \pi \# \bar{m}_t \in \mathcal{P}(\mathcal{S}), \qquad \langle \varphi, m_t \rangle = \int_{\mathcal{S}} \varphi \, dm_t = \int_{\Xi} \varphi\big(\Psi_t(\xi)\big)\, G(d\xi), \quad (13)$$

the projection onto $(\hat{\theta}, \alpha, a)$ of the full-state cross-section $\bar{m}_t := \bar{\Psi}_t \# G$; where an aggregate uses only $(\hat{\theta}, a)$ it takes the further marginal, but the cross-section itself carries the tolerance coordinate throughout. The aggregates we form, fear, dispersion, the consensus, are test functions $\varphi(\hat{\theta}, a)$ of these coordinates, so m_t carries them. In particular the *belief functional* $w(\hat{\theta})$, the value the held model reports, the perceived intensity $\hat{\lambda}$ in the crash channel and the perceived drift $\hat{\mu}$ in the trend, is one such test function, so the consensus belief $\langle w, m_t \rangle$ and the cross-sectional disagreement $\mathrm{Var}_{m_t}(w)$ used below are carried by m_t. But the boundary-crossing and re-selection that *move* the cross-section depend on the record and ranking (R_t, ρ_t), which are not coordinates of $\mathcal{S}$. The forward equation of §4.1 is for that reason the *pathwise* time-derivative of the deterministic pushforward $\Psi_t\#G$ (Proposition 3), not a closed generator equation on $\mathcal{S}$: the full state on $\bar{\mathcal{S}}$ is Markov, piecewise-deterministic in the crash channel, a diffusion with boundary reset in the trend, but along the common path (R_t, ρ_t) are deterministic functions of ξ and t, so the projected dynamics follow from it.

Proposition 3 (Pushforward cross-section)**.** *For every bounded measurable φ,* (13) *holds pathwise: under Assumption 2, m_t is the exact image of the tolerance law G under the common flow Ψ_t, with no continuum law of large numbers invoked. The realized non-revision fraction at t—the G-mass of types whose held model is not revising in the instant at t—is a pathwise functional of X, an instantaneous quantity, distinct from the survival of a holding spell from adoption to age a. The ex-ante population survivor fraction to age a is the G-mixture $\int_\Xi S_\xi(a)\, G(d\xi)$ of the single-type laws* (C.2)*, not the single-type $S(a)$; the realized and ex-ante objects agree only in expectation over paths.*

The force of Proposition 3 is that heterogeneity, not noise averaging, generates the cross-section: each type is a deterministic function of the same history, so the belief distribution is a change of variables away from the tolerance distribution, and every later statement can be made pathwise without solving for an equilibrium average.

4.1 What moves an aggregate

Between revisions a belief state only ages; at a revision it jumps. Lifting this to the population along the common path, differentiating $\Psi_t\#G$ along the common path (Proposition 3) gives the forward equation for any index, three channels on the head count, (14), and a fourth, selection, on the wealth weighting, (15).

Corollary 3 (Four channels)**.** *Let $\varphi \in C_b^1(\mathcal{S})$. Along the common path, for the cross-section m_t with overlapping-generations turnover—death at rate p independent of the record, birth into the type's newborn state $s_\circ(\xi) = (\arg\max_\theta \rho_0(\theta),\, \alpha(\xi),\, 0)$, each agent reborn with her own tolerance, the newborn cross-section $m_\circ := s_\circ\#G$, the closed population being the case $p = 0$—the head-count average obeys*

$$d\langle\varphi, m_t\rangle = \underbrace{\langle\partial_a\varphi,\, m_t\rangle\, dt}_{\textit{transport}} + \underbrace{\int(\varphi\circ J - \varphi)\, d\Phi_t}_{\textit{revision flux}} + \underbrace{p\big(\langle\varphi, m_\circ\rangle - \langle\varphi, m_t\rangle\big)\, dt}_{\textit{turnover}}\ ; \qquad (14)$$

and, in the closed population ($p = 0$) under the wealth weighting of (E.1)*, the wealth-weighted average obeys,* between common disasters—*the weights' motion is continuous there; at a common disaster wealth shares and beliefs jump together, the joint atom being the* (17) *object, outside this display—*

$$d\langle \varphi, m_t^w \rangle = \langle \partial_a \varphi,\, m_t^w \rangle\, dt + \int (\varphi \circ J - \varphi)\, d\Phi_t^w + \underbrace{d\Sigma_t(\varphi)}_{selection}, \tag{15}$$

the same integrals under the tilt, Φ_t^w the flux weighted by the crossers' wealth shares, plus the selection term of (E.1)*, where ∂_a is the ageing drift; Φ_t the* crossing flux—*a measure over the pre-jump* full *states (equivalently, over crossing types, so that $\varphi \circ J$, which consults the stopped record and carried ranking, is well defined) of held models reaching their band edge* (4) *along the common path, recording where each crosses so that $\int (\varphi \circ J - \varphi)\, d\Phi_t$ is well defined for the state-dependent increment—and J the re-selection-and-restart map* (8) *of the crossing mass; p is the turnover rate, $m_\circ$ the newborn cross-section, and Σ_t the selection term of* (E.1)*.*

The four channels, three on the head count and selection on the wealth weighting between common disasters, are the exhaustive list of why a consensus belief moves: records age (transport), investors revise (flux), the population is replaced (turnover), and, if beliefs are wealth-weighted for pricing, the weights drift with relative performance (selection). The first two carry the enlarged state's record coordinates, transport aging the record and flux being its crossing, while turnover and selection would act on any population; and it is the flux, whose size turns on where each crosser's record sits, that fails to close on the reported marginal m_t, the population image of the single-agent insufficiency (Theorem 1). The split inside the flux is the disaster asymmetry of §3.4 at the population level: lower exits at *predictable* times, disaster-driven upper exits at *inaccessible* ones. The timing split is the decomposition's own; reading it directionally—downgrades at the floor, upgrades at the ceiling—is the conditional statement of Proposition D.1, holding under local misspecification, no earlier false alarm, and a positive model-changing crossing set, while the flux itself counts every boundary crossing, self-re-adoptions and false alarms included.

The flux Φ_t is in general not Lebesgue-absolutely-continuous; at a common disaster, a *common-surprise* date, it carries an *inaccessible atom* $\Delta\Phi_u = b_u$, the crossing mass, while dispersed trend crossings and the scheduled floor-steps of disaster-synchronized cohorts supply its continuous and predictable parts (Appendix E).

Corollary 4 (Boundary-mass sufficient statistic)**.** *At a common-surprise date u, the jump in any linear index is the integral of the test-function increment over the crossing mass alone,*

$$\Delta\langle\varphi, m_u\rangle = \int (\varphi \circ J - \varphi)\, db_u, \tag{16}$$

so the aggregate response is a functional of the boundary mass b_u as a measure*—one that records each crosser's re-selection image J—rather than of its total occupation alone: the effective driver is the sub-mass whose re-selection* moves the test function *$(\varphi \circ J \neq \varphi)$: for a belief-only index $\varphi = w(\hat{\theta})$ this is the model-changing sub-mass, self-re-adopters contributing zero, while for an age-dependent φ the record restart $(a \to 0)$ makes them contribute. For a nonlinear aggregate the response is the full increment, not* Γ' *times* (16)*.*

Equation (16) is the belief-side counterpart of the sufficient statistic governing aggregate adjustment in the generalized-hazard literature (Caballero and Engel, 1999; Alvarez et al., 2022): there the response to a shock is carried by the mass at the adjustment threshold, here by the mass at the test boundary. The size of a fear wave is set at the boundary, by the crossing mass and where its re-selection sends it, not by scalar occupation. A complacent market is not for that reason safe: when the consensus has settled below the true disaster rate, the count driven by the true rate overtakes the too-low ceilings, leaving a mass poised at the boundary: a latent fragility the *level* does not reveal, since two consensuses at the same level can carry different masses poised to revise. An ordinary disaster then moves the aggregate sharply where, against a consensus with its mass interior, it would barely register.
We name the condition this isolates, that a common surprise carry positive boundary mass, an *environment-specific* condition on the population's record configuration rather than a behavioral parameter, and invoke it in the applications below. Writing ∂ for the band-edge set (the upper edge $\{T_a = u_\alpha(a)\}$ in the crash channel) and b_u for the crossing measure at u,

$$\text{(B+)} \qquad b_u(\partial) > 0 \quad \text{at the surprise date } u \tag{17}$$

requires positive mass *crossing* at u, not merely occupying a boundary set beforehand. Absent (B+) a common disaster moves the aggregate only through the smooth channels; with it (and provided a positive-measure set of the crossing mass re-selects to a different rate) the crash channel delivers a jump the trend channel cannot, its crossings dispersed across an atomless G with no synchronizing common date; with atoms in G, or trend crossings synchronized by a shared reset age, the contrast is between inaccessible and predictable jumps rather than jump versus none.

These are the aggregate's moving parts.

5 Conclusion

This paper develops a continuous-time theory of surprise-governed belief revision. Starting from Ortoleva's hypothesis-testing primitive, we show that the held belief is not a sufficient statistic for its own revision. The state enlarges to include the surviving record under which the belief has endured, making the revision hazard, its recurrence, and the aggregate cross-section implications of the rule rather than assumed objects.

The framework also scales naturally to the aggregate. Because investors observe the same history and differ only in their tolerance for surprise, the cross-section is the image of one primitive distribution under a common flow. In the rare-disaster application, this implies that consensus fear rises on disasters and declines only through scheduled revisions. The rule supplies the sign and the timing; the tolerance law supplies the magnitudes; the survey-facing patterns and the levels themselves are derived in the working-paper version.

The direction we find most promising is the endogenous case. Throughout this paper the data that investors test are taken as given. In many environments, however, beliefs influence prices, policy, or behavior, and these outcomes in turn become the record against which beliefs are tested. Whether the same primitive continues to deliver disciplined dynamics once the tested environment is itself endogenous is, we believe, the central open question, and the companion paper's subject. That the question can be posed without altering the primitive is the sense in which the rule provides a foundation rather than a reduced form.

A Foundations: statements from the working paper

This appendix states the foundational results and displays that the paper's arguments invoke, and gives the one-step proof of Proposition 1. The lemmas' proofs, and the full development of the continuous-time rule's grounding—disputability of the menu, the complete-class and highest-density selection of the band, what a trigger must test, the chained ranking, forced restart, probation onset, and the lifetime reading of the tolerance—are in the working-paper version (**?**).
The two canonical channel laws are

$$dX_t = \mu\, dt + \sigma\, dW_t, \qquad \theta = \mu \in \{\mu_1, \dots, \mu_K\} \subset \mathbb{R},\ \sigma > 0 \text{ known}, \tag{A.1}$$

$$N \text{ a Poisson process of rate } \lambda, \qquad \theta = \lambda \in \{\lambda_1, \dots, \lambda_K\} \subset (0, \infty), \tag{A.2}$$

Lemma A.1 (Disputability)**.** *For the data* (A.1)–(A.2) *observed continuously on* $(\mathcal{F}_t)$*:* (i) *distinct drifts $\mu \neq \mu'$ induce mutually equivalent laws on $\mathcal{F}_t$ for every $t > 0$, with density*

$$\frac{d\mathbb{P}^{\mu'}}{d\mathbb{P}^{\mu}}\bigg|_{\mathcal{F}_t} = \exp\!\Big(\tfrac{\mu'-\mu}{\sigma^2}\,(X_t - \mu t) - \tfrac{(\mu'-\mu)^2}{2\sigma^2}\,t\Big); \tag{A.3}$$

(ii) *distinct rates $\lambda \neq \lambda'$ induce mutually equivalent laws on $\mathcal{F}_t$ for every $t > 0$, with density*

$$\frac{d\mathbb{P}^{\lambda'}}{d\mathbb{P}^{\lambda}}\bigg|_{\mathcal{F}_t} = \Big(\tfrac{\lambda'}{\lambda}\Big)^{N_t} e^{-(\lambda'-\lambda)t}; \tag{A.4}$$

(iii) *two laws differing only in the diffusion coefficient $\sigma \neq \sigma'$ are mutually singular on $\mathcal{F}_t$ for every $t > 0$. Hence the drift and the jump intensity are disputable on the menu, the diffusion coefficient is not.*

Lemma A.2 (Band form)**.** *Fix an age $a > 0$ and a held model $\hat\theta$, and consider testing $H_0 : \theta = \hat\theta$ against $H_1 : \theta \neq \hat\theta$ on the basis of T_a, in either canonical channel: $\bar X_a \sim \mathcal{N}(\theta, \sigma^2/a)$ on the trend record, $N_a \sim$* Poisson(θa) *on the crash record. The tests whose acceptance region is an interval in T_a, randomized at most at its endpoints, form an essentially complete class: for every test there is an interval test of no larger size and of power no smaller at every alternative. Every admissible test is almost everywhere equal to an interval test.*

Proof. The classical complete-class theorems (Birnbaum, 1955; Matthes and Truax,

1967) give the statement; we display the argument. Both channels are one-parameter exponential families in T_a, with densities $p_\theta(t) = h(t)\exp(\eta(\theta)\,t - A(\eta(\theta)))$ against a σ-finite μ—Lebesgue on the trend record, counting on the crash record—and η strictly increasing ($\eta = a\theta/\sigma^2$; $\eta = \log(\theta a)$). Fix a test φ of H_0, a $[0,1]$-valued function of T_a, and write $c_0 := \mathbb{E}^{\hat\theta}[\varphi]$, $c_1 := \mathbb{E}^{\hat\theta}[T_a\varphi]$ for its rejection mass and rejection T-moment under the held model. Among tests of rejection mass c_0 the T-moment is extremized by the two one-sided threshold tests (the Neyman–Pearson lemma with T_a as the objective), and sliding the rejection mass from one tail to the other at fixed total moves the moment continuously between the extremes, endpoint randomization absorbing the lattice atoms. So either (c_0, c_1) is extremal—then φ is almost everywhere that one-sided threshold test, itself an interval test—or there is an interval test ψ, rejecting outside $[C_1, C_2]$ with randomization only at the endpoints, matched on both moments: $\mathbb{E}^{\hat\theta}[\psi] = c_0$, $\mathbb{E}^{\hat\theta}[T_a\psi] = c_1$. For any alternative $\theta' \neq \hat\theta$ put $c := \eta(\theta') - \eta(\hat\theta) \neq 0$ and let $\ell(t) = k_1 + k_2 t$ be the chord of $t \mapsto e^{ct}$ through $t = C_1, C_2$: by strict convexity $e^{ct} < \ell(t)$ on (C_1, C_2) and $e^{ct} > \ell(t)$ off $[C_1, C_2]$, so $\psi = 1$ exactly where $p_{\theta'} > (\tilde k_1 + \tilde k_2 t)\, p_{\hat\theta}$ and $\psi = 0$ where the inequality reverses, the endpoints carrying the equality set ($\tilde k_i := k_i e^{A(\eta') - A(\hat\eta)}$ up to the common positive factor). Hence, pointwise, $(\psi - \varphi)\big(p_{\theta'} - \tilde k_1 p_{\hat\theta} - \tilde k_2\, t\, p_{\hat\theta}\big) \geq 0$, and integrating against μ,

$$\mathbb{E}^{\theta'}[\psi] - \mathbb{E}^{\theta'}[\varphi] \;\geq\; \tilde k_1\big(\mathbb{E}^{\hat\theta}[\psi] - \mathbb{E}^{\hat\theta}[\varphi]\big) + \tilde k_2\big(\mathbb{E}^{\hat\theta}[T_a\psi] - \mathbb{E}^{\hat\theta}[T_a\varphi]\big) = 0:$$

ψ has the same size and power no smaller at every alternative, the essential completeness. If φ is admissible this comparison forces $\mathbb{E}^{\theta}[\psi - \varphi] = 0$ at every θ; $\psi - \varphi$ is a bounded function of T_a, complete for the exponential family (Lehmann and Romano, 2005, Theorem 4.3.1), so $\varphi = \psi$ μ-almost everywhere. □

Lemma A.3 (Bridge to the highest-density region)**.** *Fix an age a and a held model with predictive law for T_a.* (i) *If that law has a density $f_a^{\hat\theta}$ continuous at the realization, Ortoleva's one-shot test at recording resolution δ—abandon when the realized cell has held-model probability at most $c\delta$—converges as $\delta \downarrow 0$ to the density-cutoff rule: abandon iff $f_a^{\hat\theta}(T_a) \leq c$.* (ii) *Suppose every level set $\{f_a^{\hat\theta} = c\}$, $c > 0$, is Lebesgue-null—the density has no flat plateau. Then a cutoff $c_\alpha(a)$ calibrating $A_\alpha(a) = \{f_a^{\hat\theta} \geq c_\alpha(a)\}$ of* (4) *to predictive mass exactly $1-\alpha$ exists, and $A_\alpha(a)$ has minimum Lebesgue measure among all Borel sets of held-model mass at least $1-\alpha$, uniquely so up to Lebesgue-null modification.* (iii) *For the atomic law* (A.2)*, the same calibration yields* *the probability-ordered Sterne region of mass at least $1-\alpha$, of minimum cardinality.*

Proof. (i). Let $x = T_a$ and $C_\delta(x) = [x-\delta/2,\, x+\delta/2]$. By continuity of $f_a^{\hat\theta}$ at x,

$$\mathbb{P}^{\hat\theta}\big(T_a \in C_\delta(x)\big) = \int_{C_\delta(x)} f_a^{\hat\theta}(y)\,dy = f_a^{\hat\theta}(x)\,\delta\,(1+o(1)) \quad (\delta \downarrow 0). \tag{A.5}$$

The resolution-δ test abandons iff this cell probability is at most $c\delta$, i.e., by (A.5), iff $f_a^{\hat\theta}(x)(1+o(1)) \le c$; letting $\delta \downarrow 0$ gives abandon whenever $f_a^{\hat\theta}(x) < c$ and retain whenever $f_a^{\hat\theta}(x) > c$. At the cutoff $f_a^{\hat\theta}(x) = c$ the leading-order cell probability equals $c\delta$ and the test sits on its boundary; for the record densities of the main channels—the Gaussian trend statistic and other unimodal continuous laws, non-constant on the level set so it carries no Lebesgue mass—$\{f_a^{\hat\theta} = c\}$ is null (a continuous density with a flat plateau at height c would be the exception, excluded here), so retaining iff $f_a^{\hat\theta}(x) > c$ and iff $f_a^{\hat\theta}(x) \ge c$ coincide, and we take the closed region $\{f_a^{\hat\theta} \ge c\}$ of (4), fixing the $o(1)$ ambiguity; the atomic Sterne region (Lemma D.1) is closed by construction, the cutoff atom included to meet the mass floor.

(ii). With $f = f_a^{\hat\theta}$ and $c = c_\alpha(a)$, the super-level set $A^\star = \{f \ge c\}$ calibrated to mass $1-\alpha$ has minimum Lebesgue measure among Borel sets of held-model mass at least $1-\alpha$—the highest-density-region property (Box and Tiao, 1973; Hyndman, 1996). (For the unimodal continuous record densities here $\{f = c\}$ is Lebesgue-null and a unique c attains mass exactly $1-\alpha$.) Both steps are the generalized Neyman–Pearson lemma with Lebesgue measure—σ-finite, as the lemma requires—in the role of the alternative (Lehmann and Romano, 2005, Theorem 3.2.1); we display the argument. For the exact calibration, $c \mapsto \mathbb{P}^{\hat\theta}(f_a^{\hat\theta} \ge c)$ is continuous—since $\mathbb{P}^{\hat\theta}(f_a^{\hat\theta} = c) = c\,\mathrm{Leb}\{f_a^{\hat\theta} = c\} = 0$—and decreases from 1 to 0, so $c_\alpha(a)$ exists. For uniqueness, let B have held-model mass at least $1-\alpha$ and $\mathrm{Leb}(B) = \mathrm{Leb}(A_\alpha(a))$; abbreviate $A = A_\alpha(a)$, $c = c_\alpha(a)$. On $B \setminus A$ the density is strictly below c, on $A \setminus B$ it is at least c, and the mass constraint gives $\mathbb{P}^{\hat\theta}(A \setminus B) \le \mathbb{P}^{\hat\theta}(B \setminus A)$; hence $c\,\mathrm{Leb}(A \setminus B) \le \mathbb{P}^{\hat\theta}(A \setminus B) \le \mathbb{P}^{\hat\theta}(B \setminus A) < c\,\mathrm{Leb}(B \setminus A)$ whenever $\mathrm{Leb}(B \setminus A) > 0$, contradicting $\mathrm{Leb}(B \setminus A) = \mathrm{Leb}(A \setminus B)$; so both differences are Lebesgue-null.

(iii). For the atomic law (A.2), ordering the integers by descending mass (ties by smaller count) and taking the shortest initial segment to reach $1-\alpha$ gives the probability-ordered Sterne region (Sterne, 1954), of minimum cardinality among sets of mass at least $1-\alpha$: any set B of that mass with $|B| < |A|$ must omit some accrued atom $n \in A$ and include an atom $m \notin A$ with $p_m \le p_n$ (accrual is by descending

mass), so exchanging m for n raises the mass without raising cardinality; iterating, A's initial segment of size $|B|$ already has mass at least that of B, yet by minimality of the accrual it has mass $< 1-\alpha$, a contradiction. The boundary atom at the accrual cutoff is included—the closed Sterne interval of Lemma D.1, not the strict super-level set—so the region is uniquely defined. $\square$

Proof of Proposition 1. Parts (i)–(iii) are the cited lemmas. The one assembly step: under unimodality every superlevel set $\{f_a^{\hat{\theta}} \geq c\}$, $c > 0$, is an interval, so the minimal region of Lemma A.3(ii) lies *inside* the essentially complete class of Lemma A.2 rather than merely being dominated by a member of it, and the two lemmas select the same object. $\square$

Lemma A.4 (Chaining)**.** *Let* $0 = \sigma_0 < \sigma_1 < \cdots < \sigma_k$ *be successive revision times and apply* (8) *at each. Then for every pair* $\theta, \theta_0 \in \Theta$ *with* $\rho_0(\theta_0) > 0$ *(a model of zero initial mass keeps* $\rho_t = 0$ *under the multiplicative update and is dropped from the ratios),*

$$\frac{\rho_{\sigma_k}(\theta)}{\rho_{\sigma_k}(\theta_0)} = \frac{\rho_0(\theta)}{\rho_0(\theta_0)} \left.\frac{d\mathbb{P}^{\theta}}{d\mathbb{P}^{\theta_0}}\right|_{\mathcal{F}_{\sigma_k}}, \tag{A.6}$$

so the ranking at σ_k *is the single Bayes update of* ρ_0 *on the entire history up to* σ_k*. The held model is the full-history maximum-a-posteriori choice; only the test record restarts, the ranking* ρ *does not.*

Lemma A.5 (Restart removes an adoption-time degeneracy)**.** *If a freshly adopted model were monitored against the inherited record rather than a restarted one, it could lie outside its own acceptance region at the instant of adoption, forcing an immediate revision with no new data. Restarting the test record at each adoption removes this degeneracy.*

For the held law Poisson(10) at tolerance $\alpha = 0.1$ and age 1, the probability-ordered Sterne region invoked in the proofs is

$$A_{\alpha}^{(\lambda_2)}(1) = \{5, 6, \ldots, 15\}, \qquad \mathbb{P}^{10}\big(5 \leq N_1 \leq 15\big) \approx 0.922 \geq 0.9, \tag{A.7}$$

Lemma A.6 (Tempo / probation onset)**.** (i) *In the crash channel* (A.2)*, with* $a_0 = 0$ *the holding time satisfies* $\tau > 0$ *a.s.: continuous monitoring from adoption produces no instantaneous rejection.* (ii) *In the trend channel* (A.1)*, when the held model is the truth, continuous monitoring from* $a = 0$ *rejects it at arbitrarily small ages,*

$\inf\{a > 0 : T_a \notin A_\alpha(a)\} = 0$ *a.s.; a probation onset* $a_0 > 0$ *is therefore required for* $\tau > 0$*.*

Lemma A.7 (Time-uniform reading)**.** *Let* θ_0 *be the held model and* $\{\omega_\theta\}_{\theta \neq \theta_0}$ *weights with* $\omega_\theta \geq 0$*,* $\sum \omega_\theta = 1$*. The rule that abandons* θ_0 *the first time the menu likelihood-ratio process*

$$\Lambda_t := \sum_{\theta \neq \theta_0} \omega_\theta \, \frac{d\mathbb{P}^\theta}{d\mathbb{P}^{\theta_0}}\Big|_{\mathcal{F}_t} \tag{A.8}$$

reaches $1/\alpha$ *has lifetime type-I error at most* α*: the probability of ever abandoning a correct* θ_0 *is at most* α*. This rule consults the alternatives. By the admissibility of anytime-valid inference, any test with such a lifetime guarantee is of nonnegative-supermartingale (e-process) form; the finite-menu likelihood-ratio mixture* (A.8) *is one such supermartingale, sufficient for the guarantee but not the only one—conservative rules and valid supermartingales not expressible as a mixture over the menu alternatives also control the lifetime error, and a supermartingale need not be a likelihood ratio to any named menu rate. We cite the supermartingale characterization and do not rely on it.*

The *standardized record*, computed under the correctly held drift $\hat{\mu} = \mu^\star$ (under misfit it gains the drift term $\sqrt{a}\,(\mu^\star - \hat{\mu})/\sigma$, the decomposition the proof of Corollary 2(b) carries), is

$$\frac{\bar{X}_a - \hat{\mu}}{\sigma/\sqrt{a}} = \frac{(X_a - \mu^\star a)/a}{\sigma/\sqrt{a}} = \frac{\sigma W_a/a}{\sigma/\sqrt{a}} = \frac{W_a}{\sqrt{a}}, \tag{A.9}$$

B Proof of Theorem 1

Proof of Theorem 1. Part (i): insufficiency. The revision time under any rule of Definition 1 is the first failure of the retention test on the since-adoption record $X_{(\sigma,t]}$, a functional of the adoption date σ and the public path; a fixed prior updated on $\mathcal{F}_t$ is a functional π_t of the public path alone, identical across adoption dates. We exhibit an event of positive probability on which two holders of the same $\hat{\theta}$ share π_t yet carry different conditional revision intensities; the revision law then cannot be $\sigma(\pi_t)$-measurable.

Work in the crash channel at held rate $\hat{\lambda}$ with band $[l(a), u(a)]$ (Lemma D.1). *Within an age.* Fix an age a with $u(a) - l(a) \geq 2$, off the finitely many step ages of the band (such ages exist: the band (A.7) has width ten). Off the step ages,

$u(a+\,da) = u(a)$ for all small da; one arrival moves a ceiling-holder to $u(a)+1 > u(a)$ and a slack-two holder to at most $u(a)-1$, and two arrivals in $(a, a+\,da]$ have probability $O(\,da^2)$, so

$$\mathbb{P}^{\lambda^\star}\big(N_{a+\,da} > u(a+\,da) \,\big|\, N_a = u(a)\big) = \lambda^\star\,da + o(\,da), \qquad \mathbb{P}^{\lambda^\star}\big(N_{a+\,da} > u(a+\,da) \,\big|\, l(a) \le N_a \le u(a)-2\big) = \tag{B.1}$$

the held model is insufficient within an age. *Across ages on one history.* Take $\hat\lambda = \lambda^\star = 10$ and $\alpha = 0.1$, adoption dates $\sigma_A < \sigma_B$ with ages $a_A = 1$ and $a_B = \frac{1}{2}$ at t (chosen off both bands' step ages, as we may): the bands are $A_\alpha(1) = \{5, \dots, 15\}$ of mass ≈ 0.922 (A.7) and $A_\alpha(\frac{1}{2}) = \{2, \dots, 9\}$, the Sterne region of Poisson(5), of mass $\approx 0.928 \ge 0.9$. The records are counts over the nested windows $(\sigma_A, t] \supset (\sigma_B, t]$, whose two halves are independent Poisson(5) increments, so the event

$$E = \big\{N_{(\sigma_B, t]} = 7\big\} \cap \big\{N_{(\sigma_A, \sigma_B]} = 8\big\}, \qquad \mathbb{P}^{\lambda^\star}(E) = e^{-5}\frac{5^7}{7!} \cdot e^{-5}\frac{5^8}{8!} > 0, \tag{B.2}$$

puts holder B at $N_{a_B} = 7$ with $l(\frac{1}{2}) = 2 \le 7 \le u(\frac{1}{2}) - 2 = 7$, strictly inside her band, and holder A at $N_{a_A} = 7 + 8 = 15 = u(1)$, the ceiling of his. The endpoint counts alone do not yet make A and B *current* holders at t—some paths in E exit a band earlier—so refine E by an in-band staircase: on the finite grid formed by both holders' scheduled band-step ages in $(\sigma_A, t]$, prescribe a nondecreasing integer path lying in the ceiling-lagged, floor-led bands of both windows at every grid age, with the endpoint values of E. Such a path exists: joint feasibility is a corridor statement. Both windows carry the one band family $[l, u]$ of Lemma D.1 at $\hat\lambda = 10$, so writing $\mu := \hat\lambda(x - \sigma_B) \in (0, 5]$ for the younger window's expected count at calendar age $x \in (\sigma_B, t]$, the total count must lie in $[\max\{l(\mu{+}5),\, l(\mu){+}8\},\, \min\{u(\mu{+}5),\, u(\mu){+}8\}]$, the older window's band intersected with the younger's shifted by the prescribed early count 8. On $\mu \in (0, 10]$ both endpoints step only upward—the region changes of the probability-ordered prefix, computed directly as for (A.7), are all upward steps, five of them joint floor-and-ceiling shifts at rank-swap ages (D.1)—so the corridor's lower edge is $l(\mu) + 8$, since $l(\mu + 5) \le l(10) = 5 < 8$, and nonemptiness reduces to $u(\mu + 5) \ge l(\mu) + 8$. Where $l(\mu) \le 1$ this holds because $u(\mu + 5) \ge u(5) = 9$; and $l(\mu) = 2$ forces $\mu > 4$—the Sterne region of Poisson(4) is $\{1, \dots, 7\}$, of mass ≈ 0.931—whence $u(\mu + 5) \ge u(9) = 13 \ge 10$, the Sterne region of Poisson(9) being $\{4, \dots, 13\}$, of mass ≈ 0.905. Since $l(5) = 2$, the two cases exhaust the range; at the seams the corridor pins the prescription, $[8, 8]$ as $\mu \downarrow 0$ and $[10, 15]$ at $\mu = 5$, matching the endpoint values of E; and on $(\sigma_A, \sigma_B]$ only the older window's band constrains, its cells nonempty since $l \leq u$. A nondecreasing integer path through nonempty cells with upward-stepping bounds exists—the lower envelope $\max\{8,\, l(\mu) + 8\}$, its final arrivals to 15 placed after the older window's ceiling reaches 15, which it does strictly before age one since $u(10) = 15$ and t is off the step ages—and the refined event $E' \subseteq E$ prescribes finitely many Poisson increments over disjoint intervals, each of positive probability, so $\mathbb{P}^{\lambda^\star}(E') > 0$; on E' no earlier exit occurs, the grid conditions forbidding crossings at real ages exactly as in (D.18). Two scopings make the excluded object exact. The σ-field is $\sigma(\pi_t)$—functionals of the public path under the fixed prior: the witness exhibits two admissible full states on one path with the same π_t and different exit intensities, so no $\sigma(\pi_t)$-measurable description carries the hazard. And the adoption dates $\sigma_A < \sigma_B$ are *initialized* states of the canonical construction—initial conditions of the flow, not claimed as a positive-probability configuration of a single rule-generated path from a common prior; adjoining the private adoption date, the supplementary variable of part (i), is exactly what restores the Markov description—the theorem's point, not its concession. □

Corollary 2 then removes even the initialization: the configuration is realized by the rule's own dynamics from a common start—permanently in the trend channel, recurrently in the crash channel (Lemmas B.1 and B.2 below). Both hold $\hat{\lambda}$; both observe the same public path, so π_t is the same random variable for both; by (B.1) their upper-exit intensities at t are 0 and $\lambda^\star$. Since $\mathbb{P}^{\lambda^\star}(E') > 0$, the conditional revision law is not $\sigma(\pi_t)$-measurable; what separates the holders is the private adoption date σ, equivalently the age $a = t - \sigma$, which π_t does not carry. (The integer endpoints can coincide over age ranges, so the witness is the event E', not $a_A \neq a_B$ alone.) The argument used only reset-on-adoption against the posterior's reset-freeness, so it holds for the whole record-resetting class.

Part (ii): irreducibility. The revision time is the first exit of T_a from $A_\alpha(a)$; the band is fixed by $(\hat{\theta}, a)$ and the data drive T forward given (T_a, a), so the revision law is a measurable function of $(\hat{\theta}, T_a, a)$. No coordinate drops: T_a *given* $(\hat{\theta}, a)$, the within-age pair of Part (i); a *given* $(\hat{\theta}, T_a)$, the band moves with age—stepping in the crash channel, contracting like $a^{-1/2}$ in the trend channel: a summary on the age-a edge, $|T - \hat{\mu}| = z_\alpha \sigma / \sqrt{a}$, is interior at any earlier age $a' < a$, since $z_\alpha \sigma / \sqrt{a} < z_\alpha \sigma / \sqrt{a'}$ by (6); $\hat{\theta}$ *given* (T_a, a), distinct held models impose bands that separate: trend-channel bands have distinct centers $\hat{\mu}_1 \neq \hat{\mu}_2$, and crash-channel ceilings separate linearly in age,

$$u_{\hat{\lambda}_1}(a) - u_{\hat{\lambda}_2}(a) = (\hat{\lambda}_1 - \hat{\lambda}_2)\, a + O(\sqrt{a}) \longrightarrow \infty \qquad (\hat{\lambda}_1 > \hat{\lambda}_2)$$

by the leading form of Lemma D.1, so the band functions differ from some age on even where the integer endpoints coincide at isolated ages. This uses age-moving bands and model-specific separation, so it holds in the two channels and in any with a nondegenerate age-dependent band and a menu-discriminating statistic, not by record-resetting alone. The comparison underlying non-measurability is between *admissible* full states sharing (T_a, a): with distinct held models the continuation first-exit laws differ from the separation age on, and such same-record two-holder configurations are admissible as initialized states of the canonical construction, and the overlap domain is nonempty: at $\hat{\lambda}_1 = 10$, $\hat{\lambda}_2 = 9$, $\alpha = 0.1$, age 1, the bands are $A^{(10)}_{0.1}(1) = \{5, \ldots, 15\}$ (A.7) and $A^{(9)}_{0.1}(1) = \{4, \ldots, 13\}$, the Sterne region of Poisson(9) of mass $\approx 0.905 \geq 0.9$, so every count $n \in \{5, \ldots, 13\}$ lies in both, and two initialized age-1 holders at the same $(n, 1)$ carrying the two models are admissible states whose continuation first-exit laws differ from the separation age on. On sub-domains where the bands are disjoint, (T_a, a) identifies $\hat{\theta}$ and non-measurability there is vacuous relabeling; the claim is irreducibility on the overlap domain, occupied under any initialization charging both models. States outside the held band exit immediately by convention and are excluded from the domain.

Part (iii): closure. Between revisions $\hat{\theta}$ and ρ are constant, the age flows ($\dot{a} = 1$), and T_a evolves by the data's Markov transition given (T_a, a); at the first exit the state jumps and the record resets, $a \to 0$. By the stopped sufficiency of Assumption 1 the re-ranking at τ depends on the record only through (T_τ, τ), so ρ' and $\hat{\theta}' = \arg\max \rho'$ are functions of (ρ, T_τ, τ). Deterministic flow between jumps, a first-exit jump time, and a state-function reset are the transitions of a piecewise-deterministic Markov process (Davis, 1984) in the crash channel and of a boundary-reset diffusion in the trend channel; with the data Markov the enlarged state $(\hat{\theta}, T_a, a, \rho)$ is therefore Markov (Küchler and Sørensen, 1998), revision its first exit from (4). It does not regress to a finer object: ρ is a finite sufficient statistic for the history's effect on the ranking (Appendix A), and $(\hat{\theta}, T_a, a)$ alone suffices for the revision law, ρ entering only the post-revision adoption. All claims are over a positive time interval, never by conditioning on an instantaneous event (§3.2).

Adaptedness versus factorization. For a single agent with fixed initialization σ_t is itself $\mathcal{F}_t$-measurable—the last exit time, a functional of the data—so the revision time and the compensator of the revision point process are $\mathcal{F}_t$-adapted. The non-representability is therefore not a failure of adaptedness but of *factorization*: the revision law is $\sigma(\hat{\theta}_t, T_{a_t}, a_t)$-measurable but not $\sigma(\beta_t)$-measurable, the two σ-algebras being in general incomparable—the posterior carries the menu weights, the record carries the age. (A literal filtration enlargement appears only under randomized initialization, when the same fact appears as the revision law's projection onto the posterior-generated subfiltration being a coarsened average.)

Realized, not merely constructed. Proposition 2(b) turns the held-model half of the claim from a constructed possibility into a path regularity: along almost every trajectory the model is re-adopted infinitely often at a range of ages, so the same held belief recurs paired with distinct hazards on the agent's own path. The stronger statement—two agents at one date, the same fixed-prior posterior, different conditional revision laws—is Corollary 2, proved below from the two desynchronization lemmas.

The witness generated, not initialized. The lemmas and proofs below upgrade the remark: under tolerance heterogeneity the excluded configuration arises along the rule's own path—and persists in the trend channel, recurs in the crash channel.

Lemma B.1 (Common-path desynchronization; trend channel)**.** *Let two investors run the rule of Definition 1 in the disputed-trend channel with probation $a_0 > 0$ on the common path, with common menu and common prior, and band levels $z_1 \leq z_2$ (tolerances $\alpha_1 \geq \alpha_2$). Write $\sigma^{(i)}(t)$ for investor i's current adoption date and $a_t^{(i)} = t - \sigma^{(i)}(t)$ for her age.*

(a) *If $z_1 < z_2$ and the investors have shared every adoption date so far, then, conditional on the past, the next regime is again common if and only if—up to $\mathbb{P}^\star$-null boundary events—the standardized record breaches the wider band at the probation onset, an event of conditional probability at most $\bar{q} := \max_{\theta\in\Theta} \mathbb{P}\big(|Z + m_\theta\sqrt{a_0}\,| \geq z_2\big) < 1$, with $Z \sim \mathcal{N}(0,1)$ and m_θ the held model's standardized drift misfit; on the complement the z_1-investor exits strictly first. The first date $S^\dagger$ at which the two adoption dates differ is therefore $\mathbb{P}^\star$-a.s. finite.*

(b) *From any configuration at a stopping time s with $\sigma^{(1)}(s) \neq \sigma^{(2)}(s)$—any band levels, any held models—no two subsequent exits of the two investors fall on the* *same date, $\mathbb{P}^\star$-a.s. Combined with* (a)*: $a_t^{(1)} \neq a_t^{(2)}$ for every $t \geq S^\dagger$, $\mathbb{P}^\star$-a.s.*

Proof. Throughout, $X_t = \mu^\star t + \sigma B_t$ is the common path, B a standard Brownian motion, and exits follow the closed-band, strictly-outside convention of (5): for a regime adopted at the date ς with held drift $\hat{\mu}$, the standardized record at age a is

$$V_a = \frac{\bar{X}_a - \hat{\mu}}{\sigma/\sqrt{a}} = \frac{X_{\varsigma+a} - X_\varsigma - \hat{\mu} a}{\sigma\sqrt{a}} = \frac{B_{\varsigma+a} - B_\varsigma}{\sqrt{a}} + m\sqrt{a}, \qquad m = \frac{\mu^\star - \hat{\mu}}{\sigma}, \tag{B.3}$$

the band is $\{|V_a| \leq z_i\}$, and the exit age is $\tau^{(i)} = \inf\{a > a_0 : |V_a| > z_i\}$.

Step 0 (common regimes are identical regimes). Suppose the two investors have shared every adoption date through their k-th. Their k stopped records coincide—the same path over the same windows—so their chained rankings coincide by k applications of Theorem 1(iii) from the common ρ_0, and their (k+1)-st held models coincide by (8) under the setup's tie convention. Along a shared stretch the two therefore carry the same $(\hat{\theta}, \rho, T_a, a)$ and differ only in the level tested, $z_1 < z_2$.

Step 1 (one common regime: the three onset events). Let a common regime begin at the stopping time ς with held drift $\hat{\mu}$. By (B.3) and the independent increments of B, conditionally on $\mathcal{F}_\varsigma$ the onset value is $V_{a_0} \sim \mathcal{N}(m\sqrt{a_0}, 1)$, so $\mathbb{P}(|V_{a_0}| \in \{z_1, z_2\} \mid \mathcal{F}_\varsigma) = 0$ and the following three events are exhaustive up to a null set.

(1a) On $\{|V_{a_0}| > z_2\}$. V is continuous and strictly outside both bands at a_0, so there is $\varepsilon > 0$ with $|V_a| > z_2 > z_1$ on $[a_0, a_0 + \varepsilon)$, and both exit ages equal $\inf\{a > a_0 : |V_a| > z_i\} = a_0$. The stopped pairs are (T_{a_0}, a_0) for both, the rankings coincide by Step 0, the re-selections (8) coincide, and the next regime is again common.

(1b) On $\{z_1 < |V_{a_0}| < z_2\}$. Investor 1 exits at a_0 as in (1a). Investor 2 is strictly inside her band at a_0; by continuity there is $\varepsilon > 0$ with $|V_a| < z_2$ on $[a_0, a_0 + \varepsilon]$, so $\tau^{(2)} \geq a_0 + \varepsilon > a_0 = \tau^{(1)}$: the adoption dates part at $\varsigma + a_0$.

(1c) On $\{|V_{a_0}| < z_1\}$. Both survive the onset, and $\tau^{(1)} < \infty$ $\mathbb{P}^\star$-a.s. by Proposition 2(b). First, $\{a : |V_a| > z_2\} \subseteq \{a : |V_a| > z_1\}$ gives $\tau^{(1)} \leq \tau^{(2)}$. Second, the boundary evaluation: if $|V_{\tau^{(1)}}| > z_1$, continuity gives $\varepsilon > 0$ with $|V_a| > z_1$ on $(\tau^{(1)} - \varepsilon, \tau^{(1)}]$, contradicting the definition of $\tau^{(1)}$ as an infimum exceeding those times; if $|V_{\tau^{(1)}}| < z_1$, continuity gives a neighborhood of $\tau^{(1)}$ on which $|V| < z_1$, contradicting the existence of times $a_n \downarrow \tau^{(1)}$ with $|V_{a_n}| > z_1$; hence $|V_{\tau^{(1)}}| = z_1$. Third, strictness: $|V_{\tau^{(1)}}| = z_1 < z_2$ and continuity give $\varepsilon > 0$ with $|V_a| < z_2$ on $[\tau^{(1)}, \tau^{(1)} + \varepsilon]$, so $\tau^{(2)} \geq \tau^{(1)} + \varepsilon > \tau^{(1)}$: investor 1 exits alone and the dates part. (Across a continuum of levels the same monotonicity appears in the working-paper version's aggregate

appendix.)

Step 2 (part (a): the per-regime bound and its iteration). By Step 1, the regime is followed by another common regime exactly on $\{|V_{a_0}| > z_2\}$, up to a null set, and

$$\mathbb{P}\big(|V_{a_0}| > z_2 \,\big|\, \mathcal{F}_{\varsigma}\big) = \mathbb{P}\big(|Z+m\sqrt{a_0}| \geq z_2\big) =: q(\hat{\mu}) \leq \bar{q} := \max_{\theta\in\Theta} q(\hat{\mu}_\theta) < 1, \qquad Z \sim \mathcal{N}(0,1), \tag{B.4}$$

the maximum over the finite menu; each $q(\hat{\mu}) < 1$ because the nondegenerate Gaussian law of $Z + m\sqrt{a_0}$ assigns positive mass to the open interval $(-z_2, z_2)$. Let ς_n be the n-th common adoption date and C_n the event that the first n regimes all end as in (1a); then $C_n \in \mathcal{F}_{\varsigma_{n+1}}$ and $C_{n+1} = C_n \cap \{|V_{a_0}^{(n+1)}| > z_2\}$ with $V^{(n+1)}$ the $(n{+}1)$-st regime's record, so by the tower property and (B.4),

$$\mathbb{P}(C_{n+1}) = \mathbb{E}\Big[\mathbf{1}_{C_n}\,\mathbb{P}\big(|V_{a_0}^{(n+1)}| > z_2 \,\big|\, \mathcal{F}_{\varsigma_{n+1}}\big)\Big] \leq \bar{q}\,\mathbb{P}(C_n), \qquad \text{whence} \qquad \mathbb{P}(C_n) \leq \bar{q}^{\,n} \text{ for all } n.$$

Therefore $\mathbb{P}\big(\bigcap_n C_n\big) = \lim_n \mathbb{P}(C_n) = 0$ and the number N of common regimes is a.s. finite. Each (1a) regime has length exactly a_0, and the final regime—of type (1b) or (1c)—ends at the a.s.-finite $\tau^{(1)}$ of Step 1, so $S^\dagger$, the exit date of that final common regime, is a sum of finitely many a.s.-finite lengths: $S^\dagger < \infty$ a.s.

Step 3 (part (b): conditional near-atomlessness). Fix the stopping time s with $\sigma^{(1)}(s) < \sigma^{(2)}(s)$, write $d = \sigma^{(2)}(s) - \sigma^{(1)}(s) > 0$, held drifts $\hat{\mu}_1, \hat{\mu}_2$, onset dates $o_i = \sigma^{(i)}(s) + a_0$, and let $\tau^{(1)}, \tau^{(2)}$ denote, in calendar time, the two investors' next exits after s. Let R be any $\mathcal{F}_s$-measurable random time with $R > s$. On $\{\tau^{(i)} = R,\ \tau^{(i)} > o_i\}$ the exit is strictly after the onset, so the boundary evaluation of Step 1(1c) gives $|V_R^{(i)}| = z_i$, where, decomposing at s,

$$V_R^{(i)} = \frac{\big(X_s - X_{\sigma^{(i)}} - \hat{\mu}_i(s - \sigma^{(i)})\big) + \big(X_R - X_s - \hat{\mu}_i(R - s)\big)}{\sigma\sqrt{R - \sigma^{(i)}}} :$$

the first bracket and the denominator are $\mathcal{F}_s$-measurable, and the second bracket equals $\sigma(B_R - B_s) + (\mu^\star - \hat{\mu}_i)(R - s)$, conditionally Gaussian given $\mathcal{F}_s$ with variance $\sigma^2(R - s) > 0$. A nondegenerate Gaussian assigns zero mass to the two-point set $\{\pm z_i\sigma\sqrt{R - \sigma^{(i)}} - (\text{first bracket})\}$, so

$$\mathbb{P}\big(\tau^{(i)} = R,\ \tau^{(i)} > o_i \,\big|\, \mathcal{F}_s\big) = 0. \tag{B.5}$$

Step 4 (part (b): the coincidence polynomial). Suppose $\tau^{(1)} = \tau^{(2)} = t$ with both exits past their onsets; put $v = t - \sigma^{(2)}(s) > 0$. Writing (B.3) for each investor at the common date t,

$$\sigma\sqrt{v+d}\, V_t^{(1)} = X_t - X_{\sigma^{(1)}} - \hat{\mu}_1(v+d), \qquad \sigma\sqrt{v}\, V_t^{(2)} = X_t - X_{\sigma^{(2)}} - \hat{\mu}_2 v,$$

and subtracting eliminates the common value X_t:

$$\sigma\sqrt{v+d}\, V_t^{(1)} - \sigma\sqrt{v}\, V_t^{(2)} = \underbrace{\big(X_{\sigma^{(2)}} - X_{\sigma^{(1)}}\big) - \hat{\mu}_1 d}_{=:\,c} + \underbrace{(\hat{\mu}_2 - \hat{\mu}_1)}_{=:\,e}\, v, \tag{B.6}$$

with c, e $\mathcal{F}_s$-measurable. The boundary evaluation gives $V_t^{(1)} = \varepsilon_1 z_1$ and $V_t^{(2)} = \varepsilon_2 z_2$ for signs $\varepsilon_i \in \{\pm 1\}$, so (B.6) reads $\varepsilon_1 z_1 \sigma\sqrt{v+d} = (c+ev) + \varepsilon_2 z_2 \sigma\sqrt{v}$. Squaring once,

$$z_1^2\sigma^2(v+d) = (c+ev)^2 + 2\varepsilon_2 z_2 \sigma(c+ev)\sqrt{v} + z_2^2\sigma^2 v;$$

isolating the term in $\sqrt{v}$ and squaring again, every such v, for either sign pair, satisfies $P(v) = 0$ with

$$P(v) := \Big(\underbrace{z_1^2\sigma^2(v+d) - z_2^2\sigma^2 v - (c+ev)^2}_{=:\,Q(v)}\Big)^2 - 4z_2^2\sigma^2 v\,(c+ev)^2. \tag{B.7}$$

P has $\mathcal{F}_s$-measurable coefficients, and it is not the zero polynomial when $d > 0$; expanding the leading terms case by case,

$$\begin{aligned}
&e \neq 0: && P(v) = e^4 v^4 + (\text{lower order}),\\
&e = 0,\ z_1 \neq z_2: && P(v) = (z_1^2 - z_2^2)^2\sigma^4 v^2 + (\text{lower order}),\\
&e = 0,\ z_1 = z_2,\ c \neq 0: && P(v) = -4z_1^2\sigma^2 c^2\, v + (z_1^2\sigma^2 d - c^2)^2,\\
&e = 0,\ z_1 = z_2,\ c = 0: && P(v) \equiv z_1^4\sigma^4 d^2 > 0,
\end{aligned} \tag{B.8}$$

using in the third line that $Q(v) = z_1^2\sigma^2 d - c^2$ is constant there, and in the fourth that additionally $c = 0$. In every case $P \not\equiv 0$, so P has at most four roots, each an $\mathcal{F}_s$-measurable random time. Therefore

$$\{\tau^{(1)} = \tau^{(2)},\ \tau^{(1)} > o_1,\ \tau^{(2)} > o_2\} \subseteq \bigcup_{R\,\in\,\mathrm{roots}(P)} \{\tau^{(2)} = R,\ \tau^{(2)} > o_2\},$$

a union of at most four events, each null given $\mathcal{F}_s$ by (B.5). For the remaining exit-kind cases: on $\{\tau^{(1)} = o_1,\ \tau^{(2)} = o_2\}$ a coincidence forces $o_1 = o_2$, i.e. $d = 0$, excluded; on $\{\tau^{(1)} = o_1,\ \tau^{(2)} > o_2\}$ a coincidence is the event $\{\tau^{(2)} = o_1,\ \tau^{(2)} > o_2\}$, null by (B.5) with $R = o_1 > s$, and symmetrically with the roles exchanged. Summing the four cases,

$$\mathbb{P}\big(\tau^{(1)} = \tau^{(2)} \,\big|\, \mathcal{F}_s\big) = 0 \qquad \text{on } \{\sigma^{(1)}(s) \neq \sigma^{(2)}(s)\}. \tag{B.9}$$

Step 5 (part (b): iteration along the path). After $S^\dagger$ the set E of the two investors' exit dates is a.s. locally finite (Lemma D.2 for each investor; a union of two locally finite sets is locally finite), so its points can be listed $S^\dagger = s_0 < s_1 < s_2 < \cdots$, each a stopping time. Let $A_k = \{$the two investors' next exits after s_k coincide$\}$ and $G = \{$some two exits after $S^\dagger$ coincide$\}$. If a coincidence occurs at a date $t > S^\dagger$, then $t \in E$ and, at the largest listed point $s_k < t$, both investors' next exits after s_k fall at t; hence $G \subseteq \bigcup_{k\geq 0} A_k$. By induction on k: on $A_0^c \cap \cdots \cap A_{k-1}^c$ every point of E up to and including s_k is a lone exit, so the adoption dates at s_k differ—at $s_0 = S^\dagger$ the exiter's date is $S^\dagger$ while the other's is the last common adoption date, strictly earlier—and (B.9) applied at s_k gives $\mathbb{P}\big(A_k \mid \mathcal{F}_{s_k}\big) = 0$ there. Hence $\mathbb{P}\big(A_k \cap A_0^c \cap \cdots \cap A_{k-1}^c\big) = 0$ for every k, and

$$\mathbb{P}(G) \leq \sum_{k\geq 0} \mathbb{P}\big(A_k \cap A_0^c \cap \cdots \cap A_{k-1}^c\big) = 0.$$

On G^c the two adoption-date processes are right-continuous step functions changing only at points of E, never at a common point after $S^\dagger$; they differ at $S^\dagger$, and a re-equalization would require a common change, so $\sigma^{(1)}(t) \neq \sigma^{(2)}(t)$, i.e. $a_t^{(1)} \neq a_t^{(2)}$, for every $t \geq S^\dagger$. □

The lemma answers, for a pair on one path, the question the state-dependent adjustment literature asks of (S, s) cross-sections—whether heterogeneity of positions within bands sustains itself (**?**). There the cross-section is kept dispersed by idiosyncratic shocks; here band-width heterogeneity alone makes the staggering absorbing along a single common path, with no idiosyncratic noise anywhere in the model.

Lemma B.2 (Common-path staggering; crash channel)**.** *Let two investors run the rule of Definition 1 in the disputed-crash-rate channel on the common path, with common menu and common prior, and band levels $z_1 < z_2$, the Sterne bands nested at every age (Lemma D.1). Then:*

(a) *Every synchronized phase—a maximal stretch on which the two share every adoption date—ends at a lone exit in $\mathbb{P}^\star$-a.s. finite time; conditional on the* *past, each common regime ends in a lone exit with probability at least $\bar{\varepsilon} > 0$, uniform over the finite menu.*

(b) *Staggering is not absorbing: with positive $\mathbb{P}^\star$-probability a common disaster finds both counts at their ceilings, at differing adoption dates, and exits both investors at one date; at any such joint exit at which both re-select the same model—as at every exit past consolidation with the truth on the menu—the pair returns to a common regime.*

(c) *Within stretches of differing adoption dates, ceiling-versus-interior configurations recur without bound $\mathbb{P}^\star$-a.s.: infinitely many dates, unbounded above, at which one investor's count sits at her ceiling, the other's is strictly interior, and an open interval follows on which neither band steps.*

Proof. Exits follow the closed-band convention of (5) with the right-continuous step convention of Lemma D.1; the bands are nested at every age, $l_2(a) \le l_1(a) \le u_1(a) \le u_2(a)$ (Lemma D.1), and Step 0 of the proof of Lemma B.1 applies verbatim: along a shared stretch the two investors carry the same $(\hat{\theta}, \rho, N_a, a)$ and differ only in the band tested.

Step 1 (part (a): the lone-exit event D). Let a common regime begin at the stopping time ς with held rate $\hat{\lambda}$, and let N_a be the shared post-adoption count. By Lemma D.1, $u_i(a) = \hat{\lambda}a + z_i\sqrt{\hat{\lambda}a} + o(\sqrt{a})$, so $u_2(a) - u_1(a) \to \infty$ and there is a finite $A(\hat{\lambda})$ with $u_2(a) - u_1(a) \ge 2$ for all $a \ge A(\hat{\lambda})$. Fix $a^* > A(\hat{\lambda})$ off the finitely many step ages of either band on $[0, a^* + 1]$ (Lemma D.1: finitely many per compact) and $\delta > 0$ with $(a^*, a^* + \delta]$ free of step ages of either band. List both bands' step ages in $(0, a^*]$, together with a^*, as $0 = g_0 < g_1 < \cdots < g_M = a^*$, and choose a nondecreasing integer staircase $0 = n_0 \le n_1 \le \cdots \le n_M = u_1(a^*)$ with $l_1(g_r^+) \le n_r \le u_1(g_{r-1}^+)$ for every r (take $n_r = u_1(g_{r-1}^+)$, nondecreasing by the monotone staircases of Lemma D.1; a^* generic, off the finite step-age set), so the count's cell extremes are certified against the ceiling in force during the cell—such a staircase exists because along the staircase each floor reaches a given integer only after the corresponding ceiling has passed it (Lemma D.1), the construction of the refinement E' in the proof of Theorem 1. Define

$$D = \left\{N_{g_r} - N_{g_{r-1}} = n_r - n_{r-1},\ r = 1, \ldots, M\right\} \cap \left\{N_{a^*+\delta} - N_{a^*} = 1\right\}.$$

On D no exit precedes the final arrival: within each cell (g_{r-1}, g_r) both bands are constant, the count rises from n_{r-1} to n_r, and $l_1(g^+_{r-1}) \leq n_{r-1} \leq N_a \leq n_r \leq u_1(g^+_{r-1})$ holds throughout the cell because the staircase endpoints lie in the narrower band and the count is monotone; membership in the narrower band implies membership in the wider; and at each step age g_r the post-step membership is the grid condition itself, exactly as in (D.18). The single arrival in $(a^*, a^* + \delta]$ takes the count to $u_1(a^*) + 1$, which exceeds $u_1(a^*)$—investor 1 exits—and is at most $u_2(a^*) - 1$—investor 2 remains *strictly interior* and, the gap being step-free, does not exit. The regime therefore ends at a lone exit on D. The cells and the gap are disjoint intervals, so under $\mathbb{P}^{\lambda^\star}$ the increments are independent Poisson variables and

$$\mathbb{P}^{\lambda^\star}(D) = \Big[\prod_{r=1}^{M} e^{-\lambda^\star \Delta_r} \frac{(\lambda^\star \Delta_r)^{n_r - n_{r-1}}}{(n_r - n_{r-1})!}\Big] \cdot \lambda^\star \delta \, e^{-\lambda^\star \delta} =: \varepsilon(\hat{\lambda}) > 0, \qquad \Delta_r = g_r - g_{r-1},$$

and $\bar{\varepsilon} := \min_{\theta \in \Theta} \varepsilon(\hat{\lambda}_\theta) > 0$ over the finite menu. The grid and staircase depend only on $\hat{\lambda}$ and the two tolerances, and D depends only on the post-ς increments, so by the strong Markov property $\mathbb{P}(\text{the regime ends at a lone exit} \mid \mathcal{F}_\varsigma) \geq \mathbb{P}(D \mid \mathcal{F}_\varsigma) = \varepsilon(\hat{\lambda}) \geq \bar{\varepsilon}$. With C_n the event that a synchronized phase's first n regimes all end jointly, the iterated conditioning of the display in Step 2 of the proof of Lemma B.1, with $1 - \bar{\varepsilon}$ in place of $\bar{q}$, gives $\mathbb{P}(C_n) \leq (1 - \bar{\varepsilon})^n$, so the phase spans a.s. finitely many regimes; each regime has a.s.-finite length (Proposition 2(b)), so the phase ends, at a lone exit, in a.s.-finite time.

Step 2 (part (b): the arrival ladder). Run the event D of Step 1 in the first regime, whose held model $\arg\max \rho_0$ and rate $\hat{\lambda}_0$ are deterministic. On D, at the arrival date $s_0 \in (a^*, a^* + \delta]$: investor 1 exits and re-selects some model of rate $\hat{\lambda}'$ (random, menu-valued); investor 2 holds her original window with count $c = u_1(a^*) + 1$ and slack $k := u_2(a^*) - c \geq 1$ to her ceiling, finite. Two deterministic gaps control the construction: every fresh regime's band, whatever its menu rate, has region $\{0\}$ on an initial age span—by Lemma D.1 the held law at age 0 is a point mass at 0—whose length, the first band-step age, is bounded below by $\bar{t} :=$ min over the finite menu of the first step ages, $\bar{t} > 0$; and investor 2's band has finitely many step ages on any compact, so there is $\bar{g} > 0$ with $(a^*, a^* + \bar{g}]$ free of her step ages. Shrink δ in Step 1 so that $\delta < \frac{1}{2} \min(\bar{t}, \bar{g})$, and fix the deterministic dates $s_j = a^* + \delta + j \cdot \frac{\delta}{k+1}$, $j = 1, \ldots, k+1$, all inside $(a^*, a^* + 2\delta] \subseteq (a^*, a^* + \bar{g}]$ and with consecutive gaps below $\bar{t}$. Augment D with the requirement of exactly one arrival in each $(s_j - \frac{\delta}{2(k+1)}, s_j]$, $j = 1, \ldots, k+1$, and no arrivals in $(a^* + \delta,\ s_{k+1}]$ outside those sub-intervals; the augmented event $D^\dagger$ again prescribes finitely many Poisson increments over disjoint intervals, so $\mathbb{P}^{\lambda^\star}(D^\dagger) > 0$. On $D^\dagger$, by induction on $j = 1, \ldots, k$: just before the j-th ladder arrival investor 1 is in a regime younger than $\bar{t}$, so her region is $\{0\}$, her count is $0 = u$, and the arrival exits her alone—investor 2's count becomes $c + j \le u_2(a^*) - 1 < u_2$ for $j < k$ and exactly $u_2(a^*)$ at $j = k$, in the closed band either way, her band step-free on the span—after which investor 1 re-selects and restarts, her new region again $\{0\}$ for at least $\bar{t}$, covering the next gap. After the k-th ladder arrival the configuration is: investor 1 fresh at her ceiling 0; investor 2 *at* her ceiling $u_2(a^*)$; adoption dates s_k and 0, differing. The $(k+1)$-st arrival, landing before either band steps, takes investor 1's count to $1 > 0$ and investor 2's to $u_2(a^*) + 1 > u_2(a^*)$: a common disaster exits both at one date, from differing adoption dates. This proves the first claim with positive probability. For the second: at any joint exit at which the two re-selections return the same model, both records restart at the common date with the common model, and Step 0 of the proof of Lemma B.1 applies forward: a common regime. Past consolidation with the truth on the menu, Proposition 2(a) makes every re-selection $\theta^\star$, so the proviso holds at every joint exit there.

Step 3 (part (c)). Fix the a.s. event on which every synchronized phase ends at a lone exit (part (a)), both investors' exits recur with locally finite dates (Proposition 2(b), Lemma D.2), and the two-sided nesting of Lemma D.1 holds at every age. Call a stretch *synchronized* only under full operative-state coincidence, the same held model and adoption date; equal-date episodes with different held models or rankings count as desynchronized, which strengthens the conclusion, and beyond the deterministic age a_0 at which the two ceilings differ by an integer, Step 1's exit is lone. The path decomposes into alternating synchronized and desynchronized stretches. Two exhaustive cases: if the total number of stretches is finite, the last is desynchronized—a final synchronized stretch would end at a lone exit by part (a) and be followed by a desynchronized one—and is unbounded, hence infinite (Case B); otherwise the stretches alternate without bound and the synchronized ones recur infinitely often (Case A).

Case A. Let ς_k be the adoption date opening the k-th synchronized stretch's first regime, $\mathcal{G}_k = \mathcal{F}_{\varsigma_k}$, and A_k the event D of Step 1 run in that regime, defined on $\{\text{at least } k \text{ synchronized stretches}\} \in \mathcal{G}_k$. Step 1 gives $\mathbb{P}(A_k \mid \mathcal{G}_k) \ge \bar{\varepsilon}$ there, so on $\{\text{infinitely many synchronized stretches}\}$ we have $\sum_k \mathbb{P}(A_k \mid \mathcal{G}_k) = \infty$; by Lévy's extension of the Borel–Cantelli lemma (Williams, 1991, §12.15), the events

$\{\sum_k \mathbb{P}(A_k \mid \mathcal{G}_k) = \infty\}$ and $\{A_k$ infinitely often$\}$ agree up to a null set, so D occurs at infinitely many stretch openings. Each occurrence opens a desynchronized stretch in the configuration of Step 1: investor 1 fresh, her region $\{0\}$ on an initial span so her count 0 sits at her ceiling; investor 2 strictly interior; and the open interval from the opening to the earlier of the first subsequent arrival and the first subsequent step age of either band is nonempty—the first arrival after a stopping time is a.s. strictly later, and the finitely many step ages on any compact leave the next one strictly later—and on it the counts and bands, hence the configuration, are constant.

Case B. Within the infinite desynchronized stretch every exit is lone, and each investor restarts infinitely often; let r_k be the k-th restart date of either investor within the stretch, $\mathcal{H}_k = \mathcal{F}_{r_k}$, and F_k the event that the first arrival after r_k lands before the restarted investor's first band-step age, an age bounded below by $\bar{t}$ of Step 2. Then $\mathbb{P}(F_k \mid \mathcal{H}_k) \geq 1 - e^{-\lambda^\star \bar{t}} =: \tilde{p} > 0$, and by the same lemma (Williams, 1991, §12.15), on the event that the stretch is infinite, F_k occurs infinitely often. On F_k the restarted investor's region is still $\{0\}$ when the arrival lands, so her count 0 sits at her ceiling and the arrival exits her; were the other investor also at her ceiling at that instant, the arrival would take both counts strictly above their ceilings and the exit would be joint, ending the stretch—excluded in Case B—so the other investor, in her closed band and not at its ceiling, satisfies $N \leq u - 1$: strictly interior. The configuration therefore holds on the open interval from the later of r_k and the last step age of either band before the arrival, to the arrival, on which the counts and bands are constant; the interval is nonempty because the arrival is a.s. neither the restart date nor a step age. In both cases the configuration dates are unbounded—infinitely many of them among locally finite event dates—and lie within desynchronized stretches, so the adoption dates differ on them. □

Proof of Corollary 2. Part (i), Step 1 (same model, same posteriors). With the truth on the menu the trend channel's entropy rate is $\kappa(\theta) = (\mu^\star - \hat{\mu}_\theta)^2/(2\sigma^2)$, zero at $\theta^\star$ and positive elsewhere, so the projection is $\theta^\circ = \theta^\star$, uniquely. By Proposition 2(a) applied to investor i there is an a.s.-finite T_i° with $\arg\max \rho_\varsigma^{(i)} = \theta^\star$ at every revision date $\varsigma \geq T_i^\circ$; by Proposition 2(b) investor i has an a.s.-finite revision date $\varsigma_i^\# \geq T_i^\circ$, and every regime she begins at or after $\varsigma_i^\#$ holds $\theta^\star$. Both investors observe one path, and a posterior formed by updating a fixed prior on $\mathcal{F}_t$ is a measurable functional of the path up to t, hence literally one random variable shared by the two.

Part (i), Step 2 (ages differ). $z_\alpha = \Phi^{-1}(1 - \alpha/2)$ is strictly decreasing in α, so distinct tolerances give $z_1 < z_2$ and Lemma B.1 applies: $S^\dagger < \infty$ a.s. and $a_t^{(1)} \neq a_t^{(2)}$ for every $t \geq S^\dagger$. Set $S = \max\bigl(S^\dagger, \varsigma_1^{\#}, \varsigma_2^{\#}\bigr)$, a.s. finite. For $t > S$: both hold $\theta^\star$, the ages differ, and Step 1's posterior identity holds.

Part (i), Step 3 (the windows). Let $s > S$ be an exit date of investor i and let $\tau^{j\text{-next}}$ be investor j's first exit after s; the window is $W_s = (s,\, (s + a_0) \wedge \tau^{j\text{-next}})$. It is nonempty and open a.s.: $\tau^{j\text{-next}} \geq s$, and $\tau^{j\text{-next}} = s$ is a coincident exit, null by Lemma B.1(b). Fix $t \in W_s$ and let $\tau^{(i)}, \tau^{(j)}$ denote the two investors' first exits strictly after t. Since $t < \tau^{j\text{-next}}$, investor j has not exited in $(s, t]$, so $\sigma^{(j)}(t) = \sigma^{(j)}(s) < s$—adoption at s itself being the null coincidence—and her onset date satisfies

$$o_j = \sigma^{(j)}(t) + a_0 < s + a_0.$$

Part (i), Step 4 (the zero-mass side). Investor i's current regime was adopted at s, and pathwise $\tau \geq a_0$ in (5), so $\tau^{(i)} \geq s + a_0$ identically on the regime; hence for every Borel $A \subseteq (t,\, s + a_0)$,

$$\mathbb{P}\bigl(\tau^{(i)} \in A \,\big|\, \mathcal{F}_t\bigr) = 0.$$

Part (i), Step 5 (the positive-mass side). Two exhaustive sub-cases. *If* $t \geq o_j$: fix any $h \in (0,\, s + a_0 - t)$. Investor j is past probation and in her closed band at t; on $\{|V_{t+h}^{(j)}| > z_j\}$ she is outside her band at $t + h$, so her first exit after t is at most $t + h$: $\{|V_{t+h}^{(j)}| > z_j\} \subseteq \{\tau^{(j)} \in (t, t + h]\}$. Decomposing $V_{t+h}^{(j)}$ at t as in Step 3 of the proof of Lemma B.1, its $\mathcal{F}_t$-conditional law is Gaussian with variance $h/(t + h - \sigma^{(j)}) > 0$, so

$$\mathbb{P}\bigl(\tau^{(j)} \in (t, t+h] \,\big|\, \mathcal{F}_t\bigr) \geq \mathbb{P}\bigl(|V_{t+h}^{(j)}| > z_j \,\big|\, \mathcal{F}_t\bigr) > 0.$$

If $t < o_j$: choose $h \in [\, o_j - t,\ s + a_0 - t)$, nonempty since $o_j < s + a_0$. Exit before o_j is barred by probation, and on $\{|V_{o_j}^{(j)}| > z_j\}$ the record is strictly outside at the first tested age, so the infimum in (5) equals o_j and $\tau^{(j)} = o_j \in (t, t + h]$. The $\mathcal{F}_t$-conditional law of $V_{o_j}^{(j)}$ is Gaussian with variance $(o_j - t)/a_0 > 0$ by the same decomposition, so

$$\mathbb{P}\bigl(\tau^{(j)} \in (t, t+h] \,\big|\, \mathcal{F}_t\bigr) \geq \mathbb{P}\bigl(|V_{o_j}^{(j)}| > z_j \,\big|\, \mathcal{F}_t\bigr) > 0.$$

In both sub-cases investor j's conditional law assigns positive mass to a subset of

$(t,\, s + a_0)$ on which investor i's assigns zero (Step 4): the two $\mathcal{F}_t$-conditional laws differ at every $t \in W_s$.

Part (i), Step 6 (infinite total measure). List the two investors' post-S exit dates jointly as $e_0 < e_1 < e_2 < \cdots$, locally finite and infinite in number, with gaps $g_k = e_{k+1} - e_k$. The window opened at e_k has length at least $\min(a_0, g_k)$: it is capped by $e_k + a_0$ and by the *other* investor's next exit, which is not before e_{k+1}. Among any three consecutive dates e_k, e_{k+1}, e_{k+2}, two belong to the same investor, whose consecutive exits are at least a_0 apart; hence $g_k + g_{k+1} \geq a_0$ for every k. If both $\min(a_0, g_k)$ and $\min(a_0, g_{k+1})$ are below a_0 their sum is $g_k + g_{k+1} \geq a_0$; otherwise one of them equals a_0; either way $\min(a_0, g_k) + \min(a_0, g_{k+1}) \geq a_0$, so the first $2n$ windows have total length at least $n\, a_0 \to \infty$.

Part (i), Step 7 (the alternation). On each W_s opened by investor i's exit, investor i's conditional law is the one vanishing on $(t,\, s + a_0)$; the roles are exchanged on the windows opened by investor j's exits, which also recur by Proposition 2(b). The tolerance pair, the held models, and the shared posteriors are fixed throughout, so the identity of the sooner-supported law is carried by the ages and records alone.

Part (ii). Let S be as in Step 2. By Lemma B.2(c) the ceiling-versus-interior configurations recur with dates unbounded above; all but finitely many of their open intervals lie past S, and each lies within a stretch of differing adoption dates, so on it both investors hold $\theta^\star$, the posterior identity of Step 1 holds, and the regime ages differ. Fix such an interval, a date t in it, and $h > 0$ with $(t, t+h]$ inside the interval; write $N(t, t+h]$ for the number of arrivals in the window, Poisson of mean $\lambda^\star h$ and independent of $\mathcal{F}_t$. Neither band steps on the interval, so scheduled exits are excluded there and exits occur only at arrivals. For the at-ceiling investor, any arrival takes her count one above her ceiling, strictly outside the closed band, and absent arrivals she does not move: $\{\tau \in (t, t+h]\} = \{N(t, t+h] \geq 1\}$, so

$$\mathbb{P}\big(\tau \in (t, t+h] \,\big|\, \mathcal{F}_t\big) = 1 - e^{-\lambda^\star h}.$$

For the strictly interior investor, $N_t \leq u - 1$, so the first arrival takes her count to at most u, inside the closed band, and an exit in the window requires at least two arrivals: $\{\tau \in (t, t+h]\} \subseteq \{N(t, t+h] \geq 2\}$ and

$$\mathbb{P}\big(\tau \in (t, t+h] \,\big|\, \mathcal{F}_t\big) \leq 1 - e^{-\lambda^\star h} - \lambda^\star h\, e^{-\lambda^\star h} < 1 - e^{-\lambda^\star h},$$

the gap $\lambda^\star h\, e^{-\lambda^\star h}$ being strictly positive at every $h > 0$. The two conditional laws therefore differ at every t of the interval. At each interval's opening—the event D of Lemma B.2 in Case A, the post-restart event F in Case B—the at-ceiling investor is the one whose regime has just begun: the just-revised investor is the exposed one, the inversion of part (i)'s probation immunity. $\square$

C The revision hazard as a measure

Section 3.2 derives the hazard from the first-exit law and states the marginal/conditional distinction; this appendix records the hazard measure and its channel split. Throughout, the investor's type $(\hat\theta, \alpha)$ is fixed: every object below is a fibre object at that type, the tolerance entering through the band endpoints u_α, l_α, and we suppress the type in H, S, and $h(a)$ only where the fibre is unambiguous—no cross-type average appears in this appendix. In the disputed-crash-rate channel the band endpoints (7) step at discrete ages (Lemma D.1), so revision carries point masses and the full object is the generalized hazard measure

$$H(da) = h(a)\, da + \sum_j p_j\, \delta_{t_j}(da), \qquad p_j = \mathbb{P}\big(\tau = t_j \mid \tau > t_j^-\big), \tag{C.1}$$

with $\{t_j\}$ the deterministic band-step dates, and model-retention survival the product integral

$$S(a) = \exp\Big(-\int_0^a h(u)\, du\Big) \prod_{t_j \le a} (1 - p_j), \qquad \Delta S(t_j) = -p_j\, S(t_j^-). \tag{C.2}$$

The marginal hazard splits by channel. *Crash channel:* the inaccessible upper exits carry a genuine state-dependent jump rate $\chi(\hat\theta, x, a; \alpha) = \lambda^\star\, \mathbf{1}\{x = u_\alpha(a)\}$—the next disaster is the exit only when the count already sits on the ceiling—and the absolutely continuous fibre hazard $h(\hat\theta, \alpha, a)$ is the average of this rate over the survivor law $\nu_a(\cdot \mid \hat\theta, \alpha, a)$ of the record position,

$$h(\hat\theta, \alpha, a) = \int \chi(\hat\theta, x, a; \alpha)\, \nu_a(dx \mid \hat\theta, \alpha, a) = \lambda^\star\, \nu_a(\{u_\alpha(a)\} \mid \hat\theta, \alpha, a), \tag{C.3}$$

an ordinary integral precisely because χ is a rate. *Trend channel:* the exit is a *predictable* boundary hit with no absolutely continuous rate, so there is no position-conditional rate to fibre-average; the marginal holding-time hazard, where it exists, is generated by boundary *flux*—the rate at which ν_a leaks across the moving edge—not by integrating a finite χ. We use χ in the trend channel only as a proximity heuristic, never as a rate entering (C.3). The compensator of the revision point process is accordingly mixed: absolutely continuous only where exits are inaccessible (the crash channel's disaster-triggered upper exits), singular where predictable (the trend boundary and the scheduled floor steps). So the fibre hazard $h(\hat{\theta}, \alpha, a)$ is recovered from the holding-time *law* (9)—the marginal age-a revision rate of the type, an average over surviving record positions—while conditional on the full enlarged state the operative objects are the crash-channel rate $\chi(\hat{\theta}, x, a; \alpha)$ together with the scheduled atoms and the trend boundary flux; it is not the survival-conditional expectation of a pathwise intensity, which in the trend channel does not exist. We keep the distributional measure H and the pathwise first-exit distinct throughout; (C.1) is well defined as a measure, and we claim no closed form for h in general.

Proof of Corollary 1. (i). From a fresh adoption of $\hat{\theta}$ at tolerance α the record's law is the environment's, and the acceptance family $A_\alpha(\cdot)$ is a deterministic functional of $(\hat{\theta}, \alpha)$: the highest-density regions of the held law (Lemma A.3), in the crash channel the Sterne staircase of Lemma D.1 with endpoints and step ages deterministic in $(\hat{\theta}, \alpha)$. The first-exit time (5) of a law-determined record from a $(\hat{\theta}, \alpha)$-determined family has a law determined by $(\hat{\theta}, \alpha)$; its hazard (9) at age a and, in the crash channel, the atom dates t_j and masses p_j of (C.1) are therefore functions of $(\hat{\theta}, \alpha, a)$.

(ii), crash channel. Fix an age a with $u_\alpha(a) - l_\alpha(a) \geq 2$ off the finitely many step ages—such ages exist, the width growing like $2z_\alpha\sqrt{\hat{\lambda}a}$ (Lemma D.1). The in-band staircases of the refinement E' in the proof of Theorem 1 reach the ceiling $u_\alpha(a)$, and likewise any slack-two integer $x \leq u_\alpha(a) - 2$, while surviving, so both positions carry positive survivor mass under ν_a. Fix $h > 0$ below the distance to the next step age and write $N(t, t+h]$ for the arrival count, Poisson of mean $\lambda^\star h$, independent of $\mathcal{F}_t$. From the ceiling, an arrival exits and nothing else moves the count: $\{\tau \in (t, t+h]\} = \{N(t, t+h] \geq 1\}$ and the conditional window probability is $1-e^{-\lambda^\star h} = \lambda^\star h + o(h)$. From a slack-two position, the first two arrivals leave the count at most $u_\alpha(a)$, inside the closed band, so $\{\tau \in (t, t+h]\} \subseteq \{N(t, t+h] \geq 3\}$ and the conditional window probability is at most $1-e^{-\lambda^\star h}(1+\lambda^\star h+\frac{1}{2}(\lambda^\star h)^2) = O(h^3) = o(h)$. A conditional law admitting a $(\hat{\theta}, \alpha, a)$-measurable version is ν_a-a.s. equal to a function constant across record positions in the fibre, hence assigns the two atoms—each of

positive ν_a-mass—one common window probability; the displayed probabilities differ for small h, so no such version exists.

(ii), trend channel. Fix $a > a_0$ and pass to the logarithmic-age coordinates used throughout Appendix D.6, in which the *noise component* of the standardized record is the stationary Ornstein–Uhlenbeck $dV_u = -\frac{1}{2}V_u\,du + dW_u$, W a standard Brownian motion, and a calendar window $(t, t+h]$ is the log-age window of deterministic length $\eta = \log\big((a+h)/a\big) > 0$; the exit-time claims transfer along this deterministic, strictly increasing time change. Write $q_\eta(y)$ for the conditional probability of exit within the window from the in-band position y; by the Markov property of the enlarged state (Theorem 1(iii)) the $\mathcal{F}_t$-conditional window-exit probability equals q_η evaluated at the current position, a.s. Fix $\eta' \in (0, \frac{1}{2})$. Three displayed bounds.

Near the edge. Let $y \in (z_\alpha - \delta, z_\alpha)$ with $\delta := \frac{1}{2}z_\alpha\eta$. On the event that no exit has occurred by log-age $u \le \eta$, $|V_r| \le z_\alpha$ for $r \le u$, so

$$V_u = y - \tfrac{1}{2}\int_0^u V_r\,dr + W_u \;\ge\; y - \tfrac{1}{2}z_\alpha\eta + W_u,$$

so on $\big\{\sup_{u\le\eta} W_u > (z_\alpha - y) + \frac{1}{2}z_\alpha\eta\big\}$, at the first u at which W_u exceeds that level either the exit has already occurred or $V_u > z_\alpha$ strictly, and the exit occurs by η in either case; by the reflection principle, whose supremum law is atomless (Karatzas and Shreve, 1991, Chapter 2),

$$q_\eta(y) \;\ge\; \mathbb{P}\Big(\sup_{u\le\eta} W_u \ge (z_\alpha - y) + \tfrac{1}{2}z_\alpha\eta\Big) = 2\,\bar\Phi\Big(\frac{(z_\alpha - y) + \frac{1}{2}z_\alpha\eta}{\sqrt{\eta}}\Big) \;\ge\; 2\,\bar\Phi\big(z_\alpha\sqrt{\eta}\big),$$

using $z_\alpha - y < \delta = \frac{1}{2}z_\alpha\eta$. Since $2\bar\Phi(x) \to 1$ as $x \downarrow 0$, choose $\eta \le \eta_1$ so that $2\bar\Phi(z_\alpha\sqrt{\eta}) \ge 1 - \eta'$: then $q_\eta > 1 - \eta'$ on $(z_\alpha - \delta, z_\alpha)$.

Near the center. Let $|y| \le z_\alpha/4$ and $\eta \le 1$. With $\varphi(u) = \sup_{r\le u}|V_r|$, the same identity gives $\varphi(u) \le |y| + \sup_{r\le u}|W_r| + \frac{1}{2}\int_0^u \varphi(r)\,dr$, and Grönwall's inequality yields $\varphi(\eta) \le \big(|y| + \sup_{u\le\eta}|W_u|\big)e^{\eta/2}$. Hence

$$q_\eta(y) \;\le\; \mathbb{P}\Big(\sup_{u\le\eta}|W_u| \;\ge\; z_\alpha e^{-1/2} - \tfrac{z_\alpha}{4}\Big) \;\le\; 4\,\bar\Phi\Big(\frac{z_\alpha\big(e^{-1/2} - \frac{1}{4}\big)}{\sqrt{\eta}}\Big),$$

the last step bounding the two-sided supremum by twice the one-sided reflection bound; the right side tends to 0 as $\eta \downarrow 0$, so choose $\eta \le \eta_2$ with $q_\eta < \eta'$ on $[-z_\alpha/4,\, z_\alpha/4]$. Fix $\eta \le \min(\eta_1, \eta_2)$, so both bounds hold at one window.

The survivor law charges both. For any nonempty open $I \subseteq (-z_\alpha, z_\alpha)$, the event that the record stays strictly inside the band through age a and ends in I contains a uniform tube around a continuous in-band path ending in I; by the Cameron–Martin–Girsanov theorem (Karatzas and Shreve, 1991, §3.5) the record's law on the window is equivalent to that of its driftless counterpart, under which every such tube has positive probability. Hence $\nu_a(I) > 0$; in particular $\nu_a\big((z_\alpha - \delta, z_\alpha)\big) > 0$ and $\nu_a\big([-z_\alpha/4, z_\alpha/4]\big) > 0$.

Conclusion. A conditional law admitting a $(\hat{\theta}, \alpha, a)$-measurable version is ν_a-a.s. equal to one window probability across positions; the displayed bounds put the probability above $1 - \eta'$ on one positive-ν_a set and below $\eta' < 1 - \eta'$ on another. No such version exists. $\square$

D Proofs for Section 3.3

This appendix proves the results of Section 3.3. Throughout, the held model is $\hat{\theta}$ (a drift $\hat{\mu}$ or a rate $\hat{\lambda}$), the truth is $\mathbb{P}^\star$ (drift $\mu^\star$ or rate $\lambda^\star$), and the band is (6)–(7). We first record the structure of the crash-channel band, used in Proposition 2(b) and Lemma D.3.

Monotonicity. Both endpoints are nondecreasing staircases: for $n_1 < n_2$ the ratio $p_m(n_2)/p_m(n_1) = m^{n_2-n_1} n_1!/n_2!$ increases strictly in m, so each pair's Sterne order flips once, upward, at $m^*(n_1, n_2) = (n_2!/n_1!)^{1/(n_2-n_1)}$, a set finite on compacts; between flips the order is fixed and $\partial_m P_m([l, u]) = p_m(l-1) - p_m(u) \leq 0$ by the initial-segment property, so the region's mass decays, and the region extends exactly at the downward crossings of $1 - \alpha$—the exact-attainment ages—by adjoining the fixed next-ranked point at its margin, while at each boundary-straddling flip both endpoints move up by the one-integer swap (D.1). Jumps are confined to these two finite sets, the exact-attainment and rank-swap ages of the right-continuity convention.

Lemma D.1 (Sterne band structure)**.** *In the disputed-crash-rate channel the band endpoints $l_\alpha(a), u_\alpha(a)$ of (7) are deterministic step functions of age—piecewise constant with finitely many jumps on any compact interval, taken right-continuous by convention at the step ages (a choice that fixes the band only at those finitely many ages; under an infimum definition of the first-exit time it leaves the holding-time* law *unchanged, but it is not immaterial to* membership at the exit time—*a crash-channel lower exit occurs exactly when the deterministic floor steps up past an unchanged integer count, i.e.* at *a step age—and to the bookkeeping of predictable atoms, whose dates are these floor-steps)—and nested in tolerance:* $A_{\alpha'}(a) \subseteq A_\alpha(a)$ *for* $\alpha' > \alpha$ *at every* a. *As* $a \to \infty$ *the endpoints obey the normal approximation* $l_\alpha(a), u_\alpha(a) = \hat\lambda a \mp z_\alpha\sqrt{\hat\lambda a} + o(\sqrt{a})$, *a half-width* $z_\alpha\sqrt{\hat\lambda a} = O(\sqrt{a})$ *about the held mean* $\hat\lambda a$.

Proof. The held law of N_a is Poisson$(\hat\lambda a)$ with masses $p_n(a) = e^{-\hat\lambda a}(\hat\lambda a)^n/n!$. The Sterne region accrues integers in order of descending $p_n(a)$ until the cumulative mass first reaches $1-\alpha$. The region changes with a only when either two integers swap mass rank, $p_n(a) = p_m(a)$ for some $n \neq m$, or the accrued mass crosses $1-\alpha$ at a new integer. The rank-swap condition is

$$p_n(a) = p_m(a),\ n < m,\ a > 0 \iff (\hat\lambda a)^{m-n} = \frac{m!}{n!} \iff a = \frac{1}{\hat\lambda}\Big(\frac{m!}{n!}\Big)^{1/(m-n)}, \quad \text{(D.1)}$$

a single age for each pair, where the equivalence divides by $(\hat\lambda a)^n$ and so requires $a > 0$; at $a = 0$ the held law is the point mass $p_0 = 1$, with no rank-swap. On a compact interval $[0, A]$ the band lies in $\{0, \ldots, N_A^{\max}\}$ with $N_A^{\max}$ finite: writing $U_A := \min\{k : \mathbb{P}^{\mathrm{Poisson}(\hat\lambda A)}(N \le k) \ge 1-\alpha\}$, the benchmark set $\{0, \ldots, U_A\}$ has mass at least $1-\alpha$ under Poisson$(\hat\lambda a)$ for every $a \le A$ (the Poisson family is stochastically increasing in its mean), so the minimal-cardinality region has at most $U_A + 1$ atoms; it contains the modal integer $\lfloor \hat\lambda a \rfloor \le \hat\lambda A$, where the accrual starts, so its ceiling is at most $N_A^{\max} := \lfloor \hat\lambda A \rfloor + U_A + 1$. Hence only finitely many pairs (n, m) and finitely many rank-swap ages (D.1) arise; between consecutive rank-swap ages the descending-mass *ordering* is constant. With the ordering fixed the region can still change at a *cumulative-threshold* age, where the accrued mass of the leading segment first reaches $1-\alpha$ at a new integer. But for a fixed ordering the cumulative mass of any leading segment is a finite sum of terms $e^{-\hat\lambda a}(\hat\lambda a)^n/n!$, real-analytic and non-constant in a, so it equals $1-\alpha$ at finitely many ages on $[0, A]$; finitely many cumulative-threshold ages thus arise per ordering, and finitely many in all. Between consecutive ages of either kind the region is constant. Right-continuity is by the convention of taking the region at a: at a cumulative-mass threshold age or a boundary-straddling rank-swap age (D.1), where the minimal Sterne region jumps, the post-step region is in force from the step age onward—the convention the crash-channel lower-exit and the holding-time transfer matrix below both use, so that a scheduled floor or ceiling step takes effect *at* its date rather than just after it. The convention is a tie-break confined to the finitely many exact-attainment and boundary-straddling rank-swap ages. At an exact-attainment age two probability-ordered regions of mass at least $1-\alpha$ coexist—the minimal one and its one-integer enlargement—and the right-continuous choice takes the enlargement at the date. At a rank-swap age (D.1) whose tied pair straddles the region boundary, two *minimal* probability-ordered regions of equal mass coexist—the pre-swap interval and its one-integer shift, exchanging the lowest included integer for the one above the ceiling—and the right-continuous choice takes the post-swap interval. Minimal cardinality holds at every age off the exact-attainment set, so the bookkeeping at the predictable step dates and Sterne minimality never conflict off these finite sets. Formally two objects coexist: the fixed-age minimal region of Definition 1 and the right-continuous *process* version the dynamics use, equal off the finite set of exact-attainment and boundary-straddling rank-swap ages, differing at the former by the enlargement and at the latter by the one-integer shift of both endpoints; the lower-exit atoms and the transfer matrices below use the process version. For nestedness, $\alpha' > \alpha$ gives a smaller target mass $1-\alpha' < 1-\alpha$; since both regions accrue integers in the same descending-mass order, the α'-region is an initial segment of the α-region, so $A_{\alpha'}(a) \subseteq A_\alpha(a)$ (with equality when the two targets fall within the same mass step). For the asymptotic half-width, the highest-density region of Poisson$(\hat\lambda a)$ of mass $1-\alpha$ converges, under the normal approximation Poisson$(\hat\lambda a) \approx \mathcal{N}(\hat\lambda a, \hat\lambda a)$ valid as $\hat\lambda a \to \infty$ by the local central limit theorem for lattice laws (Petrov, 1975, Ch. VII), to the Gaussian highest-density interval $\hat\lambda a \pm z_\alpha\sqrt{\hat\lambda a}$, giving $l_\alpha(a), u_\alpha(a) = \hat\lambda a \mp z_\alpha\sqrt{\hat\lambda a} + o(\sqrt{a})$. □

D.1 Non-explosion of revisions

Lemma D.2 (Non-explosion of revisions)**.** *Run the per-test rule of Definition 1 in either canonical channel at a fixed tolerance, restarting the record at each adoption. The revision times satisfy $\sigma_k \uparrow \infty$ $\mathbb{P}^\star$-almost surely—the regimes do not accumulate at a finite time. The argument uses only a per-channel lower-tail floor on holding times—deterministic ($\tau_k \geq a_0$) in the trend channel, a positive lower-tail probability in the crash channel (Lemma A.6)—and the finiteness of the menu; it invokes neither consolidation nor the finiteness of any individual τ_k.*

Proof. Each σ_k is a stopping time, and the fundamental has independent increments (Brownian-with-drift in the trend channel, Poisson in the crash channel), so it is strong Markov at σ_k; with the record restarted at adoption (Lemma A.5) the conditional law of $\tau_k = \sigma_{k+1} - \sigma_k$ given $\mathcal{F}_{\sigma_k}$ depends only on the post-revision held model $\hat{\theta}_{\sigma_k}$, which ranges over the *finite* menu Θ; the data are time-homogeneous, so this conditional law is one of finitely many fresh-start laws. Each carries a lower-tail floor (Lemma A.6): $\tau_k \geq a_0$ surely in the trend channel, and $\mathbb{P}(\tau_k \geq \delta \mid \mathcal{F}_{\sigma_k}) \geq e^{-\lambda^\star \delta}$ in the crash channel for all small δ (the wait to the first disaster off the singleton band $\{0\}$). Hence there are $\delta > 0$ and $q \in (0, 1]$ with $\mathbb{P}(\tau_k \geq \delta \mid \mathcal{F}_{\sigma_k}) \geq q$ for every k, so $\sum_k \mathbb{P}(\tau_k \geq \delta \mid \mathcal{F}_{\sigma_k}) \geq \sum_k q = \infty$ a.s., and Lévy's conditional Borel–Cantelli lemma (Williams, 1991, §12.15) gives $\tau_k \geq \delta$ for infinitely many k a.s. Therefore $\sum_k \tau_k = \infty$ and $\sigma_k \uparrow \infty$ a.s. The argument invokes only the floor and the finite menu—not the finiteness of any individual τ_k, nor consolidation—so it is available wherever non-explosion is needed below, with no circularity. □

D.2 Consolidation

Proof of Proposition 2, part (a). Work within $\operatorname{supp}(\rho_0)$: a model with $\rho_0(\theta) = 0$ keeps $\rho_t(\theta) = 0$ under the multiplicative update and never leads, and $\theta^\circ \in \operatorname{supp}(\rho_0)$ since $\rho_0(\theta^\circ) > 0$. By the chaining identity (Lemma A.4), the re-ranking at any revision equals re-ranking on the entire history to that date; in pairwise-ratio form, for $\theta, \theta^\circ \in \operatorname{supp}(\rho_0)$ the ranking ratio carried at a revision time t is

$$\log \frac{\rho_t(\theta)}{\rho_t(\theta^\circ)} = \log \frac{\rho_0(\theta)}{\rho_0(\theta^\circ)} + \log \Lambda_t^{\theta/\theta^\circ}, \qquad \Lambda_t^{\theta/\theta^\circ} := \frac{d\mathbb{P}^\theta}{d\mathbb{P}^{\theta^\circ}}\bigg|_{\mathcal{F}_t}, \tag{D.2}$$

the full-history likelihood ratio, the normalizing constant having cancelled. Between revisions ρ is frozen (Definition 1); the object that evolves continuously is the latent ratio $\Lambda_t^{\theta/\theta^\circ}$, and the law of large numbers below is a statement about it. Under each menu law the record has *stationary independent increments* (Brownian motion with drift, or a Poisson process), so the log-likelihood-ratio process $t \mapsto \log \Lambda_t^{\theta/\theta^\circ}$ is, under $\mathbb{P}^\star$, an additive functional with stationary independent increments. Write $\kappa(\theta) := \lim_{t\to\infty} \frac{1}{t} \operatorname{KL}(\mathbb{P}_t^\star \| \mathbb{P}_t^\theta)$ for the per-unit-time relative entropy *rate*—the finite-horizon divergence over horizon length, the finite object here, since the full-path laws are mutually singular and the infinite-horizon divergence is not finite. The one-unit log-likelihood-ratio increment has finite first moment under $\mathbb{P}^\star$ in both channels—it is Gaussian on the trend record and an affine function of a Poisson count on the crash record, by (A.3)–(A.4)—so the rates $\kappa(\theta), \kappa(\theta^\circ)$ are finite. Then the drift per unit time is

$$\mathbb{E}^\star\big[d \log \Lambda_t^{\theta/\theta^\circ}\big] = \big(\kappa(\theta^\circ) - \kappa(\theta)\big)\, dt. \tag{D.3}$$

By the strong law of large numbers for processes with stationary independent increments with integrable unit increment (Sato, 1999, Thm. 36.5)—applied directly to calendar time rather than to a fixed grid—

$$\frac{1}{t} \log \Lambda_t^{\theta/\theta^\circ} \xrightarrow{\text{a.s.}} \kappa(\theta^\circ) - \kappa(\theta) \qquad (t \to \infty). \tag{D.4}$$

By the uniqueness of θ° as the Kullback–Leibler minimizer (in rate), $\kappa(\theta) > \kappa(\theta^\circ)$ for $\theta \neq \theta^\circ$, so the limit (D.4) is strictly negative; this is the continuous-time form of the concentration of the misspecified posterior on the KL projection (Berk, 1966). Hence $\log \Lambda_t^{\theta/\theta^\circ} \to -\infty$ a.s. for every $\theta \neq \theta^\circ$ in the (finite) support. This is a statement about the latent ratio in continuous *calendar* time, and it invokes no revisions. For each $\theta \neq \theta^\circ$ there is therefore an a.s. finite time after which $\log(\rho_0(\theta)/\rho_0(\theta^\circ)) + \log \Lambda_t^{\theta/\theta^\circ} < 0$; taking the maximum over the finite support, there is an a.s. finite calendar time T° such that, for every $t \geq T^\circ$ and every $\theta \neq \theta^\circ$,

$$\log \frac{\rho_0(\theta)}{\rho_0(\theta^\circ)} + \log \Lambda_t^{\theta/\theta^\circ} < 0. \tag{D.5}$$

This is an inequality for the *latent* full-history odds; the carried ranking ρ is frozen between revisions, and equals the latent odds exactly at every revision time (Lemma A.4), so (D.5) is transferred to ρ at the revision times and only there. Hence at *any* revision occurring at a time $\sigma_k \geq T^\circ$, $\arg\max_\theta \rho_{\sigma_k}(\theta) = \theta^\circ$ a.s.—the lemma's conditional implication, which asserts nothing about whether such a revision occurs (that is supplied by part (b); non-explosion, Lemma D.2, uses only the holding-time floor of Lemma A.6 and the finite menu, not part (a)). When the truth is on the menu, θ° is the truth, the unique zero of $\mathrm{KL}(\mathbb{P}^\star \| \cdot)$. Positive prior mass on the consolidation target is necessary—$\rho_0(\theta^\circ) > 0$, not full support: if $\rho_0(\theta^\circ) = 0$ then $\rho_t(\theta^\circ) = 0$ for all t by the multiplicative update, and θ° is never selected; absent the truth from the support, consolidation is to the KL-closest model within $\mathrm{supp}(\rho_0)$. $\square$

D.3 Recurrence

Proof of Proposition 2, part (b). Trend channel, correct held model ($\hat{\mu} = \mu^\star$). By (A.9) the band is the event $\{|W_a|/\sqrt{a} \le z_\alpha\}$, i.e. Brownian motion against the square-root boundary $z_\alpha\sqrt{a}$, the first-exit problem of Breiman (1967). Put $\tau' = \log a$ and

$$V_{\tau'} := e^{-\tau'/2}\, W_{e^{\tau'}}, \qquad \text{so that} \quad R_a = \frac{W_a}{\sqrt{a}} = V_{\log a}. \tag{D.6}$$

The time-changed process $M_{\tau'} := W_{e^{\tau'}}$ has quadratic variation $d\langle M\rangle_{\tau'} = d(e^{\tau'}) = e^{\tau'}\, d\tau'$, so $dM_{\tau'} = e^{\tau'/2}\, dB_{\tau'}$ for a standard Brownian motion B. By Itô's product rule applied to (D.6),

$$dV_{\tau'} = -\tfrac{1}{2} e^{-\tau'/2} M_{\tau'}\, d\tau' + e^{-\tau'/2}\, dM_{\tau'} = -\tfrac{1}{2} V_{\tau'}\, d\tau' + dB_{\tau'}, \tag{D.7}$$

so V is an Ornstein–Uhlenbeck process with mean-reversion $\frac{1}{2}$ and unit diffusion, whose invariant law is $\mathcal{N}\big(0, 1/(2\cdot\frac{1}{2})\big) = \mathcal{N}(0,1)$. The exit set $E = \{|v| > z_\alpha\}$ has invariant measure

$$\int_{\{|v|>z_\alpha\}} \varphi(v)\, dv = 2\,\Phi(-z_\alpha) = \alpha > 0, \tag{D.8}$$

φ the standard normal density and the last equality by $z_\alpha = \Phi^{-1}(1-\alpha/2)$. The OU process is a positive-recurrent (ergodic) diffusion (Meyn and Tweedie, 2009), so by the ergodic theorem

$$\frac{1}{T}\int_0^T \mathbf{1}\{|V_s| > z_\alpha\}\, ds \ \xrightarrow{\text{a.s.}}\ \alpha > 0, \tag{D.9}$$

hence $\{\tau' : |V_{\tau'}| > z_\alpha\}$ is unbounded a.s.; by continuity of V the first entry $H := \inf\{\tau' > \log a_0 : |V_{\tau'}| > z_\alpha\}$ is finite a.s., and therefore the holding time $\tau = e^H < \infty$ a.s.

Crash channel, correct held model ($\hat{\lambda} = \lambda^\star$). The centered count $M_a := N_a - \lambda^\star a$ has, at integer ages, increments $\xi_i := (N_i - N_{i-1}) - \lambda^\star$ that are i.i.d., mean zero, variance $\lambda^\star$. By the Hartman–Wintner law of the iterated logarithm (Hartman and Wintner, 1941) applied to the integer-age walk M_n, and interpolated over unit intervals—within which the increment is at most the unit-interval count, whose maximum to age a grows as $O(\log a)$ a.s. by Borel–Cantelli against the Poisson tail, and

$\log a = o(\sqrt{a \log\log a})$—one has

$$\limsup_{a\to\infty} \frac{N_a - \lambda^\star a}{\sqrt{2\lambda^\star a\ \log\log a}} = 1, \qquad \liminf_{a\to\infty} \frac{N_a - \lambda^\star a}{\sqrt{2\lambda^\star a\ \log\log a}} = -1 \quad \text{a.s.} \tag{D.10}$$

By Lemma D.1 the band half-width about the held mean is of order $z_\alpha\sqrt{\lambda^\star a} = O(\sqrt{a})$, and since

$$\frac{\sqrt{2\lambda^\star a\ \log\log a}}{z_\alpha\sqrt{\lambda^\star a}} = \frac{\sqrt{2\log\log a}}{z_\alpha} \to \infty \quad (a\to\infty), \tag{D.11}$$

$N_a - \lambda^\star a$ exceeds the ceiling deviation $+z_\alpha\sqrt{\lambda^\star a}$ for a sequence of ages tending to infinity a.s. by (D.10); the first such age is finite, so $\tau < \infty$ a.s.

Incorrect held model (held rate $\neq$ truth). The two paragraphs above dispose of a *correct* held model in either channel. An incorrect held model exits as well: its standardized statistic carries a divergent drift. In the crash channel, holding a rate $\hat\lambda \neq \lambda^\star$, $N_a - \hat\lambda a = (\lambda^\star - \hat\lambda)a + M_a$; by the law of large numbers $N_a/a \to \lambda^\star$ a.s., so $M_a = o(a)$ a.s., and with the band half-width $z_\alpha\sqrt{\hat\lambda a} = O(\sqrt{a}) = o(a)$ (Lemma D.1) the signed distance from the count to the relevant edge is

$$N_a - \big(\hat\lambda a \pm z_\alpha\sqrt{\hat\lambda a}\big) = (\lambda^\star - \hat\lambda)a + M_a \mp z_\alpha\sqrt{\hat\lambda a} = (\lambda^\star - \hat\lambda)a + o(a) \ \xrightarrow{\text{a.s.}}\ \pm\infty$$

according as $\hat\lambda \lessgtr \lambda^\star$, so the count crosses the ceiling (if $\hat\lambda < \lambda^\star$) or the floor (if $\hat\lambda > \lambda^\star$) at a finite age a.s. In the trend channel, holding $\hat\mu \neq \mu^\star$, $R_a = \sqrt{a}\,(\mu^\star - \hat\mu)/\sigma + W_a/\sqrt{a}$ with $|W_a|/\sqrt{a} \le \sqrt{2\log\log a}\,(1+o(1)) = o(\sqrt{a})$ a.s. by the law of the iterated logarithm, so $R_a = \sqrt{a}\,(\mu^\star - \hat\mu)/\sigma\,(1+o(1)) \to \pm\infty$ and crosses the fixed band $\pm z_\alpha$ at a finite age a.s.

So *every* held model exits a.s.: a correct rate by the Ornstein–Uhlenbeck/law-of-the-iterated-logarithm recurrence above, an incorrect rate by this divergent drift—and this is established *without* reference to which model is in force at any particular revision, so it invokes neither consolidation nor the existence of later revisions.

Iteration and non-explosion. Each revision time σ_k is a stopping time, and the fundamental's independent increments make it strong Markov at σ_k; with the record restarted at adoption (Lemma A.5) the regime adopted at σ_k monitors a fresh record, and the conditional law of the holding time $\tau_k = \sigma_{k+1} - \sigma_k$ given $\mathcal{F}_{\sigma_k}$ depends only on the post-revision held model $\hat\theta_{\sigma_k}$. Every regime exits a.s. in finite time—a correct held model by the Ornstein–Uhlenbeck/law-of-the-iterated-logarithm arguments above, an incorrect one by the divergent-drift argument above—so each $\tau_k < \infty$ a.s., and by induction $\sigma_k < \infty$ for every finite k. Each τ_k is also strictly positive (Lemma A.6): the trend channel carries a probation onset $a_0 > 0$, so $\tau_k \geq a_0$; the crash channel monitors the singleton band $\{0\}$ at small ages, so an exit there waits for the first disaster.

By Lemma D.2—whose lower-tail floor (Lemma A.6) and finite menu are exactly the strict positivity and restart structure just established—the revision times do not accumulate: $\sigma_k \uparrow \infty$ a.s. With each $\tau_k < \infty$ a.s. this gives infinitely many a.s.-finite regimes, so the model is abandoned infinitely often, yielding (a) and (b) directly, without conditioning on the future. Part (a) then composes: it supplies an a.s. finite time T° after which every revision re-selects θ°, and since $\sigma_k \uparrow \infty$ there is a finite K with $\sigma_K \geq T^\circ$, so every regime adopted from σ_K on begins with θ° and is abandoned again—a correct-on-average model revised forever. This proves (i), (ii), and the post-consolidation statement. □

D.4 The holding-time law

Lemma D.3 (Holding-time law)**.** *Partition* $[0, t]$ *as* $0 = t_0 < t_1 < \cdots < t_{J-1} < t_J = t$*, where* $t_1, \ldots, t_{J-1}$ *are the ages in* $(0, t)$ *at which the Sterne band* (7) *changes (finitely many) and* $t_J = t$ *is the terminal endpoint. In the disputed-crash-rate channel, the model-retention survival is the finite matrix product*

$$\mathbb{P}(\tau > t) = e_0^\top \Big(\prod_{j=1}^{J} e^{Q_j s_j} \, P_j \Big) \mathbf{1}, \tag{D.12}$$

one factor per interval $(t_{j-1}, t_j]$*, taken left to right in increasing* j*, where:* $s_j = t_j - t_{j-1}$ *is the interval length;* Q_j *is the sub-generator on the band* $[l_j, u_j]$ *in force on the open interval* (t_{j-1}, t_j)*—*P_j *imposing the right-continuous band* at t_j*—of the disaster count killed on leaving that band, with entries*

$$(Q_j)_{n,n} = -\lambda^\star, \quad (Q_j)_{n,n+1} = \lambda^\star \ (n < u_j), \qquad l_j \leq n \leq u_j, \tag{D.13}$$

the up-move from u_j *being the killing;* P_j $(j < J)$ *is the transfer matrix embedding the surviving states of band* j *into band* $j+1$ *at the true band change* t_j*, and* $P_J = I$ *at the terminal endpoint* $t_J = t$ *(replaced by the transfer matrix in the coincidental* *case that a band change falls exactly at t); e_0 places the count at its adoption value and $\mathbf{1}$ sums over surviving states. The expression is exact—no approximation—and finite at every age, with $(e^{Q_j s})_{n,n'} = e^{-\lambda^\star s}(\lambda^\star s)^{n'-n}/(n'-n)!$ for $l_j \le n \le n' \le u_j$.*

Proof of Lemma D.3. Fix t and partition $[0,t]$ as $0 = t_0 < t_1 < \cdots < t_{J-1} < t_J = t$, with $t_1, \ldots, t_{J-1}$ the band-step dates of Lemma D.1 in $(0,t)$ (finite in number) and $t_J = t$ the terminal endpoint. On $(t_{j-1}, t_j]$ the band is the constant interval $[l_j, u_j]$. The disaster count N, killed on leaving $[l_j, u_j]$, is a pure-birth chain on $\{l_j, \ldots, u_j\}$: from $n < u_j$ it moves to $n+1$ at rate $\lambda^\star$; from u_j the next disaster is the killing (upper exit). This is an age-inhomogeneous phase-type construction—a finite matrix-product whose state space changes at the band-step ages, not a time-homogeneous phase-type with a fixed generator and exponential tail (Neuts, 1981). Its sub-generator is (D.13). The probability of surviving the interval, transporting the count's distribution from t_{j-1} to t_j, is the killed semigroup $e^{Q_j s_j}$, $s_j = t_j - t_{j-1}$. At the step date t_j the band changes to $[l_{j+1}, u_{j+1}]$ and the surviving states embed into the new state space through the transfer matrix P_j (the identity on retained states, zero on states the new floor or ceiling excludes—a lower exit at the step). By this convention $e^{Q_j s_j}$ carries the count over the half-open $(t_{j-1}, t_j]$ under the old band and P_j imposes the new band *at* t_j, so the deterministic floor-step lower exit is counted as an exit at t_j and the product below returns the post-atom survival $\mathbb{P}(\tau > t_j)$; the left-limit $\mathbb{P}(\tau > t_j^-)$ omits P_j and exceeds it by that atom's mass. (Endpoint convention, explicitly: the product returns the right-continuous, post-atom survival $\mathbb{P}(\tau > t_j)$; the pre-step law is its left limit—the same right-continuity as Lemma D.1.) Composing from $j = 1$ to J, starting from e_0 (the count at its adoption value) and summing over surviving states with $\mathbf{1}$,

$$\mathbb{P}(\tau > t) = e_0^\top \Big(\prod_{j=1}^{J} e^{Q_j s_j} P_j \Big) \mathbf{1}, \tag{D.14}$$

with $s_j = t_j - t_{j-1}$, the transfer P_j ($j < J$) at the true band change t_j, and $P_J = I$ at the terminal endpoint $t_J = t$ (the transfer matrix in the coincidental case of a band change exactly at t). This is (D.12). For the entries of $e^{Q_j s}$, the matrix Q_j is upper bidiagonal with $-\lambda^\star$ on the diagonal and $\lambda^\star$ on the superdiagonal, so $\Phi(s) := e^{Q_j s}$ solves $\Phi'(s) = Q_j \Phi(s)$, $\Phi(0) = I$, i.e. for $l_j \le n \le n' \le u_j$,

$$\frac{d}{ds}\Phi_{n,n'}(s) = -\lambda^\star \, \Phi_{n,n'}(s) + \lambda^\star \, \Phi_{n+1,n'}(s). \tag{D.15}$$

The candidate $\Phi_{n,n'}(s) = e^{-\lambda^\star s}(\lambda^\star s)^{n'-n}/(n'-n)!$ satisfies (D.15): with $k = n' - n$,

$$\frac{d}{ds}\Big[e^{-\lambda^\star s}\tfrac{(\lambda^\star s)^k}{k!}\Big] = -\lambda^\star e^{-\lambda^\star s}\tfrac{(\lambda^\star s)^k}{k!} + \lambda^\star e^{-\lambda^\star s}\tfrac{(\lambda^\star s)^{k-1}}{(k-1)!} = -\lambda^\star\,\Phi_{n,n'}(s) + \lambda^\star\,\Phi_{n+1,n'}(s), \tag{D.16}$$

since $\Phi_{n+1,n'}$ corresponds to $k-1$ (with the conventions $1/(k-1)! := 0$ at $k=0$, so the diagonal entry solves $\Phi'_{n,n} = -\lambda^\star\Phi_{n,n}$, and $\Phi_{n,n'} := 0$ for indices outside $[l_j, u_j]$, so the top *starting* state $n = u_j$ takes no super-diagonal inflow), and $\Phi_{n,n'}(0) = \mathbf{1}\{n = n'\}$ gives $\Phi(0) = I$. So $(e^{Q_j s})_{n,n'} = e^{-\lambda^\star s}(\lambda^\star s)^{n'-n}/(n'-n)!$, the Poisson probability of $n' - n$ births in time s, as claimed. The product (D.14) is finite and exact at every t. $\square$

D.5 A stochastic-process toolkit

Three lemmas isolate the boundary-crossing steps the tail and re-selection results rest on, each with checkable hypotheses. Two discrete errors of the counting channel are named where they enter: the $o(\sqrt{a})$ gap between integer Sterne endpoints and continuous square-root boundaries, and the one-count jump overshoot at an upper exit.

Lemma D.4 (Poisson exit through square-root boundaries)**.** *Let N be a Poisson process of rate $\lambda > 0$, let $l \le u$ be nondecreasing integer staircases with $l(a) = \lambda a - z\sqrt{\lambda a} + o(\sqrt{a})$ and $u(a) = \lambda a + z\sqrt{\lambda a} + o(\sqrt{a})$ for some $z > 0$, and let $T = \inf\{a > a_0 : N_a \notin [l(a), u(a)]\}$, with any finite-age convention on $[0, a_0]$. Then*

$$\mathbb{P}(T > t) = t^{-\theta\,(1+o(1))}, \qquad \theta = \lambda_0(-z, z), \tag{D.17}$$

the principal Dirichlet eigenvalue of $\frac{1}{2}(\partial_{vv} - v\,\partial_v)$ on $(-z, z)$, and the monotone-envelope brackets of the proof have exactly regularly varying survival tails of that index. Only the asymptotic standardized endpoints $\mp z$ enter: any two staircase families with the same asymptotes have tails of the same index.

Proof. Center at the unit skeleton: the centered count at integer ages, $S_k := (N_k - \lambda k)/\sqrt{\lambda}$, $k \in \mathbb{N}$, is a random walk with i.i.d. mean-zero, unit-variance increments (centered Poisson variables, all moments finite, so $\mathbb{E}[S_1^2 \log^+|S_1|] < \infty$), and by hypothesis the band endpoints are $\lambda a \mp z\sqrt{\lambda a} + o(\sqrt{a})$, so in the centered coordinate the retention region at age k is an interval with endpoints $g_{1,2}(k) = \mp z\sqrt{k} + o(\sqrt{k})$.

Because N is nondecreasing and the ceiling and floor move only upward and only at scheduled ages (by hypothesis), the continuous-age exit is bracketed by integer-age skeleton events: writing u, l for the ceiling and floor,

$$\big\{N_k \in [\,l(k{+}1),\, u(k{-}1)\,]\ \forall k \le n\big\} \ \subseteq\ \{\tau > n\} \ \subseteq\ \big\{N_k \in [\,l(k),\, u(k)\,]\ \forall k \le n\big\}, \tag{D.18}$$

the left event forbidding any crossing at real ages: on $(k, k{+}1]$ the count satisfies $N_s \le N_{k+1} \le u(k) \le u(s)$ against the ceiling and $l(s) \le l(k{+}1) \le N_k \le N_s$ against the floor. (Before the probation onset the events coincide by convention; a finite-age modification and the restriction to integer t shift constants only, survival being monotone in t.) The centered bracketing boundaries carry the staircase-against-drift oscillation and need not be monotone, so bracket once more by monotone envelopes—for the centered ceiling, $\bar{g}_2(k) = \max_{j \le k} g_2(j)$ and $\underline{g}_2(k) = \inf_{j \ge k} g_2(j)$, and symmetrically for the floor—each monotone with the unchanged asymptote $\mp z\sqrt{k}$, the resulting survival events nested by inclusion around those of (D.18). Each outer problem then satisfies hypotheses (H) of Greenwood and Perkins (1983): i.i.d. mean-zero, unit-variance increments with $\mathbb{E}[X^2 \log^+ X] < \infty$; boundaries $f_1 \le 0 \le f_2$ with f_1 eventually non-increasing, f_2 eventually nondecreasing, and $f_i(k)/\sqrt{k} \to c_i$ finite (their case ($\mathrm{H_a}$))—the envelopes are monotone outright, and clipping the finitely many early ages at which the sign condition could fail is a finite-age modification of the kind already noted. Theorem 5(b) there gives $\mathbb{P}(T > n) = n^{-\lambda_0(c_1,c_2)}\pi(n)$ with π slowly varying, and their Proposition 2 identifies $-\lambda_0(c_1, c_2)$ as the largest eigenvalue of $\frac{1}{2}(\partial_{xx} - x\,\partial_x)$ on (c_1, c_2) with Dirichlet conditions at finite endpoints; see Novikov (1982) for one-sided nonlinear boundaries by martingale methods and Denisov et al. (2025) for the surrounding modern literature. Both brackets have $(c_1, c_2) = (-z, z)$, so both survival probabilities are regularly varying with the one index $\lambda_0(-z, z)$; the inner and outer skeleton survivals are squeezed between them, giving (D.17). The one-count overshoot at an upper exit affects the exit position, not the survival event, and does not enter. The hypotheses of case ($\mathrm{H_a}$) consume only the asymptotic standardized endpoints, which proves the final claim. □

Lemma D.5 (Crossing at the detection scale, crash channel)**.** *Let N have rate $\lambda^\star$ and let the held rate be $\hat\lambda \ne \lambda^\star$ with detection scale $a^\star = z_\alpha^2 \hat\lambda/(\lambda^\star - \hat\lambda)^2$. Throughout, the regime is local misspecification—$|\lambda^\star - \hat\lambda| \to 0$, equivalently $a^\star \to \infty$, at fixed $\lambda^\star$—and every $o(1)$ and limit statement is taken in that regime. Write $D(a) := (\lambda^\star - \hat\lambda)\sqrt{a/\hat\lambda}$*

for the standardized drift—so $|D(a)| = z_\alpha\sqrt{a/a^\star}$*—and* $W_a := (N_a - \lambda^\star a)/\sqrt{\hat\lambda a}$ *for the standardized fluctuation, and let* E *be the* no-false-alarm *event that* W *does not exit* $(-z_\alpha, z_\alpha)$ *before the deterministic horizon* $9a^\star$*. Fix* $\varepsilon_0 \in (0, \frac{1}{2})$ *and set* $E' := E \cap \{T > a^{\star\,1-\varepsilon_0}\}$*; all conclusions below are on* E'*, the excluded early false alarms carrying probability tending to* α*. In the understated case* $\hat\lambda < \lambda^\star$*, on* E*:* (i) *the first exit is the* upper *boundary—a floor exit would need* $W_T \le -z_\alpha - D(T) < -z_\alpha$*;* (ii) *the crossing identity*

$$\sqrt{T/a^\star} = \frac{z_\alpha - W_{T-} + \varepsilon_T}{z_\alpha}, \qquad \varepsilon_T = o(1) + O\big((\hat\lambda T)^{-1/2}\big), \tag{D.19}$$

holds, ε_T *collecting the Sterne endpoint error and the one-count overshoot in standardized units, whence,* $W_{T-} \ge -z_\alpha$ *on* E*,*

$$T \;\le\; 4\,a^\star\big(1 + o(1)\big) \qquad \textit{deterministically on } E; \tag{D.20}$$

(iii) $\mathbb{P}\big(T \le a^{\star\,1-\varepsilon} \mid E'\big) \to 0$ *for every* $\varepsilon > 0$*, so* $\log T/\log a^\star \to 1$ *in probability on* E'*. The overstated case* $\hat\lambda > \lambda^\star$ *is the mirror at the floor—a scheduled exit, no overshoot—with the same bound* (D.20)*; there the ceiling is drift-opposed and a* jump-driven *upper exit remains possible on* E*, confined on* E' *to* $T = O(a^{\star\,1/2})$ *and of conditional probability tending to zero, so the first exit is the floor with conditional probability tending to one on* E'*, not pathwise (proof,* Mirror*).*

Proof. First, $T \le 9a^\star$ pathwise on E: were the record still in-band at $a = 9a^\star$, where $D = 3z_\alpha$, the standardized ceiling $z_\alpha + o(1)$ would force $W_a \le -2z_\alpha + o(1) < -z_\alpha$, contradicting E; so E's deterministic window covers $[0, T]$, and every conditioning below is survival of W over that window. *(i).* Before T the drift is nonnegative, so a floor exit, $D(a) + W_a \le -z_\alpha + o(1)$, forces $W_a \le -z_\alpha - D(a) + o(1) < -z_\alpha$ off a null correction, excluded on E. *(ii).* At the upper exit the count equals the ceiling plus the single triggering jump, so in standardized units $D(T)+W_T = z_\alpha+\varepsilon_T$ with ε_T as stated; the triggering jump moves W by $(\hat\lambda T)^{-1/2}$, so $W_T = W_{T-} + O((\hat\lambda T)^{-1/2})$ and $D(T) = z_\alpha - W_{T-} + \varepsilon_T$, which is (D.19); $W_{T-} \ge -z_\alpha$ on E gives $D(T) \le 2z_\alpha + \varepsilon_T$, hence (D.20). *(iii).* A crossing at $T \le a^{\star\,1-\varepsilon}$ forces, by (D.19), $W_{T-} \ge z_\alpha\big(1 - a^{\star\,-\varepsilon/2}\big) - |\varepsilon_T|$—an approach within $\eta := z_\alpha a^{\star\,-\varepsilon/2}(1 + o(1))$ of the upper edge without exiting. The transfer to the diffusion is a strong approximation, not a fixed-horizon limit. The Poisson increments have exponential moments, so the Komlós–Major–Tusnády coupling applies:

on a common probability space, $\sup_{u\le a}|N_u-\lambda^\star u-B(\lambda^\star u)|=O(\log a)$ a.s. (Komlós et al., 1975, 1976; Csörgő and Révész, 1981), so the standardized coupling error is $O(\log a/\sqrt{a})$; the standardized fluctuation W carries variance $\lambda^\star/\hat\lambda=1+o(1)$ in the declared regime, so the coupled Brownian term is $\sqrt{\lambda^\star/\hat\lambda}$ times the standard one, the scaling absorbed into the boundary corrections—both negligible against the layer width η for $\varepsilon<\frac12$, and the event $\{T\le a^{\star\,1-\varepsilon}\}$ shrinks in ε, so that case suffices. In log age the Brownian term is *exactly* the stationary Ornstein–Uhlenbeck ($e^{-\tau'/2}B(e^{\tau'})$, the Lamperti map), and conditioning the killed Ornstein–Uhlenbeck on not exiting is the Doob h-transform by ψ_0—the Q-process, positive recurrent on the bounded interval (Pinsky, 1985; Collet et al., 1995, 2013); the discrete side's conditioned position law is that of Greenwood and Perkins (1983, Theorem 5(a)), density ψ_0 against the Gaussian speed measure, vanishing linearly at the edges. Under the transform an excursion from a fixed interior layer reaches the η-layer with probability $\asymp\eta/\eta_0$ against an order-one unconditioned passage, and positive recurrence makes the expected number of such excursions over the log-horizon $(1-\varepsilon)\log a^\star$ of order $\log a^\star$, so by the union bound $\mathbb{P}(T\le a^{\star\,1-\varepsilon}\mid E)\le C\log(a^\star)\,a^{\star\,-\varepsilon/2}\to 0$. *Mirror.* For $\hat\lambda>\lambda^\star$ the drift is negative; the floor exit is scheduled (no jump, no overshoot), and the identity reads $\sqrt{T/a^\star}=(z_\alpha+W_T+\varepsilon_T)/z_\alpha$ with $W_T\le z_\alpha$ on E, giving (D.20). An *upper* exit on E is not excluded pathwise: the triggering disaster can carry W over the ceiling from $W_{T-}\ge z_\alpha+|D(T)|-(\hat\lambda T)^{-1/2}$, which requires $(\hat\lambda T)^{-1/2}>|D(T)|=z_\alpha\sqrt{T/a^\star}$, hence $T<a^{\star\,1/2}/(z_\alpha\hat\lambda^{1/2})=O(a^{\star\,1/2})$; on E' the residual range is $(a^{\star\,1-\varepsilon_0},O(a^{\star\,1/2})]$, where $\eta\le\hat\lambda^{-1/2}a^{\star\,-(1-\varepsilon_0)/2}\to 0$ uniformly and the edge-layer estimate of (iii) gives conditional probability $o(1)$; the finite-age endpoint error $r(T)=O((\hat\lambda T)^{-1/2})=o(|D(T)|)$ on the same range, since $|D(T)|\ge z_\alpha a^{\star\,-\varepsilon_0/2}$, so exact step-boundary exits are likewise excluded. On E' the first exit is therefore the scheduled floor with conditional probability tending to one, and *(iii)* holds at the floor by the same coupling with the reflected boundary. □

Lemma D.6 (Directional re-selection at the detection scale)**.** *In the setting of Lemma D.5, with a finite menu, the carried odds dominated as in Assumption D.1(ii), and condition* (U) *holding at the exit, condition on the trimmed no-false-alarm event E' of Lemma D.5.* Understated, $\hat\lambda<\lambda^\star$: *the exit is the ceiling, at a disaster, and the just-ended record satisfies*

$$\frac{N_T}{T}=\hat\lambda+z_\alpha\sqrt{\hat\lambda/T}\,\big(1+o(1)\big)\;\ge\;\hat\lambda+\tfrac12(\lambda^\star-\hat\lambda)\big(1-o(1)\big),\tag{D.21}$$

the inequality by (D.20)*. The pairwise likelihood ratio of the record between $\theta > \hat{\lambda}$ and $\hat{\lambda}$ favours θ iff N_T/T exceeds*

$$\frac{\theta - \hat{\lambda}}{\log(\theta/\hat{\lambda})} = \hat{\lambda} + \tfrac{1}{2}(\theta - \hat{\lambda}) + O\big((\theta - \hat{\lambda})^2/\hat{\lambda}\big), \tag{D.22}$$

and favours $\hat{\lambda}$ over every menu rate below it since $N_T/T > \hat{\lambda}$; so the re-selected rate satisfies $\hat{\theta}' > \hat{\lambda}$ strictly on $E' \cap \{T \leq (4-\epsilon)a^\star\}$ for each $\epsilon > 0$. The boundary sliver $\{T/a^\star \in [4-\epsilon,\, 4(1+o(1))]\}$ requires, by (D.19)*, W_{T-} within $O(\epsilon)$ of the drift-opposed edge and carries conditional mass vanishing as $\epsilon \downarrow 0$—the conditioned position density vanishes at the edges (Greenwood and Perkins, 1983, Theorem 5(a)); an exact tie self-re-adopts by the incumbent-retention priority order.* Overstated, $\hat{\lambda} > \lambda^\star$: *the mirror at the floor, scheduled, with $N_T/T \leq \hat{\lambda} - \frac{1}{2}(\hat{\lambda} - \lambda^\star)(1 - o(1))$ and $\hat{\theta}' < \hat{\lambda}$, with conditional probability tending to one on the same event (the jump-driven upper sliver of Lemma D.5 vanishing).*

Proof. The exit side is Lemma D.5(i). At the ceiling $N_T = u_\alpha(T)+1$, and Lemma D.1's leading form gives the equality in (D.21); (D.20) gives $z_\alpha\sqrt{\hat{\lambda}/T} \geq \frac{1}{2}z_\alpha\sqrt{\hat{\lambda}/a^\star}\,(1 - o(1)) = \frac{1}{2}(\lambda^\star - \hat{\lambda})(1 - o(1))$, the inequality. For (D.22), $\log \Lambda_T^{\theta/\hat{\lambda}} = N_T \log(\theta/\hat{\lambda}) - T(\theta - \hat{\lambda})$ is positive iff N_T/T exceeds the displayed ratio, whose expansion follows from $\log(\theta/\hat{\lambda}) = \frac{\theta-\hat{\lambda}}{\hat{\lambda}}\big(1 - \frac{\theta-\hat{\lambda}}{2\hat{\lambda}} + O((\theta-\hat{\lambda})^2/\hat{\lambda}^2)\big)$ inverted. On $E \cap \{T \leq (4-\epsilon)a^\star\}$, $N_T/T - \hat{\lambda} \geq \frac{z_\alpha\sqrt{\hat{\lambda}/a^\star}}{2\sqrt{1-\epsilon/4}}\,(1 - o(1))$, which exceeds $\frac{1}{2}(\lambda^\star - \hat{\lambda}) + O((\lambda^\star - \hat{\lambda})^2/\hat{\lambda})$ once the misfit is small, so the record strictly favours over $\hat{\lambda}$ every menu rate up to $\hat{\lambda} + 2(N_T/T - \hat{\lambda})(1 + o(1))$ and favours $\hat{\lambda}$ over all rates below it; under (U) such a higher rate exists, the dominated carried odds cannot overturn the strict pairwise gaps (Assumption D.1(ii)), and the argmax exceeds $\hat{\lambda}$. The boundary-sliver mass and the mirror follow as stated. □

Lemma D.7 (Selection robustness)**.** *Let $\{A'_\alpha(a)\}$ be any nested family of acceptance intervals, of held-model mass exactly* $1-\alpha$ *in the continuous channel and at least* $1-\alpha$ *in the counting channel, whose standardized endpoints converge to $\mp z_\alpha$, the counting endpoints nondecreasing integer staircases. Then Theorem 2(a) holds at the level of the tail* exponent*, $\mathbb{P}(\tau > t) = t^{-\theta(\alpha)(1+o(1))}$: the trend problem has the same principal eigenvalue—the exact power law $\sim C\,t^{-\theta(\alpha)}$ requiring endpoints exactly $\mp z_\alpha$, merely-convergent endpoints controlling the exponent but not the slowly varying* *factor—and the crash tail is* $t^{-\theta(\alpha)(1+o(1))}$ *by Lemma D.4, whose hypotheses consume only the asymptotic endpoints. The finite/infinite-mean threshold at* $\theta(\alpha) = 1$ *and the recurrence of Proposition 2 are selection-invariant; the calendar of scheduled steps, the predictable-atom dates, and the finite-age geometry of the band are selection-specific. In the symmetric limits of both channels the natural selections—highest-density, equal-tailed, central—share the endpoints* $\mp z_\alpha$.

Proof. In the trend channel the standardized band lies, from some age on, between the fixed intervals $(-z_\alpha \mp \delta,\, z_\alpha \pm \delta)$ for every $\delta > 0$; domain monotonicity of the killed problem sandwiches the survival between the two fixed-band problems, whose exponents converge to $\theta(\alpha)$ by the continuity of λ_0 (Greenwood and Perkins, 1983, Proposition 2), so the exponent is unchanged, and with endpoints exactly $\mp z_\alpha$ the spectral representation (D.23) is unchanged including the constant. In the crash channel the hypotheses of Lemma D.4 are the endpoint asymptotics and the monotone staircase form, both assumed, so (D.17) gives the tail, and the mean threshold follows from the lemma's brackets as in Theorem 2(a). Recurrence in Proposition 2 uses only $\tau < \infty$ a.s. and the restart, both selection-free. The scheduled dates are the integer-attainment ages of the specific staircase and move with the selection. $\square$

D.6 Heavy tails and the detection scale

Proof of Theorem 2(a). By (D.6), holding the correct trend model, $\tau = e^H$ with $H = \inf\{\tau' > \log a_0 : |V_{\tau'}| > z_\alpha\}$ the first exit of the Ornstein–Uhlenbeck process V from $(-z_\alpha, z_\alpha)$. Monitoring begins at $\tau' = \log a_0$, so the killed-process duration is $D := H - \log a_0$, with ν_0 the law of V at the onset $\log a_0$ *restricted to the interior* $(-z_\alpha, z_\alpha)$. Under the correct model $V_{\log a_0}$ is standard normal, so the onset places mass $\alpha = \mathbb{P}(|V| > z_\alpha)$ already outside the band—those paths exit at once, $D = 0$—and only the interior sub-probability mass of total $1 - \alpha$ feeds the positive-survival expansion below (equivalently, the Dirichlet eigenfunctions, extended by zero outside the band, integrate against ν_0 over the interior alone); the atom at $D = 0$ does not affect the $s \to \infty$ tail. The generator $\mathcal{L} = \frac{1}{2}\partial_{vv} - \frac{1}{2}v\,\partial_v$ on $(-z_\alpha, z_\alpha)$ with Dirichlet boundary conditions is self-adjoint with respect to the Gaussian speed (invariant) measure $m(dv) \propto e^{-v^2/2}\,dv$, not Lebesgue measure—a regular Sturm–Liouville operator with discrete spectrum $0 < \theta_1(\alpha) < \theta_2(\alpha) < \cdots$ and a positive principal eigenfunction ψ_1, the eigenfunctions $\{\psi_k\}$ orthonormal in $L^2(m)$. The survival of the killed process admits the eigenvalue (spectral) expansion for diffusion hitting times (Kent, 1980; Alili et al., 2005)

$$\mathbb{P}(D > s) = \sum_{k\geq 1} c_k\, e^{-\theta_k(\alpha)s}, \qquad c_1 = \Big(\int \psi_1\, dm\Big)\Big(\int \psi_1\, d\nu_0\Big) > 0, \tag{D.23}$$

the duration measured from the onset, so $\mathbb{P}(D > s) \sim c_1 e^{-\theta_1(\alpha)s}$ as $s \to \infty$. Therefore, using $H = \log a_0 + D$,

$$\mathbb{P}(\tau > t) = \mathbb{P}\big(e^H > t\big) = \mathbb{P}\big(D > \log(t/a_0)\big) \sim c_1\, e^{-\theta_1(\alpha)\log(t/a_0)} = \big(c_1 a_0^{\theta_1(\alpha)}\big)\, t^{-\theta_1(\alpha)}, \tag{D.24}$$

which is (10) with $\theta(\alpha) = \theta_1(\alpha)$ and constant $C = c_1 a_0^{\theta_1(\alpha)}$ depending on the onset a_0 but not the exponent. The mean is

$$\mathbb{E}[\tau] = \mathbb{E}\big[e^H\big] = a_0 \int_0^\infty e^s \left(-d\mathbb{P}(D > s)\right) < \infty \iff \theta_1(\alpha) > 1, \tag{D.25}$$

the integral converging at $s \to \infty$ exactly when the decay rate $\theta_1(\alpha)$ in (D.23) exceeds 1.

In the crash channel, holding the correct model the count has rate $\lambda^\star$, and by Lemma D.1 the Sterne staircases are nondecreasing with endpoints $\lambda^\star a \mp z_\alpha\sqrt{\lambda^\star a} + o(\sqrt{a})$—the integer-versus-continuous gap is exactly the $o(\sqrt{a})$ the hypotheses of Lemma D.4 absorb—so the lemma applies with $(\lambda, z) = (\lambda^\star, z_\alpha)$ and (D.17) is (11), the index the same principal Dirichlet eigenvalue as the trend channel's, the operator being the same. The finite-mean threshold follows from the lemma's exactly regularly varying brackets: $\mathbb{E}[\tau] < \infty$ when $\theta(\alpha) > 1$ (the upper bracket integrable) and $\mathbb{E}[\tau] = \infty$ when $\theta(\alpha) < 1$ (the lower bracket not). The jump structure and the predictable-floor/inaccessible-ceiling split of the exit (Lemma D.1) enter the exact finite-age law (D.12) but wash out of the index.

For an incorrect held model the statistic carries a drift (Theorem 2(b)), so this heavy tail is a property of holding the correct model. $\square$

Proof of Theorem 2(b). Trend. With $\hat\mu \neq \mu^\star$, $R_a = \sqrt{a}\,(\mu^\star - \hat\mu)/\sigma + W_a/\sqrt{a}$; the deterministic term reaches the band edge z_α when $\sqrt{a}\,|\mu^\star - \hat\mu|/\sigma = z_\alpha$, i.e. at $a = z_\alpha^2\sigma^2/(\mu^\star - \hat\mu)^2$, the bounded-in-probability fluctuation $W_a/\sqrt{a} = O_P(1)$ being of the same order as the band half-width z_α—so $a^\star$ is the detection *scale*, not a sharp threshold, and the concentration is the conditional statement of Theorem 2(b). *Crash.* With $\hat\lambda \neq \lambda^\star$, $N_a - \hat\lambda a = (\lambda^\star - \hat\lambda)a + M_a$ reaches the band half-width $z_\alpha\sqrt{\hat\lambda a}$ when $|\lambda^\star - \hat\lambda|\,a = z_\alpha\sqrt{\hat\lambda a}$, i.e. at $a = z_\alpha^2\hat\lambda/(\lambda^\star - \hat\lambda)^2$. In both the deterministic drift reaches the band at the $a^\star$ of (12): in the trend channel the standardized R_a has an $O(1)$ band and $O(1)$ fluctuation against a drift $\sqrt{a}\,(\mu^\star - \hat\mu)/\sigma$ that grows like $\sqrt{a}$, while in the crash channel the unstandardized $N_a - \hat\lambda a$ has drift $(\lambda^\star - \hat\lambda)a$ against an $O(\sqrt{a})$ martingale and band. The crossing is *shallow*, so $a^\star$ pins only the *scale*: in logarithmic age $\tau' = \log a$ the trend statistic is $\delta e^{\tau'/2} + V_{\tau'}$ with $\delta = (\mu^\star - \hat\mu)/\sigma$—for $\mu^\star < \hat\mu$ reflect the record, $X \mapsto -X$, the band being symmetric, so take $\delta > 0$ without loss—and V the stationary Ornstein–Uhlenbeck process of (D.6) (order one); the deterministic part reaches z_α at $\tau'^\star = \log a^\star$ with slope $z_\alpha/2$. *Conditional on no earlier false alarm*—the event F_δ that the stationary fluctuation V does not exit the fixed band $(-z_\alpha, z_\alpha)$ before the deterministic horizon $\bar s := \tau'^\star + 2\log 2$, at which $\delta e^{\bar s/2} = 2z_\alpha$—the drifted statistic exceeds $-z_\alpha$ pointwise before $\bar s$, so has no lower exit there, and equals $2z_\alpha + V_{\bar s} > z_\alpha$ at $\bar s$, so crosses upward by then: the statistic equals z_α at the upper crossing $H \le \bar s$, i.e. $\delta e^{H/2} + V_H = z_\alpha$ with $V_H \in (-z_\alpha, z_\alpha)$ on F_δ (V has not exited by $\bar s \ge H$), and the relation is exact, requiring no expansion and no tightness of V at the random time H:

$$e^{(H-\tau'^\star)/2} = \frac{z_\alpha - V_H}{z_\alpha} \in (0,2) \qquad \Longrightarrow \qquad H \le \tau'^\star + 2\log 2 \quad \text{deterministically on the event.} \tag{D.26}$$

For the matching lower bound, a crossing at $H \le (1-\varepsilon)\tau'^\star$ forces, by the same exact relation, $V_H \ge z_\alpha(1 - e^{-\varepsilon\tau'^\star/2})$—an approach within $\eta_\varepsilon := z_\alpha e^{-\varepsilon\tau'^\star/2}$ of the absorbing edge. Under the conditioned law the edge is entrance-repelling: conditioning on survival is the Doob h-transform by the principal eigenfunction ψ_0—the positive recurrent Q-process of the killed Ornstein–Uhlenbeck on a bounded interval (Pinsky, 1985; Collet et al., 1995)—under which a passage from the fixed interior layer $z_\alpha - \eta_0$ to the η-layer carries the factor $\psi_0(z_\alpha - \eta)/\psi_0(z_\alpha - \eta_0) \asymp \eta/\eta_0$ against an order-one unconditioned passage (ψ_0 vanishes linearly at the edge), and the ψ_0-conditioned diffusion, being positive recurrent, makes an *expected* $O(\tau'^\star)$ approaches to the fixed layer over the horizon, so by the union bound

$$\mathbb{P}\big(H \le (1-\varepsilon)\tau'^\star \,\big|\, F_\delta\big) \;\le\; C\,\tau'^\star e^{-\varepsilon\tau'^\star/2} \;\longrightarrow\; 0 \qquad \text{for every } \varepsilon > 0.$$

In the crash channel the two bounds are Lemma D.5(ii)–(iii): the crossing identity (D.19) replaces the exact relation, the one-count overshoot enters only through ε_T, and the Komlós–Major–Tusnády strong approximation enters only the lower side. Hence, on the conditioning event, $H/\tau'^\star \to 1$ in probability, so the holding time concentrates on the *log* scale, $\log\tau/\log a^\star \to 1$ (equivalently $\tau = a^{\star\,1+o_{\mathbb{P}}(1)}$) as the misfit $\to 0$. The multiplicative ratio $\tau/a^\star$ is not claimed to converge: the upper side is the deterministic $\tau/a^\star = e^{H-\tau'^\star} \le 4$, the lower side only sub-polynomial. The limit is conditional on F_δ, and $\mathbb{P}(F_\delta)$ *vanishes* as the misfit $\to 0$—over the growing log-age interval before the deterministic crossing, the stationary fluctuation exits the fixed band with probability tending to one—so the *unconditional* holding time is governed by the fixed-band false-alarm time of the stationary fluctuation—an a.s. finite random time, tight relative to the diverging $a^\star$ but not concentrating around any deterministic shorter scale. Thus $\log\tau/\log a^\star \to 1$ is the no-false-alarm statement, and $a^\star$ is the conditional detection scale. For the matched comparison, work in the local (small-misfit) regime, where $\mathrm{KL}(\mathbb{P}^\star\|\mathbb{P}^{\hat\lambda}) \approx (\lambda^\star - \hat\lambda)^2/(2\lambda^\star)$ (Poisson) and likewise in the trend channel; equal KL then fixes $|\lambda^\star - \hat\lambda|$ to leading order, and the band coefficient orders $a^\star$ only at the next order. Away from this regime equal KL no longer pins the misfit term of (12), and the comparison is stated only locally. $\square$

D.7 The two deaths

The general theory of this appendix is channel-free. The crash channel adds one sharp, channel-specific feature, the *two deaths* flagged in §3.4: the *direction* of a model's error fixes the *manner* in which a long-enough record refutes it. We state the result and its scope, then prove it.

Assumption D.1 (Local misspecification / long record)**.** *The crash-channel* reselection-direction *result—Proposition D.1—is stated for the* local-misspecification *regime: $\hat\lambda$ is close enough to $\lambda^\star$ that the detection scale $a^\star = z_\alpha^2\hat\lambda/(\lambda^\star - \hat\lambda)^2$ (Theorem 2(b)) is large; the two consequences below are statements about exits* at that scale*—the conditional regime of Theorem 2(b)—while fluctuation-driven exits at short records occur and are outside them. Two consequences are used.* (i) Understated rate, $\hat\lambda < \lambda^\star$: *at an upper exit the count sits on the ceiling, $u_\alpha(\tau) = \hat\lambda\tau + z_\alpha\sqrt{\hat\lambda\tau} + o(\sqrt{\tau})$, so $N_\tau/\tau = \hat\lambda + z_\alpha\sqrt{\hat\lambda/\tau} + o(1/\sqrt{\tau})$; on the no-false-alarm event $\tau \le 4a^\star(1+o(1))$ (Lemma D.5), so $N_\tau/\tau \ge \hat\lambda + \frac{1}{2}(\lambda^\star - \hat\lambda)(1 - o(1))$, equal to $\lambda^\star$ to leading order* *at* $\tau/a^\star \to 1$ *(*$z_\alpha\sqrt{\hat\lambda/a^\star} = \lambda^\star - \hat\lambda$*); the post-jump count exceeds the ceiling by the single triggering disaster, a further* $o(1)$ *correction.* Overstated rate, $\hat\lambda > \lambda^\star$: *the detection-side exit is the* floor, $l_\alpha(\tau) = \hat\lambda\tau - z_\alpha\sqrt{\hat\lambda\tau} + o(\sqrt{\tau})$, *with the opposite sign,* $N_\tau/\tau = \hat\lambda - z_\alpha\sqrt{\hat\lambda/\tau} + o(1/\sqrt{\tau}) \le \hat\lambda - \frac{1}{2}(\hat\lambda - \lambda^\star)(1 - o(1))$, *and* $z_\alpha\sqrt{\hat\lambda/a^\star} = \hat\lambda - \lambda^\star$. (ii) *The re-selection (8) is driven by the record rather than by* ρ *when the carried log-odds at adoption are dominated by the record's pairwise separations,* $\max_{\theta\neq\theta'}|\log(\rho(\theta)/\rho(\theta'))| = o(\Delta_\tau)$ *with* Δ_τ *the smallest pairwise log-likelihood-ratio gap the just-ended record generates among the menu rates it ranks (a uniform sufficient condition: the selection turns only on the gaps to contenders that can overtake the record-favoured model)—the operative scale for a* local-alternatives *menu, whose pairwise gaps shrink relative to* τ*; for a fixed menu of separated rates the gaps are of order* τ *and the criterion is correspondingly weaker, as when the prior history is short or the carried ranking near-uniform—and* not *after consolidation, where the carried ranking already favours* θ° *by a margin growing with the accumulated history, the re-selection is* ρ*-driven and self-re-adopts* θ° *(condition* (U) *fails); the directional map below, a statement about* detectably-misspecified *held rates, then does not apply, and we restrict it to the* $o(\tau)$ *case. For* gross *misspecification the record is short, the carried prior can decide the re-selected rate, and the map is not guaranteed. The* exit-side *and* timing *statements hold conditional on no earlier false alarm (Lemma D.5) and need neither consequence (i) nor (ii); only the re-selection* direction *additionally uses them.*

One condition governs whether an exit's *direction* translates into a strictly different re-selected rate. Call the held rate *detectably misspecified* on its just-ended record when it is not the ranking-weighted front-runner of that record,

$$\hat\lambda \;\neq\; \arg\max_{\theta\in\Theta} L_\theta\big(X_{(\sigma,\sigma+\tau]}\big)\,\rho(\theta). \tag{U}$$

Condition (U) holds whenever the model is detectably misspecified—an *environment-specific* property of the held rate against its record and the menu, not a behavioral parameter—and fails ($\hat\lambda' = \hat\lambda$, self-re-adoption) when the held rate remains the record's MAP front-runner: not only when it equals the empirical rate, but whenever the empirical rate falls in its likelihood cell or the carried ranking keeps it on top, as for a consolidated or correct model, whose record's empirical rate $\to \lambda^\star$ lies in $\hat\lambda$'s likelihood cell—$\hat\lambda$ the menu rate nearest $\lambda^\star$ in Kullback–Leibler, equal to $\lambda^\star$ when the truth is on the menu—and favours $\hat{\lambda}$ exponentially over any other menu rate.

Proposition D.1 (Two deaths, disputed-crash-rate channel)**.** *In the disputed-crash-rate channel,* conditional on the record surviving long enough to detect the misfit*—to an age of order the detection scale $a^\star$ of Theorem 2(b), without an earlier exit—the* exit side *is fixed by the error direction—with conditional probability tending to one in the local-misspecification limit, on the no-earlier-false-alarm event of Lemma D.5—and the* re-selection direction *by the record's empirical rate. A held model that understates the disaster rate, $\hat{\lambda} < \lambda^\star$, is then carried to its ceiling by the rising count and dies by an* upper *exit—a disaster carrying the count above the ceiling, a totally inaccessible stopping time of the count's filtration—and re-selects* (8) *a rate weakly above $\hat{\lambda}$,* strictly *above under* (U)*—the exit record's empirical rate sitting at least halfway from $\hat{\lambda}$ toward $\lambda^\star$ (Lemma D.6; Assumption D.1). A held model that overstates it, $\hat{\lambda} > \lambda^\star$, is carried below its floor and dies by a* lower *exit—the scheduled step-up of the floor in a quiet spell, a predictable stopping time—and re-selects a rate weakly below $\hat{\lambda}$,* strictly *below under* (U)*—at least halfway down toward $\lambda^\star$ (Lemma D.6). The conditioning is essential, not cosmetic:* unconditionally*, before the misfit is detected, an early chance fluctuation can trigger either exit, and Theorem 2(b) shows these false alarms* dominate *the holding time as the misfit shrinks—so the directional map is the conditional-on-detection statement, not an unconditional claim that a misspecified model usually dies on its drift-aligned side.*

Proof of Proposition D.1. Decompose the count relative to the held mean,

$$N_a - \hat{\lambda}a = (\lambda^\star - \hat{\lambda})\, a + M_a, \qquad M_a := N_a - \lambda^\star a, \tag{D.27}$$

with M_a a mean-zero martingale, $M_a = O_{\mathbb{P}}(\sqrt{a})$ and $M_a = o(a)$ a.s. by (D.10). By Lemma D.1 the ceiling and floor sit at $\pm z_\alpha\sqrt{\hat{\lambda}a} = O(\sqrt{a})$ about the held mean.

Understatement, $\hat{\lambda} < \lambda^\star$. The drift $(\lambda^\star - \hat{\lambda})a > 0$ in (D.27) grows linearly and dominates both $M_a = o(a)$ and the $O(\sqrt{a})$ ceiling, so for all large a, $N_a - \hat{\lambda}a$ exceeds the ceiling: the count exceeds the ceiling at a finite age almost surely, so the ceiling is the boundary the drift reaches. Whether this *upper* exit is the *first* exit is settled below; the displayed crossing shows only that the ceiling is reached a.s. eventually. An upper exit occurs at a jump of N, i.e. at a disaster; the jump times of a Poisson process are totally inaccessible stopping times of its natural filtration, admitting no announcing sequence (Dellacherie and Meyer, 1982). On re-selection (8) after a conditional upper exit the just-ended record has empirical rate at least halfway from $\hat{\lambda}$ to $\lambda^\star$ ((D.21), via the deterministic crossing bound (D.20); Assumption D.1), so its likelihood favours a *strictly higher* menu rate when condition (U) holds, while a $\hat{\lambda}$ that remains the menu rate nearest the record self-re-adopts ((U) fails), $\hat{\lambda}' = \hat{\lambda}$. The drift-opposite event is a lower exit—the rising floor catching a stalled count. Since N_a is non-decreasing while the floor $\hat{\lambda}a - z_\alpha\sqrt{\hat{\lambda}a}$ rises only at rate $\hat{\lambda} < \lambda^\star$, the gap

$$N_{a'} - \big(\hat{\lambda}a' - z_\alpha\sqrt{\hat{\lambda}a'}\big) = (\lambda^\star - \hat{\lambda})a' + M_{a'} + z_\alpha\sqrt{\hat{\lambda}a'} = (\lambda^\star - \hat{\lambda})a' + o(a') \ \to\ +\infty \quad \text{a.s.}$$

grows: the count outpaces the floor, and by (D.10) it also exceeds the ceiling at a finite age a.s. A lower exit can nonetheless come *first*: the floor catches the count whenever an early downward fluctuation of M of order the band half-width $z_\alpha\sqrt{\hat{\lambda}a}$ arrives before the drift has separated the count from the floor, an event of positive probability that does not vanish with age. Which exit is first is thus a competition between the drift-driven ceiling crossing (near $a^\star$) and an early lower false alarm, and we do *not* claim the upper exit wins unconditionally. On the contrary, Theorem 2(b) shows that in the local-misspecification regime ($a^\star$ large) the fixed-band false-alarm time *dominates* the detection scale, so unconditionally the first exit is typically a false alarm, which can fall on either band. What survives is the *conditional* statement, and it is Lemma D.6: conditional on *no earlier false alarm*—the same regime as Theorem 2(b)—the first exit is the ceiling, pathwise on the conditioning event (a floor exit would require the fluctuation itself to cross the band, Lemma D.5(i)); it arrives by (D.20) no later than $4a^\star(1 + o(1))$; and the just-ended record's empirical rate (D.21) sits at least halfway from $\hat{\lambda}$ to $\lambda^\star$, so the likelihood re-selects a strictly higher menu rate under (U) and self-re-adopts otherwise. No survival-conditioned concentration estimate enters: the conditioning excludes the drift-opposed boundary pathwise, and the re-selection direction follows from the deterministic crossing bound alone.

Overstatement, $\hat{\lambda} > \lambda^\star$. Symmetrically the drift $(\lambda^\star - \hat{\lambda})a < 0$ carries the count behind the band centre, the floor gap $N_{a'} - (\hat{\lambda}a' - z_\alpha\sqrt{\hat{\lambda}a'}) = -(\hat{\lambda} - \lambda^\star)a' + o(a') \to -\infty$ a.s.; the unconstrained count falls behind the floor at a finite age almost surely—an eventual, counterfactual crossing, not yet the spell's first exit; when the lower exit *is* first, it occurs at a scheduled band-step date t_j where the floor steps up past the count. That date is deterministic in the regime's age clock; the corresponding calendar date $\sigma + t_j$ is a predictable time, σ being $\mathcal{F}_\sigma$-measurable and the date announced by $\sigma + t_j - 1/n$, and the event $\{N_{t_j^-} < l_{j+1}\}$ is $\mathcal{F}_{(\sigma+t_j)^-}$-measurable, so the lower exit is a predictable stopping time in the calendar filtration. Re-selection then returns, under (U), a rate *strictly lower* than $\hat{\lambda}$—the exit record's empirical rate at least halfway down toward $\lambda^\star$ (Lemma D.6, *Overstated*)—and self-re-adopts otherwise. Symmetrically the drift-opposite event is an upper exit—a disaster surge carrying the lagging count over the ceiling before the floor catches it—of positive, non-vanishing probability, so again the first exit is not unconditionally the lower one; conditional on no earlier false alarm the mirror of Lemma D.6 applies: the exit is the floor—an upper exit would require the fluctuation itself to cross the band, up to the *Mirror*'s vanishing jump-driven sliver—scheduled, arriving by (D.20) no later than $4a^\star(1 + o(1))$, with exit empirical rate at least halfway from $\hat{\lambda}$ down to $\lambda^\star$, so the likelihood reselects a strictly *lower* menu rate under (U); the next boundary touched is the floor, with conditional probability tending to one on the conditioning event (Lemma D.5, *Mirror*). The error-direction map—understate $\Rightarrow$ upper (inaccessible) exit $\Rightarrow$ revise toward fear, by surprise; overstate $\Rightarrow$ lower (predictable) exit $\Rightarrow$ revise toward calm, on schedule—is therefore the *conditional-on-detection* statement: it names the boundary the misfit-driven count reaches once a long enough record has detected the error, not an unconditional claim about which exit comes first, since before detection a chance false alarm can refute either type on either side (Theorem 2(b)). $\square$

E Proofs for Section 4

This appendix proves the results of Section 4. Throughout, the common path of X is fixed, $\Psi_t : \Xi \to \mathcal{S}$ is the flow it induces on belief states, and $m_t = \Psi_t \# G$.

Proof of Proposition 3. For fixed path, the band half-width $z_{\alpha(\xi)}$ depends measurably on ξ, and the successive exit times are first passage times of the common record against the ξ-dependent band, each a measurable functional of (ξ, path); hence Ψ_t is G-measurable and $m_t = \Psi_t \# G$ is a well-defined element of $\mathcal{P}(\mathcal{S})$. The identity $\langle \varphi, m_t \rangle = \int_\Xi \varphi(\Psi_t(\xi))\, G(d\xi)$ is the change-of-variables formula for a pushforward and uses no averaging over independent agent-level noise: the only randomness is the common path, held fixed. For the realized-versus-ex-ante distinction, fix an age a and let $\mathbf{1}^a_{\text{surv}}(\xi)$ indicate that type ξ's current holding spell has survived from adoption to age a—the event whose probability (C.2) computes—so that, conditional on the spell's initialization—adoption date, held model, carried ranking, the state from which (C.2) is computed—its expectation is the single-type survival law from that state, $\mathbb{E}[\mathbf{1}^a_{\text{surv}}(\xi)] = S_\xi(a)$: the indicator is the pathwise object—survival along the realized path from adoption, carrying the adoption date through the path segment it spans—while $S_\xi(a)$ averages over the record paths the initialization admits; the unconditional mixture is adoption-aligned, over G and the initialization law, not a calendar-time snapshot. The realized population survivor fraction to age a, $\int_\Xi \mathbf{1}^a_{\text{surv}}(\xi)\, G(d\xi)$, is a functional of the path, and by Fubini $\mathbb{E}\int_\Xi \mathbf{1}^a_{\text{surv}}(\xi)\, G(d\xi) = \int_\Xi \mathbb{E}[\mathbf{1}^a_{\text{surv}}(\xi)]\, G(d\xi) = \int_\Xi S_\xi(a)\, G(d\xi)$, the G-mixture of survival laws—so the realized fraction equals the ex-ante mixture only in expectation, and neither equals the single-type $S(a)$. This survivor-to-age object is distinct from the *instantaneous* non-revision fraction that enters m_t in the statement—the G-mass of types not revising in the instant at t, a separate pathwise functional with its own mean—which is precisely why the cross-section $m_t = \Psi_t \# G$ is an instantaneous pushforward of G and not a survival mixture. □

Proof of Corollary 3. We derive (14) as a direct pathwise balance, not from an extended generator. Fix the realized common path. Each type ξ then has a deterministic trajectory: along the common path the record $R_t = T_{a_t}$ is a deterministic function of ξ and t—the same data windowed from ξ's adoption date—the held model $\hat{\theta}$ is constant between revisions, and the age rises at unit rate, so $\Psi_t(\xi) = (\hat{\theta}_t(\xi), \alpha(\xi), a_t(\xi))$ is determined—$\dot{\alpha} = 0$, the tolerance a frozen coordinate—and $m_t = \Psi_t \# G$ (Proposition 3). The record's between-revision motion is a diffusion in the trend channel and piecewise-constant in the crash channel; conditioning on the common path makes it a fixed sample path either way, so no autonomous deterministic flow—and no Davis piecewise-deterministic structure—is invoked or needed. A revision is the first exit of that record from the band $A_\alpha(a)$ of (4): the first time the trajectory $R_t(\xi)$ leaves the ordinary range, at which instant $\Psi_t(\xi)$ jumps by the re-selection-and-restart map J of (8) (a new held model, the record reset to age 0, testing resuming at the probation onset a_0). No transversality of the crossing is required: the first-exit time is well defined for each type along the path, the diffusion's local recrossings of the boundary notwithstanding, since the record resets at the first exit and the old band is no longer monitored.

Differentiate $\langle \varphi, m_t \rangle = \int_\Xi \varphi(\Psi_t(\xi))\, G(d\xi)$ in t along the path. Between its own exits $\frac{d}{dt}\varphi(\Psi_t(\xi)) = \partial_a\varphi(\Psi_t(\xi))$, since $\hat{\theta}$ is fixed and $\dot{a} = 1$; at an exit φ jumps by $(\varphi \circ J - \varphi)(\Psi_{t^-}(\xi))$. Integrate over ξ (Fubini; φ bounded, G a probability measure) and collect the exits into the *crossing flux* Φ_t—the measure on pre-jump states that accumulates, by time t, the G-mass of types whose record has first reached the band edge, marked by where it crosses—so that the transport term is $\langle \partial_a\varphi, m_t\rangle\, dt$ and the boundary-crossing term is the Stieltjes integral $\int(\varphi \circ J - \varphi)\, d\Phi_t$ against Φ_t, which is locally finite—its cumulative mass on $[0, t]$ is the G-average number of crossings by t, finite on compacts by a *channel-specific* argument rather than a single floor. In the trend channel the deterministic probation floor $\tau_k \geq a_0 > 0$ (Lemma A.6) caps each type's revision count on $[0, t]$ at t/a_0, so the G-average is finite. In the crash channel, where $a_0 = 0$ and no deterministic floor holds, crossings occur only at disasters and at scheduled floor-steps: disasters arrive as a Poisson process of rate $\lambda^\star$, a.s. finitely many on $[0, t]$, and the scheduled exits are uniformly spaced from below across the continuum of types: within any spell the Sterne floor first reaches 1 only at an age $a_1(\hat{\lambda}, \alpha) > 0$ that grows as the band widens and shrinks as the held rate rises, so $a_1 \geq a_1(\lambda_K, \bar{\alpha}) > 0$ uniformly, with λ_K the largest menu rate and $\bar{\alpha} := \sup\{\alpha(\xi) : \xi \in \operatorname{supp} G\}$, below 1 by the non-degeneracy clause of Assumption 2. A spell ends at a disaster or at a scheduled exit, and any scheduled exit falls at an age of at least $a_1(\lambda_K, \bar{\alpha})$; so between consecutive disasters, a gap of length g holds at most $g/a_1(\lambda_K, \bar{\alpha}) + 1$ spells, and every type's crossing count on $[0, t]$ obeys the single integrable bound $2N_t + t/a_1(\lambda_K, \bar{\alpha}) + 1$, the G-average finite on compacts—local finiteness inherited from the driving Poisson process and the uniform first-step floor, not from a holding-time floor. In neither channel is the count bounded by 1: under recurrence it grows with the population-average number of revisions. The hazard (9) is the *ex-ante* rate of these boundary hits (Appendix C), the expectation over paths of the flux, not a pathwise intensity; Φ_t itself is the realized, pathwise crossing flux, consistent with the pathwise m_t and not with $\mathbb{E}[m_t]$ for independently-revising agents. Differentiating the fixed-population integral $\int_\Xi \varphi(\Psi_t(\xi))\, G(d\xi)$ closes with these two terms *alone*—transport and flux—because G is a fixed probability measure over a closed population: no mass enters or leaves and every type carries equal weight. The remaining two channels are not produced by this differentiation; each requires changing what is differentiated. Turnover enters once the population is *not* closed: it is a genuine rate-p birth–death jump appended to the balance—at rate p a G-uniform fraction is replaced and reinjected at the type's newborn state $s_\circ(\xi)$—

each agent reborn with her own tolerance, the newborn cross-section the pushforward $m_\circ = s_\circ \# G$—contributing $p(\langle \varphi, m_\circ\rangle - \langle \varphi, m_t\rangle)\, dt$. Selection enters once the weighting is *not* head-count. Wealth is *type-attached*, not a state field: each type ξ carries a relative wealth share $\omega_t(\xi)$ with $d\omega_t(\xi)/\omega_t(\xi) = r_t(\xi)\, dt$ between common disasters, $r_t(\xi)$ the relative return net of the wealth-weighted average, and a revision moves the type's state, not its wealth—the crossers of Φ_t^w carry their shares through J unchanged. Since $s = (\hat{\theta}, \alpha, a)$ omits the record and the type history, ω_t and r_t are not functions of s; the state-space objects are the fibre averages over $\{\Psi_t = s\}$: $\bar{\omega}_t(s)$, the G-fibre average of ω_t—equivalently the Radon–Nikodym derivative of the weighted pushforward $\Psi_t \#(\omega_t G)$ against m_t—and $\bar{r}_t^w(s)$, the wealth-weighted fibre average of r_t. With the wealth-weighted type measure $w_t(d\xi) := \omega_t(\xi)\, G(d\xi)/\langle \omega_t, G\rangle$ and $m_t^w := \Psi_t \# w_t$, disintegration over Ψ_t gives the exact tilt $m_t^w(ds) = \bar{\omega}_t(s)\, m_t(ds)/\langle \bar{\omega}_t, m_t\rangle$. Differentiating $\langle \varphi, m_t^w\rangle = \langle (\varphi \circ \Psi_t)\, \omega_t, G\rangle / \langle \omega_t, G\rangle$ by the quotient and product rules on the type space, the weight motion contributes

$$d\Sigma_t(\varphi) = \big(\langle \varphi\, \bar{r}_t^w\rangle_{m_t^w} - \langle \varphi\rangle_{m_t^w} \langle \bar{r}_t^w\rangle_{m_t^w}\big) dt = \operatorname{Cov}_{m_t^w}(\varphi,\, \bar{r}_t^w)\, dt, \tag{E.1}$$

the replicator covariance expressed in share space; the projection from the type space is exact—$\varphi \circ \Psi_t$ is constant on fibres, so the tower property gives $\operatorname{Cov}_{w_t}(\varphi \circ \Psi_t,\, r_t) = \operatorname{Cov}_{m_t^w}(\varphi,\, \bar{r}_t^w)$—and it is identically zero under the head-count weighting taken as the baseline here. Collecting: the head-count average carries the first three contributions, which is (14); and between common disasters, where the weight motion is continuous, the quotient rule above returns the same integrals under the tilt and adds $d\Sigma_t$, which is (15)—at a common disaster weights and beliefs jump together, and the joint atom is the (17) object, outside these displays. The terms of (14): the ageing drift $\langle \partial_a \varphi, m_t\rangle\, dt$ (transport); the boundary-crossing contribution $\int (\varphi \circ J - \varphi)\, d\Phi_t$ (flux; J the re-adoption map, which resets the *age* coordinate to 0—the probation onset a_0 is the distinct object gating when testing resumes on the fresh record), where Φ_t resolves into three parts, none absolutely continuous under the baseline assumptions: a *continuous* part from the heterogeneously-timed crossings of the continuous statistic, singular-continuous in general and continuous under atomless G; *predictable atoms* at scheduled lower exits when a positive cohort shares a reset age (the floor step-ups of Lemma D.1); and an *inaccessible atom* at a common disaster, where one jump pushes every count up at once and all types with ceiling below the new count exit together, $\Delta\Phi_u = b_u$ (the inaccessible upper exits of Proposition D.1); the turnover term $p(\langle\varphi, m_\circ\rangle - \langle\varphi, m_t\rangle)\, dt$; and, when the measure is wealth-weighted with $d\omega_t/\omega_t$ the type-attached relative performance, the product-rule term $d\Sigma_t(\varphi) = \mathrm{Cov}_{m_t^w}(\varphi,\, \bar{r}_t^{\,w})\, dt$ (selection; $\bar{r}_t^{\,w}$ the wealth-weighted fibre average of the relative return). The predictable atoms require a *tolerance-independent* rank-swap date: only floor-steps at such dates synchronize a positive cohort across an atomless G (the shared reset ages of Lemma D.1), while floor movements tied to tolerance-*dependent* cumulative thresholds disperse, a rank-swap step contributing an atom only over the positive-mass set of tolerances it re-orders. Predictable atoms thus require a shared reset age in either channel. Given one, the crash channel schedules them at the tolerance-independent rank-swap dates above; the trend channel schedules them at the cohort's probation expiry $\sigma + a_0$, where the tolerances whose bands the common record sits outside at a_0—a set of positive G-mass when the record clears the smallest critical values—exit at the predictable date, each exiter re-adopting on the spot, so the cohort propagates. What is channel-specific is the cohort's origin, not the atom's possibility: crossings of the continuous trend statistic are heterogeneously timed and create no shared reset age endogenously, while a common disaster synchronizes upper exits into exactly such a cohort; absent such a cohort—none arises without initialization or a prior synchronized exit—scheduled exits disperse in both channels, and predictable atoms are not generic across all floor movements.

Ex-ante crossing rate. Let $\mu_t \in \mathcal{P}(\bar{\mathcal{S}})$ denote the law at t of the full population state $\bar{s} = (\hat{\theta}, \alpha, a, R, \rho)$ of §4—the law of $\bar{\Psi}_t(\xi)$ under the data measure with the type drawn from G, the mean of the pathwise full-state cross-section $\bar{\Psi}_t \# G$—and $\pi\#\mu_t$ its $(\hat{\theta}, \alpha, a)$-marginal under the projection π of §4, the mean of the cross-section m_t of (13). The ex-ante expected crossing rate is channel-specific. In the *crash channel* the exit carries the genuine state-dependent rate χ of (C.3), so the expected crossing rate is $\langle\chi, \mu_t\rangle = \int_{\bar{\mathcal{S}}} \chi(\hat{\theta}, R, a; \alpha)\, d\mu_t(\bar{s})$, the tolerance α a coordinate of the type carried by the state; because $\chi(\hat{\theta}, x, a; \alpha) = \lambda^\star \mathbf{1}\{x = u_\alpha(a)\}$ depends on the record only through the event $\{R = u_\alpha(a)\}$ at the type's *own* ceiling, the inner integral over the (R, ρ)-fibre at fixed $(\hat{\theta}, \alpha, a)$ evaluates, type by type, to $\lambda^\star\, \nu_a(\{u_\alpha(a)\} \mid \hat{\theta}, \alpha, a) = h(\hat{\theta}, \alpha, a)$ and it pushes down to the single-spell hazard of each tolerance type, $\langle\chi, \mu_t\rangle = \langle h, \pi\#\mu_t\rangle$ with h carrying the $(\hat{\theta}, \alpha, a)$ arguments—no cross-type averaging, the fibres being tolerance-specific—the pushdown well defined because the marginal $\pi\#\mu_t$, like the cross-section m_t of (13), carries the tolerance coordinate; and $\langle\chi, \mu_t\rangle$ is the *inaccessible upper-exit* component of the crash-channel crossing rate, the scheduled floor-step exits being the predictable-atom part of Φ_t, outside χ. In the *trend channel* the exit is a predictable boundary hit carrying no *pathwise* rate χ, so the pathwise pushdown above has no trend analogue; the ex-ante marginal hazard h of (9) remains well defined from the holding-time law, but the object that closes the flux term of (14) is the boundary flux Φ_t itself, not $\langle h, m_t \rangle$. $\square$

Proof of Corollary 4. At a common-surprise date u the only mass that moves is the crossing mass b_u, carried by the post-revision map J from its pre-jump location to $J\#b_u$, the interior of m being unaffected: $m_u - m_{u-} = J\#b_u - b_u$. For a linear index,

$$\Delta\langle \varphi, m_u \rangle = \int \varphi \, d(J\#b_u) - \int \varphi \, db_u = \int \big(\varphi \circ J - \varphi\big) \, db_u, \tag{E.2}$$

which is (16). For a nonlinear aggregate $\Gamma(\langle \varphi, m \rangle)$ the jump is the finite increment $\Gamma(\langle \varphi, m_u \rangle) - \Gamma(\langle \varphi, m_{u-} \rangle)$; this is not $\Gamma'(\langle \varphi, m_{u-} \rangle) \, \Delta\langle \varphi, m_u \rangle$ because the jump in $\langle \varphi, m \rangle$ is of finite size, not infinitesimal, so the first-order expansion does not close. $\square$

References

Larbi Alili, Pierre Patie, and Jesper Lund Pedersen. Representations of the first hitting time density of an ornstein–uhlenbeck process. *Stochastic Models*, 21(4): 967–980, 2005.

Fernando Alvarez, Francesco Lippi, and Aleksei Oskolkov. The macroeconomics of sticky prices with generalized hazard functions. *Quarterly Journal of Economics*, 137(2):989–1038, 2022.

F. J. Anscombe. Fixed-sample-size analysis of sequential observations. *Biometrics*, 10(1):89–100, 1954.

Peter Armitage, C. K. McPherson, and B. C. Rowe. Repeated significance tests on accumulating data. *Journal of the Royal Statistical Society. Series A (General)*, 132(2):235–244, 1969.

Cuimin Ba. Robust misspecified models. *American Economic Review*, 116(4):1340–1379, 2026.

Isaac Baley and Javier Turén. Lumpy forecasts. CEPR Discussion Paper 19824, Centre for Economic Policy Research, 2025.

Robert H. Berk. Limiting behavior of posterior distributions when the model is incorrect. *Annals of Mathematical Statistics*, 37(1):51–58, 1966.

Allan Birnbaum. Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. *The Annals of Mathematical Statistics*, 26(1):21–36, 1955.

David Blackwell and Lester Dubins. Merging of opinions with increasing information. *Annals of Mathematical Statistics*, 33(3):882–886, 1962.

David Blackwell and David Freedman. A remark on the coin tossing game. *Annals of Mathematical Statistics*, 35:1345–1347, 1964.

Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Diagnostic expectations and credit cycles. *The Journal of Finance*, 73(1):199–227, 2018.

Pedro Bordalo, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–2782, 2020.

George E. P. Box and George C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA, 1973.

Leo Breiman. First exit times from a square root boundary. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, Part 2, pages 9–16. University of California Press, Berkeley, 1967.

Ricardo J. Caballero and Eduardo M. R. A. Engel. Explaining investment dynamics in U.S. manufacturing: A generalized (S,s) approach. *Econometrica*, 67(4):783–826, 1999.

In-Koo Cho and Kenneth Kasa. Learning and model validation. *Review of Economic Studies*, 82(1):45–82, 2015.

In-Koo Cho, Noah Williams, and Thomas J. Sargent. Escaping nash inflation. *Review of Economic Studies*, 69(1):1–40, 2002.

Pierre Collet, Servet Martínez, and Jaime San Martín. Asymptotic laws for one-dimensional diffusions conditioned to nonabsorption. *Annals of Probability*, 23(3): 1300–1314, 1995.

Pierre Collet, Servet Martínez, and Jaime San Martín. *Quasi-Stationary Distributions: Markov Chains, Diffusions and Dynamical Systems*. Springer, Berlin, 2013.

Pierre Collin-Dufresne, Michael Johannes, and Lars A. Lochstoer. Parameter learning in general equilibrium: The asset pricing implications. *American Economic Review*, 106(3):664–698, 2016.

D. R. Cox. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):433–441, 1955.

Edwin L. Crow and Robert S. Gardner. Confidence intervals for the expectation of a Poisson variable. *Biometrika*, 46(3–4):441–453, 1959.

Miklós Csörgő and Pál Révész. *Strong Approximations in Probability and Statistics*. Academic Press, New York, 1981.

M. H. A. Davis. Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society, Series B*, 46 (3):353–388, 1984.

Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential B: Theory of Martingales*. North-Holland, Amsterdam, 1982.

Denis Denisov, Alexander Sakhanenko, Sara Terveer, and Vitali Wachtel. Random walks with square-root boundaries: the case of exact boundaries $g(t) = c\sqrt{t+b} - a$. arXiv:2501.04554, 2025.

Ian Dew-Becker, Stefano Giglio, and Pooya Molavi. Learning and the emergence of nonlinearity in financial markets. NBER Working Paper 34584, National Bureau of Economic Research, 2025.

Larry G. Epstein and Michel Le Breton. Dynamically consistent beliefs must be Bayesian. *Journal of Economic Theory*, 61(1):1–22, 1993.

Ignacio Esponda and Demian Pouzo. Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, 84(3):1093–1130, 2016.

Mira Frick, Ryota Iijima, and Yuhta Ishii. Belief convergence under misspecified learning: A martingale approach. *Review of Economic Studies*, 90(2):781–814, 2023.

Drew Fudenberg, Giacomo Lanzani, and Philipp Strack. Limit points of endogenous misspecified learning. *Econometrica*, 89(3):1065–1098, 2021.

Priscilla Greenwood and Edwin Perkins. A conditioned limit theorem for random walk and Brownian local time on square root boundaries. *Annals of Probability*, 11 (2):227–261, 1983.

Peter Grünwald, Rianne de Heide, and Wouter M. Koolen. Safe testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(5):1091–1128, 2024.

Philip Hartman and Aurel Wintner. On the law of the iterated logarithm. *American Journal of Mathematics*, 63(1):169–176, 1941.

Rob J. Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50(2):120–126, 1996.

Ehud Kalai and Ehud Lehrer. Rational learning leads to Nash equilibrium. *Econometrica*, 61(5):1019–1045, 1993.

Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer, New York, 2nd edition, 1991.

John Kent. Eigenvalue expansions for diffusion hitting times. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 52(3):309–319, 1980.

János Komlós, Péter Major, and Gábor Tusnády. An approximation of partial sums of independent RV's and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32:111–131, 1975.

János Komlós, Péter Major, and Gábor Tusnády. An approximation of partial sums of independent RV's and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34:33–58, 1976.

Julian Kozlowski, Laura Veldkamp, and Venky Venkateswaran. The tail that wags the economy: Belief-driven business cycles and persistent stagnation. *Journal of Political Economy*, 128(8):2839–2879, 2020.

Uwe Küchler and Steffen L. Lauritzen. Exponential families, extreme point models and minimal space-time invariant functions for stochastic processes with stationary and independent increments. *Scandinavian Journal of Statistics*, 16(3):237–261, 1989.

Uwe Küchler and Michael Sørensen. Exponential families of stochastic processes: A unifying semimartingale approach. *International Statistical Review*, 57(2):123–144, 1989.

Uwe Küchler and Michael Sørensen. *Exponential Families of Stochastic Processes*. Springer Series in Statistics. Springer, New York, 1997.

Uwe Küchler and Michael Sørensen. On exponential families of Markov processes. *Journal of Statistical Planning and Inference*, 66(1):3–19, 1998.

Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses*. Springer, New York, 3 edition, 2005.

Ulrike Malmendier and Stefan Nagel. Depression babies: Do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics*, 126(1):373–416, 2011.

T. K. Matthes and D. R. Truax. Tests of composite hypotheses for the multivariate exponential family. *The Annals of Mathematical Statistics*, 38(3):681–697, 1967.

Gordon D. Menzies and Daniel John Zizzo. Inferential expectations. *The B.E. Journal of Macroeconomics*, 9(1), 2009.

Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2nd edition, 2009.

Marcel F. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Johns Hopkins University Press, Baltimore, 1981.

Alexander A. Novikov. The crossing time of a one-sided nonlinear boundary by sums of independent random variables. *Theory of Probability and its Applications*, 27: 688–702, 1982.

Pietro Ortoleva. Modeling the change of paradigm: Non-bayesian reactions to unexpected news. *American Economic Review*, 102(6):2410–2436, 2012.

Pietro Ortoleva. Alternatives to bayesian updating. *Annual Review of Economics*, 16, 2024.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

Valentin V. Petrov. *Sums of Independent Random Variables*. Springer, Berlin, 1975.

Ross G. Pinsky. On the convergence of diffusion processes conditioned to remain in a bounded region for large time to limiting positive recurrent diffusion processes. *Annals of Probability*, 13(2):363–378, 1985.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38(4):576–601, 2023.

Herbert Robbins. Statistical methods related to the law of the iterated logarithm. *Annals of Mathematical Statistics*, 41(5):1397–1409, 1970.

Herbert Robbins and David Siegmund. Boundary crossing probabilities for the Wiener process and sample sums. *Annals of Mathematical Statistics*, 41(5):1410–1429, 1970.

Ken-iti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, 1999.

David Siegmund. *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York, 1985.

Theodore E. Sterne. Some remarks on confidence or fiducial limits. *Biometrika*, 41 (1/2):275–278, 1954.

Jessica A. Wachter and Yicheng Zhu. Learning with rare disasters. *Quantitative Economics*, 16(4):1189–1221, 2025.

Jonathan Weinstein. Bayesian inference tempered by classical hypothesis testing. Working paper, Washington University in St. Louis, 2017.

Martin L. Weitzman. Subjective expectations and asset-return puzzles. *American Economic Review*, 97(4):1102–1130, 2007.

David Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, 1991.