Self-Fulfilling Accuracy

Dmitry Livdan* Bernardo Maciel† Norman Schürhoff‡

October 31, 2026§

Abstract

Predictive models can change the outcomes used to evaluate and retrain them. We characterize the structure, identification, and stochastic dynamics of this feedback. Under a one-sided feedback bound, a stable predictor is a strictly increasing transformation of the posterior mean of a latent fundamental. It preserves rankings and information and remains calibrated, even when reinforcing feedback raises deployment fit. Deterministic deployment generally does not separate fundamentals from feedback. Connected policy variation, randomized holdouts, and model archives provide identifying variation. Under conditionally mean-zero archive disturbances, transition differences identify nonlinear feedback differences. In a linear-Gaussian benchmark, fresh-sample retraining is stable exactly when $|k| < 1$, while pooled least squares converges for every $k < 1$ and slows divergence otherwise. Stability depends on both economic feedback and the updating rule. With stateful feedback, sufficiently large panels are stable below the population frontier, while suitable finite blocks expand in expected log norm above it.

*Haas School of Business, University of California, Berkeley, and CEPR. Email livdan@haas.berkeley.edu.
†Haas School of Business, University of California, Berkeley. Email bernardo_maciel@berkeley.edu.
‡University of Lausanne and Swiss Finance Institute and CEPR. Email norman.schuerhoff@unil.ch.
§New draft, comments welcome.

A credit score helps determine the financing a borrower receives, and that financing can change subsequent repayment. When repayment enters the next training sample, the model learns from outcomes that its own earlier predictions helped produce. The same issue arises when recommendations influence sales or clinical scores guide treatment. Recent evidence from consumer credit shows that score changes can affect subsequent default even when the underlying credit-report information is held fixed (Duarte and Fonseca, 2026). Repeated retraining carries this effect of deployment into later prediction rules.

Measured accuracy can therefore have two sources. A model may predict an outcome well because its features contain information about an underlying fundamental or because deploying the prediction helps produce the outcome. We call an increase in deployment fit arising through the second channel *self-fulfilling accuracy*. For example, favorable predictions can direct financing to borrowers whose repayment prospects are already stronger. If that financing further improves the repayments, it widens the outcome differences associated with the original prediction. The model can then explain a larger share of outcome variation while containing the same information about repayment capacity.

This distinction matters even when the model is estimated correctly. A lender testing predictions on new borrowers under the same allocation rule observes outcomes shaped by the same feedback mechanism as in training data and test outcomes. Good performance can establish that the model predicts well under the current rule. It does not, by itself, separate the repayment capacity from the contribution of financing or establish what would happen under different terms. Repeated retraining raises a further question: will successive updates converge, fluctuate around a stable benchmark, or amplify prediction errors and become unstable? Each revision changes financing decisions, and therefore the outcomes used to estimate the next model, so the answer depends on both the economic feedback and the updating rule. Evaluating the score, identifying the effect of deployment, and understanding retraining dynamics therefore require a common account of how predictions enter outcomes.

We study these questions in a model that holds the joint distribution of fundamentals, features, and outcome shocks fixed. Let $u(x) = \mathbb{E}[\theta \mid x]$ denote the posterior mean of a latent fundamental, and let $f(s)$ denote the effect of deploying score s on the outcome. The features can be rich, but their effect on the allocation operates through the scalar score. Population retraining estimates the conditional mean of the outcome generated under the current deployment. A predictor that is unchanged by this operation therefore satisfies $m(x) = u(x) + f(m(x))$. The stable prediction incorporates both the information in the features and the outcome response induced by the use of that prediction.

Our structure theorem characterizes this fixed point. If the slope of f is uniformly bounded above by a constant below one, the stable predictor is unique and takes the form $m^*(x) = \Phi(u(x))$, where Φ is strictly increasing and invertible. It preserves posterior rankings and contains exactly the same information about the fundamental as $u(x)$. It also remains calibrated to the outcomes generated under its own deployment. The feedback bound allows for negative responses of large magnitude. Consequently, the existence of a stable predictor does not guarantee that repeated retraining will approach it. An additional bound on the absolute feedback gain guarantees convergence, a distinction that becomes important for comparing learning rules.

Feedback can nevertheless change measured fit substantially. Affine positive feedback stretches the differences in the predictable component of outcomes while leaving the irreducible prediction error unchanged. If both the posterior mean and the irreducible prediction error have positive variance, the deployment R^2

increases. Under the same conditions, affine negative feedback compresses these differences and reduces R^2. An effective treatment triggered by a high risk score can thus reduce measured fit without reducing the score's fundamental information. For nonlinear responses, we give conditions under which stronger feedback also weakly increases fitted variance and deployment R^2. These comparisons show why calibration alone cannot establish that a model has learned more about the fundamental. Calibration is assessed against outcomes that already incorporate the effect of deployment.

Separating the two components requires variation in the deployed score or allocation whose effect on outcomes can be distinguished from differences in fundamentals. For example, comparing outcomes at the same feature value under different scoring rules holds the posterior mean fixed while varying the feedback response. A single deterministic deployment does not provide this comparison because each feature value receives only one allocation. Without additional restrictions, different combinations of posterior fundamentals and feedback can, therefore, generate the same observed outcomes. The structure theorem still identifies the ranking of posterior means from a stable score, and known bounds on positive feedback translate observed score differences into bounds on differences in posterior fundamentals. Recovering the feedback response requires additional variation across deployment rules, experimental allocations, or model vintages under the identifying conditions developed below.

Multiple deterministic regimes provide such comparisons when the feedback response and posterior fundamentals remain common across regimes. The conditional outcome differences then identify the feedback differences between the scores assigned to the same feature value. These comparisons form a *deployment graph*. A connected graph identifies the feedback and the posterior mean up to common-level normalization, while cycle restrictions supply specification tests. Randomized holdouts provide an experimental benchmark by assigning a fixed baseline allocation independently of the fundamentals. The threshold rules identify a feedback jump when the score-conditional fundamental mean is continuous at the cutoff. Each design supplies a different way to separate the effect of allocation from the information used to assign it.

Stored models provide another source of identifying variation. An institution can evaluate successive model vintages on a fixed set of audit features, even if it never observes the same economic unit twice. The predictions then change, while the feature value at which they are evaluated remains fixed. When retraining errors have conditional mean zero given the archive history, the transition mean satisfies $\mathbb{E}[M_{t+1}(x) \mid M_t(x) = s] = u(x) + f(s)$. Differences in this transition mean identify nonlinear feedback differences without observing the latent fundamental. The mean-zero condition is important. If the estimation bias varies with the current score, the identified contrast includes the corresponding change in bias. We characterize this distortion and provide sensitivity bounds when the variation in bias can itself be bounded.

Identification supports a counterfactual calculation. Once the posterior fundamentals have been recovered, a new allocation rule can be evaluated by solving its fixed-point equation with a specified feedback response. The calculation requires fundamentals to remain unchanged and the new response to be known over the score values reached under the counterfactual regime. It can be performed before the outcomes under that rule are observed. The common additive normalization left by connected regime comparisons or randomized holdouts does not affect the result when the factual and counterfactual responses use the same normalization.

The archive also records how the estimation error propagates through deployment. Intuitively, the effect

of a perturbation to a stable score is governed by the feedback slope. In a linear-Gaussian benchmark, this relationship becomes an exact autoregression, $\lambda_{t+1} = \rho + k\lambda_t + U_t$, where k is the feedback slope, ρ is the feature reliability, and U_t is the OLS sampling error from a fresh cross-section. Gaussian sampling gives the innovations an exact scaled Student distribution. Fresh independent cross-sections make them independent across vintages and of the previously deployed coefficient. When stationary second moments exist, the archive autocorrelation identifies k, the stationary mean identifies ρ given k, and the variance provides an additional testable restriction. Ordinary estimation noise thus supplies the variation that a deterministic deployment lacks.

The same representation gives a distinct stochastic stability frontier. For $|k| < 1$, the coefficient process is geometrically ergodic, with continuing fluctuations around the population fixed point. With at least three observations per cross-section, $k = 1$ produces a random walk with positive drift, while $k = -1$ produces unbounded oscillatory excursions. Beyond either boundary, coefficient magnitudes diverge geometrically. Retaining all past observations changes these conclusions because the influence of each new cross-section declines as the archive grows. The pooled least squares converges to the fixed point for every $k < 1$, including arbitrarily strong negative feedback. For $k = 1$, pooling slows the divergence from linear to logarithmic, and for $k > 1$ it slows the divergence from geometric to polynomial. The same economic response can therefore generate stable or unstable retraining depending on how observations are reused.

Durable allocations introduce an additional source of persistence. Investment induced by an earlier score can continue to affect outcomes after the model is updated, while persistent fundamentals keep that investment correlated with the features used in later predictions. The newly estimated coefficient then reflects both current feedback and the consequences of earlier deployments. Retraining on a fixed panel generates a random linear recurrence, so the stability depends on how successive random matrices amplify or contract perturbations. A negative top Lyapunov exponent yields a unique causal stationary and ergodic solution. We show that every population economy strictly below the stability frontier has a negative finite-panel exponent for all sufficiently large panels.

The corresponding result above the frontier remains incomplete. We establish that suitable finite blocks expand in expected log norm and with probability approaching one as the panel grows. Expansion over a finite block does not by itself establish a positive long-run exponent because the directions amplified in one block may be contracted in the next. We give a sufficient alignment condition and isolate the remaining question of whether the primitive Gaussian panel dynamics satisfy it. Establishing that condition, together with resolving the exact critical case, would complete the link between the population frontier and stochastic stability in large panels.

Related literature. The paper builds on performative prediction literature. Perdomo et al. (2020) introduce performative stability and repeated risk minimization. Subsequent work develops stochastic optimization and methods for minimizing performative risk (Mendler-Dünner et al., 2020; Miller et al., 2021; Izzo et al., 2021; Drusvyatskiy and Xiao, 2023). Hardt and Mendler-Dünner (2025) survey this literature. Recent contributions study predictability and self-validation (Perdomo, 2025), decision making under outcome performativity (Kim and Perdomo, 2023), convergence of no-regret learning to mixed performatively stable equilibria (Farina and Perdomo, 2026a), and stability with few model deployments (Farina and Perdomo, 2026b). We study what a stable predictor reveals about a fixed fundamental, which

deployment variation identifies the feedback response, and how finite-sample estimation shapes the retraining process.

Hardt (2026) is especially close to the affine benchmark. His stable-signal principle shows how regularization and normalization can recover model-independent signal directions under strong affine performative feedback. Our analysis connects the information represented by a stable score to nonidentification under deterministic deployment and to the exact stochastic law of model vintages. Estimation noise is central to that law and supplies variation that can identify the economic feedback gain.

Earlier deployments can also affect the distribution faced by the current model. Brown et al. (2022) study this possibility in a stateful environment. Gradual and persistent responses are examined by Izzo et al. (2022), Li and Wai (2022), and Ray et al. (2022). Lee and Zrnic (2026) allow performative feedback to coexist with exogenous changes in the data distribution. Our stateful analysis derives the random recurrence generated by repeated estimation on a finite panel under stationary primitives and relates its Lyapunov exponent to the population stability frontier.

The identification problem is closely related to overlap in treatment-effect econometrics (Rosenbaum and Rubin, 1983; Hahn, 1998; Crump et al., 2009; D'Amour et al., 2021). A deterministic allocation rule leaves no variation in allocation conditional on the feature. Holdouts, changes in deployment rules, and threshold comparisons provide distinct ways to obtain the variation needed to identify feedback under their respective assumptions. Mendler-Dünner et al. (2022) also study how performative responses can be anticipated from predictions. Our observational-equivalence results specify what deployment data leave unidentified and what additional variation resolves that ambiguity.

The paper also relates to strategic scoring (Frankel and Kartik, 2022; Hennessy and Goodhart, 2023; Ball, 2025). Strategic changes in the features supplied to a model can reduce their information content. Our baseline holds the feature distribution fixed and allows deployment to change the outcome conditional on the feature. This setting isolates how the measured fit can improve while the stable score retains the same fundamental information.

Finally, the retraining dynamics connect to adaptive learning and random recurrence models. Pooling produces a decreasing gain, linking its stability properties to Marcet and Sargent (1989), Sargent (1999), and Evans and Honkapohja (2001). The finite-panel stateful analysis uses the random-recurrence theory of Brandt (1986) and Bougerol and Picard (1992), together with the theory of random matrix products and subadditive growth in Furstenberg and Kesten (1960) and Kingman (1973). These tools characterize how estimation disturbances persist when each fitted model changes the outcomes used to train its successors.

1 Stable Predictors under Outcome Feedback

This section characterizes the population theory of stable predictors, which establishes the fixed point of the retraining operator and its stability.

1.1 Setup

A unit has a latent fundamental $\theta \in \mathbb{R}$, observed feature x taking values in a Polish space $\mathbb{X}$, and the outcome shock $\varepsilon \in \mathbb{R}$. Their joint law μ satisfies $\mathbb{E}[\theta^2] < \infty$, $\mathbb{E}[\varepsilon^2] < \infty$, and $\mathbb{E}[\varepsilon \mid \theta, x] = 0$, with P_X

denoting the marginal distribution of x under μ. The posterior mean of the fundamental

$$u(x) \coloneqq \mathbb{E}[\theta \mid x], \tag{1}$$

is the optimal predictor of θ under squared loss and provides the benchmark outcome prediction in the absence of deployment feedback. The full conditional distribution of θ given x need not be determined by $u(x)$. The fundamental may represent repayment capacity, productivity, demand, health risk, or similar characteristics.

The modeler deploys a measurable score $s = m(x)$, which induces an action $a = g(s)$. Let $b(a)$ denote the effect of that action on the realized outcome. The composite feedback response $f\colon \mathbb{R} \to \mathbb{R}$ is

$$f(s) = b(g(s)). \tag{2}$$

We study outcomes of the form

$$y = \theta + f(m(x)) + \varepsilon. \tag{3}$$

This formulation accommodates credit, ranking, and treatment applications without requiring separate specifications of the allocation rule g and the outcome response b.

Assumption 1. *The feedback response f is differentiable and satisfies $f'(t) \leq \kappa^+ < 1$ for every $t \in \mathbb{R}$. Define the local loop gain by $k(t) \coloneqq f'(t)$ and the uniform absolute gain by $\overline{\kappa}_{\text{abs}} \coloneqq \sup_{t\in\mathbb{R}} |f'(t)|$.*

A useful linear specification takes $g(s) = a_0 + \gamma s$ and $b(a) = \beta a$. The resulting feedback response is

$$f(s) = \beta\gamma s + \beta a_0. \tag{4}$$

A constant f is the benchmark without score-dependent feedback. Its level may shift the outcomes, but changing the deployed score then has no effect on them.

Under (2), the chain rule gives $k(t) = b'(g(t))g'(t)$ whenever derivatives exist.[1] The one-sided bound in Assumption 1 keeps the score-minus-feedback map invertible. When convergence of the retraining sequence is needed, we impose the stronger condition $\overline{\kappa}_{\text{abs}} < 1$.

The modeler retrains the model nonparametrically using squared loss. We consider deployed scores m for which $f(m(x))$ has a finite second moment. Given such a score, let $y^{(m)}$ denote the outcome generated under such deployment. The next predictor is the population regression function

$$(Tm)(x) \coloneqq \mathbb{E}\left[y^{(m)} \mid x\right] = u(x) + f(m(x))\,. \tag{5}$$

The operator separates the information $u(x)$ from the outcome response caused by the deployment $f(m(x))$. Therefore, population retraining reveals only their sum.

Definition 1. *A score m^* is performatively stable if $T\,m^* = m^*$ μ-almost surely.*

This is the fixed-point concept of Perdomo et al. (2020) for a predictor class on which the squared-loss

[1] The existence, ordering, and Lipschitz conclusions extend to Lipschitz responses satisfying the corresponding slope bound almost everywhere. Results involving local derivatives additionally require differentiability at the relevant stable scores.

regression is defined. Equality and uniqueness are understood to hold μ-almost surely. For pointwise arguments, we fix measurable versions of u and T throughout.

1.2 The structure theorem

The structure theorem characterizes the form of every performatively stable predictor and explains how feedback shapes its predictions. Its central result is that the stable score is a strictly increasing transformation of the posterior mean $u(x)$. Feedback can therefore change score levels and differences while preserving the ordering and information contained in the posterior mean. This characterization provides the foundation for distinguishing improvements in fit against deployment outcomes from improvements in information about fundamentals. The theorem also establishes that the stable predictor is unique and gives an additional condition under which repeated population retraining converges to it.

Start with the relation (5). Because (5) acts pointwise, stability requires that

$$m(x) = u(x) + f\big(m(x)\big). \tag{6}$$

The stable score equals the informational component plus the outcome response induced by that score. Equivalently,

$$u(x) = m(x) - f\big(m(x)\big). \tag{7}$$

This representation separates the informational component from the feedback component. Define

$$h(t) \coloneqq t - f(t). \tag{8}$$

The function h measures the portion of a score that remains after subtracting the outcome effect generated by deploying that score. Assumption 1 implies that h is strictly increasing. A higher score always corresponds to a higher inferred fundamental because the feedback response grows less than one-for-one with the score. The structure theorem formalizes this observation and characterizes the stable predictor.

Theorem 1 (Structure)**.** *Under Assumption 1, the following statements hold.*

(i) The map $h(t)$ is a strictly increasing bijection. Its inverse $\Phi \coloneqq h^{-1}$ is strictly increasing and differentiable, with

$$\Phi'(u) = \frac{1}{1 - k\big(\Phi(u)\big)} > 0. \tag{9}$$

Moreover, Φ is globally Lipschitz with constant $1/(1-\kappa^+)$, so

$$|\Phi(u_2) - \Phi(u_1)| \le \frac{|u_2 - u_1|}{1 - \kappa^+} \qquad \text{for all } u_1, u_2 \in \mathbb{R}. \tag{10}$$

(ii) There is a unique performatively stable score up to μ-almost-sure equality given by

$$m^*(x) = \Phi\big(u(x)\big). \tag{11}$$

(iii) Suppose in addition that $\overline{\kappa}_{\mathrm{abs}} < 1$. For any initial score m_0 with a finite second moment, the sequence

defined by $m_{n+1} = Tm_n$ converges pointwise to m^* at a geometric rate. In particular,

$$|m_n(x) - m^*(x)| \leq \overline{\kappa}_{\text{abs}}^n |m_0(x) - m^*(x)| \qquad \textit{for every } x \in \mathbb{X}. \tag{12}$$

Theorem 1 provides a complete representation of the stable predictor. The fixed-point equation in (7) becomes $h(m^*(x)) = u(x)$, and inversion gives equation (11). The stable score is therefore obtained by applying the same scalar transformation Φ to the posterior mean for every unit. The distribution of features may be high-dimensional, and a complex prediction technology may generate the posterior mean, yet feedback enters the stable score through a one-dimensional map.

The economic content of the result is easiest to see in a lending example. Two borrowers with posterior repayment capacities $u(x_1)$ and $u(x_2)$ receive stable scores $m^*(x_1) = \Phi(u(x_1))$ and $m^*(x_2) = \Phi(u(x_2))$. Since Φ is strictly increasing, the borrower with the higher expected repayment capacity always receives the higher stable score. Feedback changes the numerical distance between their scores because the induced financing affects their subsequent outcomes, but does not reverse their ranking.

Equation (9) describes the local amplification created by this feedback. A marginal increase in the posterior mean raises the stable score by $\Phi'(u) = 1/[1 - k(\Phi(u))]$. When the local loop gain is positive, the multiplier exceeds one. A unit of information about the fundamental then produces more than a unit change in the stable score because the score generates an allocation that reinforces the predicted outcome. In a credit market, a stronger repayment signal increases the score. The higher score improves the financing terms, and the resulting investment improves repayment. In a recommendation system, higher predicted demand raises visibility, which generates additional demand and magnifies the final score.

The local gain may vary across score levels. Financing conditions may respond strongly to a score near an approval threshold and more weakly elsewhere, while a platform ranking may have a larger effect on demand near the top of a search page. The transformation Φ can therefore be nonlinear even when the information summarized by $u(x)$ is one-dimensional. Its slope remains positive throughout, so the transformation may stretch different parts of the score distribution by different amounts without reversing the ordering of units.

When the feedback response is nondecreasing, the multiplier in (9) is between 1 and $1/(1 - \kappa^+)$ at every posterior mean. This remains true when the local gain varies across scores. Integrating the multiplier provides a bound on the difference between any two stable scores.

Corollary 1 (Pointwise amplification under nondecreasing feedback)**.** *Suppose $0 \leq f'(t) \leq \kappa^+ < 1$ for every $t \in \mathbb{R}$. Then, for any $u_2 > u_1$,*

$$u_2 - u_1 \leq \Phi(u_2) - \Phi(u_1) \leq \frac{u_2 - u_1}{1 - \kappa^+}. \tag{13}$$

Thus, positive feedback stretches every difference in posterior means by a factor between one and $1/(1 - \kappa^+)$. Variation in local gain determines how much each part of the score distribution stretches, while the ordering remains the same.

The final part of Theorem 1 separates existence and uniqueness from convergence of repeated retraining. The one-sided restriction in Assumption 1 guarantees that the stable score exists and is unique. Global convergence is guaranteed when the absolute feedback gain is uniformly below one. Under this stronger

condition, the map $t \mapsto u(x) + f(t)$ is a contraction for every fixed x. Each round of retraining reduces the distance to the stable score to at most $\overline{\kappa}_{\text{abs}}$ times its previous value.

The distinction matters when feedback is negative. A clinical intervention may reduce the outcome associated with a high risk score because treatment lowers the patient's subsequent risk. Retraining can then oscillate around the stable score. The absolute-gain condition ensures that these oscillations shrink rather than grow.

The uniform gap from one guarantees the global existence. Without that gap, existence does not need to hold for every posterior mean. If $\sup_t f'(t) = 1$, the derivative may remain strictly below one at every finite point while approaching one sufficiently quickly that h fails to cover the real line. For example, let $f(t) = t - \arctan(t)$. Then $h(t) = \arctan(t)$, whose range is bounded. Posterior means outside this range cannot be reconciled with any stable score. The uniform bound in Assumption 1 rules out this possibility by ensuring that $h(t)$ grows at least linearly in both tails.

These conclusions are the nonparametric counterparts of the stability condition in the affine model. When $f(t) = a + kt$, the inverse map is affine and has slope $1/(1-k)$. The condition $k < 1$ ensures the existence of a stable score, while $|k| < 1$ guarantees the convergence of repeated retraining. Related contraction conditions appear in Perdomo et al. (2020), Drusvyatskiy and Xiao (2023), and the stateful extensions discussed in Section 4. Hardt (2026) develops a complementary stable-signal perspective for affine retraining operators. Theorem 1 gives an explicit nonparametric representation of the stable predictor and makes its information content transparent.

Figure 1 illustrates the structure theorem for affine feedback $f(s) = ks$. Panel (a) plots the stable transformation $m^* = u/(1-k)$ for three feedback gains. Positive feedback stretches scores away from zero, while each transformation remains strictly increasing and invertible. Panel (b) follows the same two posterior means, $u_1 = 1/2$ and $u_2 = 1$, across the three regimes. Their stable-score difference increases from $1/2$ without feedback to 1 when $k = 1/2$ and 2 when $k = 3/4$. The higher posterior mean receives the higher score in every regime. Feedback changes the distance between scores while preserving their ordering and the information they contain about the fundamental.

The monotonicity of Φ has a stronger implication than preservation of rankings. Since the transformation is invertible, observing the stable score reveals exactly the same information as observing the posterior mean. The score may place the information on a different numerical scale, but no information is created or destroyed by the feedback transformation. This observation yields the next result.

Corollary 2 (Information invariance)**.** *Under Assumption 1,*

$$\sigma(m^*(x)) = \sigma(u(x)). \tag{14}$$

Consequently, the following statements hold.

(i) The conditional law of θ given $m^*(x)$ is equal to the conditional law of θ given $u(x)$.

(ii) Mutual information satisfies

$$I(m^*(x); \theta) = I(u(x); \theta). \tag{15}$$

(iii) Every statistic of $(m^*(x), \theta)$ that is invariant to a strictly increasing transformation of the score equals

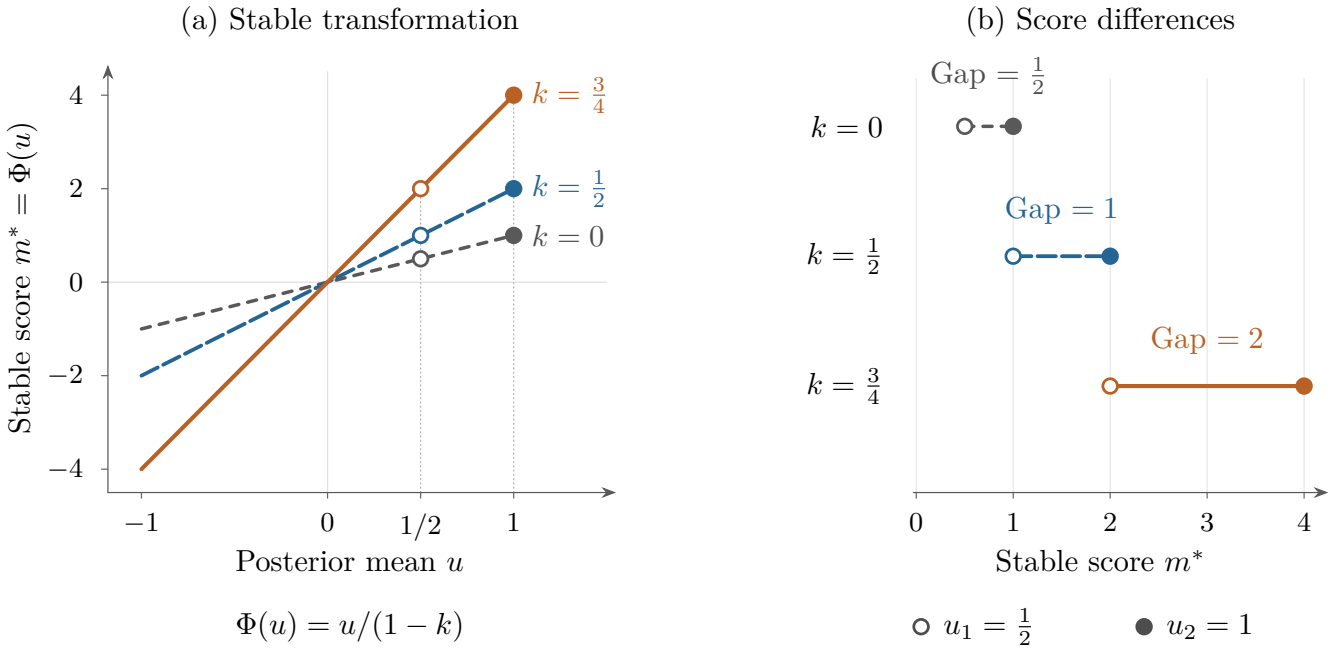


Figure 1: The structure theorem under affine feedback $f(s) = ks$. Panel (a) shows the stable transformation $\Phi(u) = u/(1-k)$ for $k = 0$, $k = 1/2$, and $k = 3/4$, with slopes 1, 2, and 4. Open and filled circles mark the stable scores associated with the same posterior means $u_1 = 1/2$ and $u_2 = 1$. Panel (b) compares these scores on a common scale. Their difference is 1/2, 1, and 2, respectively, while their ranking is unchanged. Each transformation is invertible, so the stable score contains the same information about the fundamental as the posterior mean.

> *the corresponding statistic of* $(u(x), \theta)$*. In particular, Spearman rank correlation and Kendall's tau are preserved.*

These quantities do not depend on the feedback response f*.*

Because Φ is invertible, the stable score and the posterior mean generate the same information. Nonlinear feedback can change cardinal moments such as variance and Pearson correlation, but it cannot change mutual information or rank dependence. In the affine Gaussian specialization, the squared Pearson correlation is also preserved and is equal to ρ.

Information invariance does not imply that the stable score performs poorly when evaluated against realized outcomes. The score affects those outcomes through the allocation, and the regression used for retraining incorporates this causal effect. At a fixed point, the resulting score satisfies an exact calibration condition. The next corollary records this property.

Corollary 3 (Calibration at stability)**.** *Let* m^* *be a performatively stable score satisfying* $\mathbb{E}[|m^*(x)|] < \infty$*. Then*

$$\mathbb{E}[y \mid m^*(x)] = m^*(x). \tag{16}$$

The calibration result follows from the defining property of a population squared-loss fixed point. Stability gives $m^*(x) = \mathbb{E}[y \mid x]$. Applying iterated expectations conditional on the stable score yields equation (16). The argument does not require Assumption 1, differentiability of the feedback response, or

the additive outcome equation (3). It applies to any integrable stable point of nonparametric squared-loss retraining.

Calibration is therefore an equilibrium property of deployment data. Combined with Corollary 2, it shows why calibration cannot distinguish information about fundamentals from an outcome response caused by the score. Related calibration properties of stable predictors appear in Hardt and Mendler-Dünner (2025), Perdomo (2025), and Kim and Perdomo (2023).

1.3 The feedback-invariant residual

The structure theorem and Corollaries 2 and 3 separate three properties of a stable predictor. Feedback preserves the information contained in the posterior mean, the deployed score remains calibrated against realized outcomes, and the numerical scale of the score may change. The remaining question is whether this change in scale improves conventional measures of predictive fit. The answer follows from a simple accounting property of the outcome equation.

For any deployed score m with square-integrable $f(m(x))$, equation (3) implies

$$y - \mathbb{E}[y \mid x] = \theta - u(x) + \varepsilon. \tag{17}$$

The feedback term disappears because it is measurable with respect to x. A lender may improve a borrower's repayment outcome by extending cheaper credit, and a platform may raise a product's sales by increasing its visibility, but these effects are predictable from the score and are therefore absorbed by the fitted conditional mean. They do not enter the regression residual.

Lemma 1 (Feedback-invariant residual)**.** *For any deployed score m with square-integrable $f(m(x))$, the conditional law of $y - \mathbb{E}[y \mid x]$ given x is equal to the conditional law of $\theta - u(x) + \varepsilon$ given x. This law does not depend on f or m. In particular,*

$$\sigma^2_{\mathrm{res}} \coloneqq \mathbb{E}[\mathrm{Var}(y \mid x)] = \mathbb{E}[\mathrm{Var}(\theta \mid x)] + \mathbb{E}[\mathrm{Var}(\varepsilon \mid x)] \tag{18}$$

is the same in every feedback regime.

The entire conditional residual law is invariant, not only its variance. This identity drives both the fit comparison below and the exact finite-sample retraining law in Section 3.

1.4 Self-fulfilling accuracy

Accuracy generally refers to the fit against realized deployment outcomes. *Self-fulfilling* accuracy is an increase in that fit because the score changes the outcomes it predicts. A higher score may induce an allocation that increases the subsequent outcome. Then, differences in realized outcomes become easier to predict from the score, even though the score contains the same information about the latent fundamental.

To determine how measured fit changes with feedback, this section studies the dispersion of the fitted conditional mean holding the distribution of fundamentals, features, and outcome shocks fixed. Take $\overline{\tau} > 0$

and for the regime $\tau \in [0, \overline{\tau}]$, parameterize the feedback strength by

$$f_\tau(s) = \tau f(s). \tag{19}$$

Here f is the response used above, f_0 is the zero response, and $f_1 = f$ when $1 \leq \overline{\tau}$. We impose the strength bound

$$\overline{\tau} \sup_{t \in \mathbb{R}} f'(t) < 1. \tag{20}$$

This bound ensures that $t \mapsto t - f_\tau(t)$ is a strictly increasing bijection for every $\tau \in [0, \overline{\tau}]$. Let Φ_τ denote its inverse and let the corresponding stable score be

$$m^*_\tau(x) = \Phi_\tau(u(x)). \tag{21}$$

Deployment fit in regime τ is measured by

$$R^2(\tau) \coloneqq \frac{\mathrm{Var}(m^*_\tau(x))}{\mathrm{Var}(m^*_\tau(x)) + \sigma^2_{\mathrm{res}}}. \tag{22}$$

This is the population coefficient of determination from regressing the deployed outcome on x. The stable score is the conditional mean of that outcome, while Lemma 1 gives the irreducible residual variance.[2]

At $\tau = 0$, there is no score-dependent feedback and $m^*_0(x) = u(x)$, implying that

$$R^2(0) = \frac{\mathrm{Var}(u(x))}{\mathrm{Var}(u(x)) + \sigma^2_{\mathrm{res}}}.$$

The difference $R^2(\tau) - R^2(0)$ measures the change in deployment fit within this family of regimes. Corollary 2 implies that $m^*_\tau(x)$ contains the same information about θ as $u(x)$ in every regime. A positive difference in fit can therefore arise without any improvement in information about the fundamental.

A stronger feedback response does not automatically increase score dispersion when the response is nonlinear. The stable score itself moves with τ, so the relevant derivative is evaluated at a different point of the feedback function in each regime. The following condition requires the effective local gain along the stable path to increase with feedback strength.

Definition 2 (Monotone amplification)**.** *The family $\{f_\tau = \tau f\}_{\tau \in [0, \overline{\tau}]}$ satisfies monotone amplification if, for every u, the map $\tau \longmapsto \tau f'(\Phi_\tau(u))$ is nondecreasing on $[0, \overline{\tau}]$.*

The condition accounts for changes in the stable score as the feedback strength changes. It compares the effective local gain at the score reached in each regime. Under affine feedback with a nonnegative slope, increasing τ weakly increases this gain. The condition also holds for a nonnegative convex response when f is nondecreasing. In that case, stronger feedback shifts the stable score upward until the response is at least as steep.

Theorem 2 (Self-fulfilling accuracy)**.** *Let f be nondecreasing with $\overline{\tau} \sup_{t \in \mathbb{R}} f'(t) < 1$. Assume that $\mathrm{Var}(u(x)) < \infty$ and $\mathrm{Var}(u(x)) + \sigma^2_{\mathrm{res}} > 0$. If monotone amplification holds, then $\mathrm{Var}(m^*_\tau(x))$ and $R^2(\tau)$ are nondecreasing on $[0, \overline{\tau}]$. Monotone amplification holds in each of the following cases:*

[2]The definition applies when the denominator in (22) is positive.

(a) The function f is affine with nonnegative slope.

(b) The function f is convex and nonnegative.

Suppose in case (a) that $f'(t) = k > 0$ and $\mathrm{Var}(u(x)) > 0$. Then

$$\mathrm{Var}(m^*_\tau(x)) = \frac{\mathrm{Var}(u(x))}{(1-\tau k)^2}. \tag{23}$$

This variance is strictly increasing in τ. If $\sigma^2_{\mathrm{res}} > 0$, then $R^2(\tau)$ is also strictly increasing.

Under monotone amplification, stronger feedback stretches pairwise score differences while preserving their ordering. Fitted variance rises weakly, and Lemma 1 keeps residual risk fixed. In the affine case, the multiplier $1/(1-\tau k)$ makes the source of the fit gain explicit.

Figure 2 illustrates this mechanism with $\mathrm{Var}(u(x)) = \mathrm{Var}(\theta - u(x)) = 1$ and $\varepsilon = 0$. In panel (a), increasing the effective feedback gain τk from zero to 0.75 increases fitted variance from 1 to 16, while residual variance remains 1. The predictable component therefore accounts for a larger share of outcome variation. Panel (b) shows the corresponding increase in deployment R^2 from 0.50 to $16/17 \approx 0.94$. Meanwhile, the squared correlation between the stable score and the fundamental remains 0.50. The shaded gap, $R^2(\tau) - R^2(0)$, measures the improvement in deployment fit generated by feedback.

Under the strictness conditions in Theorem 2, affine feedback with $k > 0$ strictly increases deployment fit for every $\tau > 0$. This improvement reflects greater dispersion in the predictable component of realized outcomes. Throughout the comparison, the information quantities in Corollary 2 remain constant and calibration in Corollary 3 holds in every regime. Thus, the score becomes more accurate against deployment outcomes without becoming more informative about the fundamental.

Corollary 4 (Affine feedback and deployment fit)**.** *Suppose $f(s) = a + ks$ with $k < 1$ and $\mathrm{Var}(u(x)) < \infty$. Then the stable score satisfies*

$$\mathrm{Var}(m^*(x)) = \frac{\mathrm{Var}(u(x))}{(1-k)^2}. \tag{24}$$

If $\mathrm{Var}(u(x)) > 0$, then relative to the no-feedback benchmark $k = 0$, $0 < k < 1$ strictly increases stable-score dispersion and $k < 0$ strictly lowers it. When $\sigma^2_{\mathrm{res}} > 0$, deployment R^2 changes in the same direction.

Positive feedback stretches differences in realized outcomes in the direction of the original prediction. Negative feedback compresses them. For example, an effective treatment triggered by a high risk score can reduce outcome differences across patients even though the stable score preserves their ordering and fundamental information. In either direction, the score remains calibrated to outcomes under its own deployment. Self-fulfilling accuracy is the positive-feedback case in which deployment R^2 rises while information about θ remains unchanged. A higher deployment R^2 alone therefore does not establish that the model has learned more about the fundamental.[3]

[3]Related concerns about self-validation and intervention-induced updating appear in Perdomo (2025), Liley et al. (2021), and Scarone and Baeza-Yates (2026).

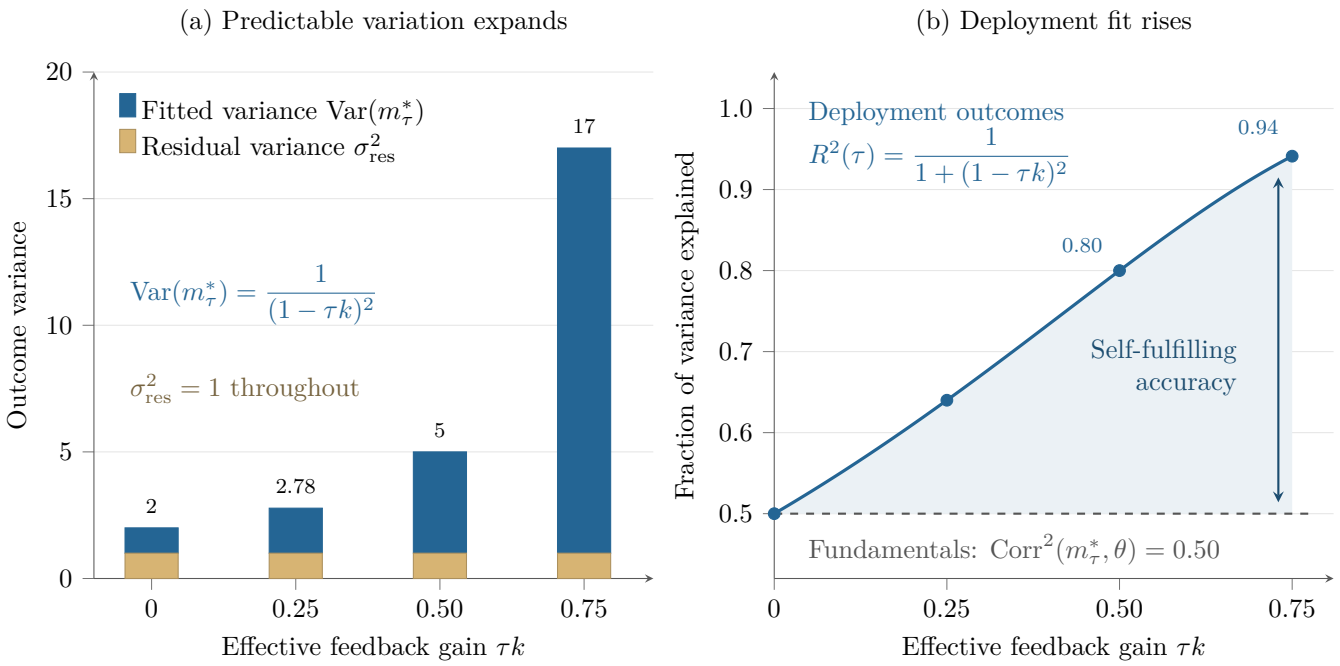


Figure 2: Self-fulfilling accuracy under affine feedback $f_\tau(s) = \tau k s$. The illustration sets $\mathrm{Var}(u(x)) = \mathrm{Var}(\theta - u(x)) = 1$ and $\varepsilon = 0$, so $\sigma^2_{\mathrm{res}} = 1$ and $m^*_\tau(x) = u(x)/(1-\tau k)$. Panel (a) decomposes outcome variance into fitted variance and residual variance. The numbers above the bars report total outcome variance. Panel (b) shows that deployment R^2 rises as feedback strengthens, while the squared correlation between the score and the fundamental remains $1/2$. The shaded gap equals $R^2(\tau) - R^2(0)$. Every stable score is an invertible transformation of the same posterior mean, so its information about the fundamental remains unchanged.

2 Identification

This section first characterizes what a single deployment leaves unidentified and what bounds on feedback can recover. It then studies identification across deterministic regimes, randomized holdouts, and threshold allocations. The final subsection uses an identified decomposition to evaluate a specified counterfactual allocation rule.

2.1 Observational equivalence and partial identification

The deployment data combine information on the fundamentals with the consequences of the allocation generated by the model. In a lending application, borrower characteristics predict repayment capacity, while the credit terms assigned by the score affect investment and subsequent repayment. The conditional mean observed by the lender is therefore

$$\mathbb{E}[y \mid x] = u(x) + f(m(x)). \tag{25}$$

The first term reflects what the feature reveals about the borrower's underlying repayment capacity. The second reflects the improvement or deterioration in repayment caused by the financing decision attached to the score. Data generated in a single deployment reveal their sum, while decomposition into these two

components requires additional variation or restrictions.

Formally, the modeler observes the joint law P of (x, y) generated by equation (3) under a fixed deployed score m. The score does not need to be performatively stable. Define a structure as a pair $S = (\mu, f)$, where μ is a joint law of (θ, x, ε) that satisfies the moment conditions from Section 1.1, and $\mathbb{E}[\varepsilon \mid \theta, x] = 0$, while f is a measurable feedback response. A structure is consistent with P if equation (3), evaluated under that structure and the deployed score m, reproduces the observed law P.

The source of the identification problem can be seen when considering any square-integrable function of the deployed score. An observed repayment advantage among highly scored borrowers can be attributed partly to stronger latent borrowers or in part to the favorable financing induced by their scores. Moving a score-dependent component from the fundamental to the feedback response leaves every observed outcome unchanged. The following theorem characterizes this observational equivalence.

Theorem 3 (Observational equivalence under deployment)**.** *Consider the unrestricted class of structures in which f is any measurable function and μ is any joint law satisfying the setup moment conditions. Assumption 1 is not imposed, and the transformations below need not preserve differentiability, monotonicity, or the loop-gain bound. Let $S = (\mu, f)$ be consistent with P, and let $c\colon \mathbb{R} \to \mathbb{R}$ be any measurable function satisfying $\mathbb{E}\left[c(m(x))^2\right] < \infty$. Define $S_c = (\mu_c, f + c)$, where μ_c is the law of $(\theta - c(m(x)), x, \varepsilon)$. Then S_c is also consistent with P. Consequently, the following statements hold.*

(i) The posterior mean is identified only up to transformations of the form

$$u(x) \longmapsto u(x) - c(m(x)). \tag{26}$$

If m is injective on the support of x, every measurable function of x can be represented as a measurable function of $m(x)$ on that support. The decomposition in equation (25) is then unrestrictedly unidentified.

(ii) Every functional of the observed law P is constant across the identified set. This includes calibration statistics, R^2, proper scoring rules, and any other validation criterion computed solely from deployment data.

(iii) The transformations

$$(\theta, u, f) \longmapsto (\theta - c(m(x)), u - c \circ m, f + c) \tag{27}$$

generate a subset of the identified set. Any structural functional that changes under one admissible transformation is therefore not identified. In particular, the fundamental-information quantities in Corollary 2 are not identified in general. Invariance under the shift family is necessary for identification, although it is not sufficient because the identified set may also contain redecompositions of the latent fundamental and the outcome shock that are not generated by deterministic score-dependent shifts.

The equivalence follows from an exact cancellation. Under the transformed structure, the observed outcome is $y = \theta - c(m(x)) + f(m(x)) + c(m(x)) + \varepsilon = \theta + f(m(x)) + \varepsilon$. The transformation changes the economic interpretation of the outcome without changing its observed distribution. It also preserves the conditional mean restriction on ε because, conditional on x, the original fundamental can be recovered from the transformed fundamental by adding $c(m(x))$.

This equivalence also explains why high deployment accuracy cannot resolve the problem. Calibration, fit, and proper scoring rules are all functionals of the observed distribution, which remains unchanged across observationally equivalent structures. A model that predicts strong fundamentals and has little effect on outcomes can generate the same deployment data as a model that predicts weaker fundamentals and creates a large outcome response through its allocation.

Theorem 3 is an absence-of-overlap result. The deployed allocation is a deterministic function of the feature because it is determined by $m(x)$. Conditional on x, the data contain no otherwise comparable units exposed to a different allocation. In a lending application, every borrower with the same observed characteristics receives the same score-dependent credit terms. The data therefore cannot reveal how the same borrower would have performed under a different credit allocation. The observed conditional mean combines borrower quality with the causal effect of financing.

This interpretation connects the theorem to Proposition 3 of Mendler-Dünner et al. (2022), which establishes the analogous identification failure in a structural causal model. The equivalence in Theorem 3 concerns an unrestricted structural class. Restrictions on the latent conditional mean or on the feedback technology can reduce the identified set by preventing the fundamental from absorbing arbitrary functions of the score.

The unrestricted identified set is large, but economically motivated restrictions can still deliver useful bounds. The next result considers a stable deployment when feedback is known to be positive, and its slope is bounded from above.

Proposition 1 (Partial identification under bounded positive feedback)**.** *Suppose the deployed score is performatively stable and the maintained class of feedback responses satisfies* $0 \le f'(t) \le \kappa^+ < 1$ *on the convex hull of the range of* m^**. For any* x_1, x_2 *with* $m^*(x_2) \ge m^*(x_1)$*,*

$$(1-\kappa^+)\big[m^*(x_2) - m^*(x_1)\big] \le u(x_2) - u(x_1) \le m^*(x_2) - m^*(x_1). \tag{28}$$

If $m^*(x)$ *has a finite second moment, then*

$$(1-\kappa^+)^2 \mathrm{Var}\big(m^*(x)\big) \le \mathrm{Var}\big(u(x)\big) \le \mathrm{Var}\big(m^*(x)\big). \tag{29}$$

Thus, stable deterministic deployment identifies the ranking of posterior fundamentals exactly and bounds their pairwise differences and variance under the stated derivative bound.

The bounds turn the failure of point identification into a quantitative identified set. If the maximal feedback gain is small, observed score differences are close to differences in posterior fundamentals. As the bound approaches one, the interval widens, and deterministic deployment becomes less informative about cardinal differences. Additional identifying variation is required to separate the fundamental and feedback components pointwise without such functional restrictions.

The identification designs below use variation in the score or allocation whose effect can be separated from changes in fundamentals. Cross-regime comparisons and randomized holdouts compare outcomes at the same feature value, holding the posterior mean $u(x)$ fixed. Model archives can provide such variation over time when the fundamental environment remains unchanged and archive disturbances are conditionally mean zero. Threshold designs instead compare units with scores approaching a cutoff from opposite sides.

These units need not have the same posterior mean. Identification requires the conditional mean of the fundamental given the score to be continuous at the cutoff, so that the jump in the conditional outcome mean isolates the discontinuity in feedback.

2.2 Identification across deterministic regimes

Consider first variation across deterministic deployment regimes. Suppose the same fundamental environment and feedback response are observed under scoring rules m_r, where $r \in \{1, \dots, R\}$. These rules need not be performatively stable. Under Assumption 1, Theorem 1 gives a unique stable predictor for a given u and f, so scoring rules that are all stable in this common environment coincide almost surely. Identifying variation can come from externally imposed scoring rules or from models deployed during adjustment toward the stable predictor. Write $g_r(x) = \mathbb{E}[y^{(m_r)} \mid x]$ for the conditional outcome mean in regime r. Let $\mathbb{X}_0 \subseteq \operatorname{supp}(P_X)$ be a common full-measure set on which the scoring rules are known, identified versions of all conditional means are fixed, and $g_r(x) = u(x) + f(m_r(x))$ holds in every regime. For any $x \in \mathbb{X}_0$ and regimes r, q, subtraction gives

$$g_r(x) - g_q(x) = f\big(m_r(x)\big) - f\big(m_q(x)\big). \tag{30}$$

The posterior mean cancels because both regimes are evaluated at the same feature value. Each comparison therefore identifies a difference in feedback between two observed scores. Combining comparisons can recover further differences and, when they connect all observed score values, identify the feedback response on that set after one level normalization.

Define the *deployment graph* $\mathcal{G}$ with vertex set $\mathcal{S} = \bigcup_{r=1}^{R} m_r(\mathbb{X}_0)$. Two scores $s, s' \in \mathcal{S}$ are joined if, for some $x \in \mathbb{X}_0$ and regimes r, q, $s = m_r(x)$ and $s' = m_q(x)$. Orienting the comparison from s' to s, its observed label is $g_r(x) - g_q(x)$. Under the common posterior mean and feedback response, equation (30) makes this label equal to $f(s) - f(s')$. Summing labels along a finite path identifies the feedback difference between its endpoints. The following theorem characterizes the resulting identification on the connected components of $\mathcal{G}$.

Theorem 4 (Identification from multiple deterministic regimes)**.** *Suppose $g_r(x) = u(x) + f(m_r(x))$ for every $x \in \mathbb{X}_0$ and every regime r, with the same u and f across regimes. Then:*

(i) Any two decompositions consistent with the conditional means have feedback responses that differ by a constant on each connected component of $\mathcal{G}$.

(ii) If $\mathcal{G}$ is connected, fixing one value $f(s_0)$ identifies f on $\mathcal{S}$ and identifies $u(x)$ for $x \in \mathbb{X}_0$ through

$$u(x) = g_r(x) - f\big(m_r(x)\big). \tag{31}$$

The right-hand side is the same in every regime.

(iii) Suppose $\mathcal{G}$ has finitely many connected components, each a Borel subset of $\mathbb{R}$. In the unrestricted class of measurable feedback responses, one additive constant per component remains free. If the admissible class imposes differentiability or derivative bounds on f, those restrictions may constrain the constants across components.

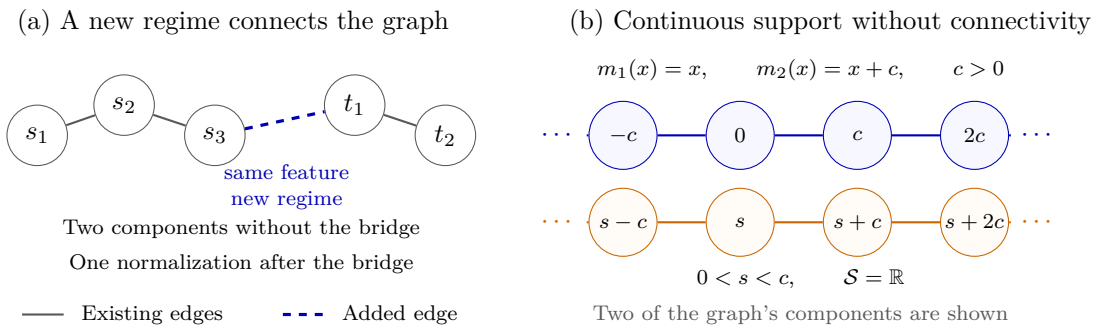


Figure 3: Deployment comparisons and graph connectivity. Each edge joins scores assigned to the same feature value under different regimes and identifies the corresponding difference in feedback. In panel (a), the solid edges form two connected components. The additional comparison shown by the dashed edge joins them, so one normalization identifies feedback at all five score values. In panel (b), the rules $m_1(x) = x$ and $m_2(x) = x + c$ on $\mathbb{R}$, with $c > 0$, generate a continuous score support but separate components of the form $\{s + jc : j \in \mathbb{Z}\}$. Two components are shown. Continuity of the score support alone does not imply connectivity of the deployment graph.

The theorem recovers feedback differences by adding observed labels along a path between two scores. Every path with the same endpoints must give the same difference. This yields restrictions that can be checked without choosing a level normalization. If several feature and regime pairs generate the same directed edge, their labels must agree. Labels must also sum to zero around any closed path.

Corollary 5 (Cycle restrictions and spanning trees)**.** *Under the conditions of Theorem 4, let $\ell(s, s')$ be the observed label on an edge directed from s' to s. Under the maintained model, $\ell(s, s') = f(s) - f(s')$, independently of the feature and regime pair that generates the edge. Then:*

(i) For every finite directed cycle $s_0, s_1, \ldots, s_L = s_0$,

$$\sum_{j=1}^{L} \ell(s_j, s_{j-1}) = 0. \tag{32}$$

(ii) The sum of edge labels is the same along any two paths with the same endpoints.

(iii) If $\mathcal{G}$ is finite with J vertices and connected, the labels on any spanning tree and one value $f(s_0)$ recover f on all J vertices. A graph with fewer than $J - 1$ distinct connecting edges is disconnected. Its edge labels and one normalization alone cannot recover feedback levels on every component. One normalization per component suffices for graph-based recovery.

Figure 3 illustrates how deployment variation connects feedback values. In panel (a), the solid edges provide comparisons within two groups of scores. Each comparison holds the feature value fixed, so the difference in conditional outcome means removes the posterior mean and identifies a difference in feedback. The dashed edge adds a comparison between the two groups. Once that difference is known, one normalization determines the feedback response at every vertex. Without the bridge, the observed edges do not determine the relative feedback levels of the two components. Additional restrictions on the response may constrain those levels.

Panel (b) shows why a continuous score support does not ensure graph connectivity. Suppose the common identified feature set is $\mathbb{X}_0 = \mathbb{R}$ and the two scoring rules are $m_1(x) = x$ and $m_2(x) = x + c$, where $c > 0$. The observed contrast is $g_2(x) - g_1(x) = f(x + c) - f(x)$. Each edge therefore joins scores that differ by c, and every finite path changes the score by an integer multiple of c. Conversely, repeated edges connect any two scores with such a difference. The connected components are thus the sets $\{s + jc : j \in \mathbb{Z}\}$. Although the observed score support is all of $\mathbb{R}$, these comparisons do not link scores in different components.

Connectivity can nevertheless arise with only two regimes. Suppose $x = (x_1, x_2)$, the common identified feature set is $\mathbb{X}_0 = I \times I$ for an interval I, and the scoring rules are $m_1(x) = x_1$ and $m_2(x) = x_2$. For any scores $s, s' \in I$, the feature value $x = (s, s')$ receives score s in the first regime and s' in the second. Every pair of score values is therefore connected by an observed comparison. One normalization then identifies f throughout I. Identification depends on which score values the deployment rules connect.

The graph formulation separates deterministic policy variation from randomization. A policy reform, staggered rule change, or cross-market difference can identify the feedback response whenever it links previously disconnected score values while leaving the primitive fundamental environment unchanged. Corollary 5 also turns the design into a specification test. Nonzero cycle sums reject a common feedback function or a regime-invariant fundamental environment. When candidate comparisons can be selected separately at additive costs, a minimum-cost spanning tree supplies the cheapest set of connecting comparisons sufficient for point identification up to one normalization. A randomized holdout provides an experimental counterpart to these regime comparisons. Its common baseline allocation connects deployed scores to a single reference point. The next subsection develops this design and then considers threshold allocations. Section 3 studies the identifying variation supplied by model archives under restrictions on estimation disturbances.

2.3 Randomized holdouts and threshold allocations

Randomized holdouts create allocation variation among units with the same feature value. A lender can continue to score every applicant while assigning a randomly selected group to a fixed baseline credit policy. The outcomes of this group reveal the conditional mean of fundamentals under a common allocation, while the difference between the deployed and baseline groups reveals the effect of the score-dependent policy.

Proposition 2 (Holdout identification)**.** *Suppose the scoring rule m and baseline score s_0 are known. Let $D \in \{0, 1\}$ be independent of (θ, x, ε), with $\mathbb{P}(D = 1) = h \in (0, 1)$. Units with $D = 1$ are scored but receive the fixed baseline allocation associated with s_0, so their outcome is*

$$y = \theta + f(s_0) + \varepsilon. \tag{33}$$

Units with $D = 0$ receive the score-dependent allocation in (3). Let $\mathcal{S}_m$ denote the support of $m(x)$. The joint law of (x, y, D) identifies the following objects.

(i) For P_X-almost every x,

$$\mathbb{E}[y \mid x, D = 1] = u(x) + f(s_0). \tag{34}$$

Thus $u(x)$ is identified almost surely once the level $f(s_0)$ is normalized.

(ii) For P_X-almost every x,

$$\mathbb{E}[y \mid x, D = 0] - \mathbb{E}[y \mid x, D = 1] = f(m(x)) - f(s_0). \tag{35}$$

If f is continuous, the function $s \mapsto f(s) - f(s_0)$ is identified at every $s \in \mathcal{S}_m$. One normalization of $f(s_0)$ then identifies both f on $\mathcal{S}_m$ and $u(x)$ almost surely.

(iii) Suppose f is differentiable and $\mathcal{S}_m$ contains an interval I. The local loop gain $k(s) = f'(s)$ is identified at every interior point of I. If $\mathbb{P}(m(x) \in \operatorname{int} I) = 1$ and $\mathbb{E}|f'(m(x))| < \infty$, the deployment-average gain $\overline{k} = \mathbb{E}[f'(m(x))]$ is also identified. These support conditions hold, for example, if $m(x)$ has an absolutely continuous distribution supported on I with a density that is positive almost everywhere on I.

The holdout design randomizes the allocation while preserving the score. Every unit can therefore be evaluated by the model, and the comparison between the deployed and baseline groups is available throughout the support of the feature. This differs from designs that randomize whether the model is evaluated at all, which would remove the score for some units and weaken the comparison across feature values.

The additive constant reflects the absence of an observed allocation with a known zero effect. If the baseline allocation satisfies $f(s_0) = 0$, then both the posterior mean and the feedback response are identified in levels. In a lending application, this normalization may correspond to a benchmark financing policy whose effect on repayment is normalized to zero. Without such a benchmark, the data identify how outcomes change with the allocation while leaving the common level of the fundamental and feedback components undetermined.

Randomized intervention-free holdouts are studied by Liley et al. (2021) and Haidar-Wehbe et al. (2025) as a method for preventing intervention-induced bias in predictive model updating. Munro (2025) studies related treatment allocation problems when agents respond strategically. Proposition 2 characterizes the structural objects recovered by such a design in the present environment. Livdan et al. (2026) study permanent and rotating holdouts in a dynamic allocation environment in which current outcomes depend on capital accumulated under earlier deployments. The two designs then recover different components of the feedback.

Randomization may be costly or institutionally infeasible. Many economic allocation rules nevertheless contain discrete changes that create local variation. Credit approval, insurance eligibility, regulatory classification, and platform access often change discontinuously when a score crosses a known threshold. A discontinuity in the conditional outcome mean identifies the feedback jump when the score-conditional fundamental mean is continuous at the cutoff. The following result also recovers posterior fundamentals under a specification in which feedback changes only through that jump.

Proposition 3 (Threshold allocations)**.** *Retain the outcome equation (3) and the setup moment conditions, but replace Assumption 1 with the following threshold specification. Suppose the known scoring rule m determines a binary allocation at a known cutoff c, and the feedback response belongs to the class*

$$f_\Delta(s) = a + \Delta \mathbb{1}\{s \geq c\}, \tag{36}$$

where a and Δ are unknown constants. Assume that m is continuous, that $m(x)$ has a density that is positive in a neighborhood of c, and that the score-conditional fundamental mean

$$r(s) := \mathbb{E}[\theta \mid m(x) = s] \tag{37}$$

has a version continuous at c. Then deployment data identify the jump through

$$\Delta = \lim_{s \downarrow c} \mathbb{E}[y \mid m(x) = s] - \lim_{s \uparrow c} \mathbb{E}[y \mid m(x) = s]\,, \tag{38}$$

where the limits are taken using the conditional-mean version $r(s) + f_\Delta(s)$. Once Δ is known, $\mathbb{E}[y \mid x] - \Delta \mathbb{1}\{m(x) \geq c\} = u(x) + a$ for almost every feature value x. Thus $u(x)$ is identified almost surely up to the additive constant a.

The identifying restriction concerns fundamentals conditional on the score. Smoothness of the primitives in feature space alone does not ensure this continuity, because the composition of units with scores near the cutoff can change discontinuously. A jump in r would therefore enter the observed outcome contrast alongside the allocation effect. Under the maintained continuity condition, threshold variation separates the discrete feedback effect from selection on fundamentals. Within the specified threshold class, subtracting this identified effect from the feature-conditional outcome mean then recovers $u(x) + a$.

The restriction also has a clear economic interpretation. Suppose a credit score determines approval at a cutoff but also affects the interest rate smoothly on either side of the cutoff. The discontinuity identifies the discrete effect of approval under the continuity condition, while the smooth financing effect remains observationally equivalent to a smooth component of borrower quality. A threshold design therefore recovers the entire feedback response only when the allocation technology is known to contain no additional score-dependent component.

Figure 4 illustrates the identification argument in Proposition 3. The observed outcome mean satisfies $\mathbb{E}[y \mid m(x) = s] = r(s) + a + \Delta \mathbb{1}\{s \geq c\}$. The figure normalizes $a = 0$, so the observed mean coincides with the fundamental mean below the cutoff and lies Δ above it once the allocation is activated. The upward slope of the dashed curve allows fundamentals to vary systematically with the score. Continuity of r at c ensures that this variation contributes no jump at the cutoff, leaving the observed discontinuity to identify Δ.

The two designs rely on different sources of identifying variation. The holdout changes the allocation by random assignment and identifies feedback differences without specifying the shape of the response. The threshold design uses a known discontinuity in the allocation rule and requires continuity of the score-conditional fundamental mean at the cutoff. This comparison identifies the feedback jump even if the response also varies continuously with the score. Recovering the entire response up to a constant, and hence recovering $u(x)$ up to that constant, additionally requires the restriction in Proposition 3 that feedback changes only through the discrete jump.

Together, Theorems 3 and 4 and Propositions 2 and 3 give a common identification principle. Identification uses variation in the score or allocation whose outcome effect can be separated from changes in fundamentals. Multiple regimes and randomized holdouts compare outcomes at the same feature value, holding the posterior mean fixed. Threshold designs instead compare scores approaching the cutoff from

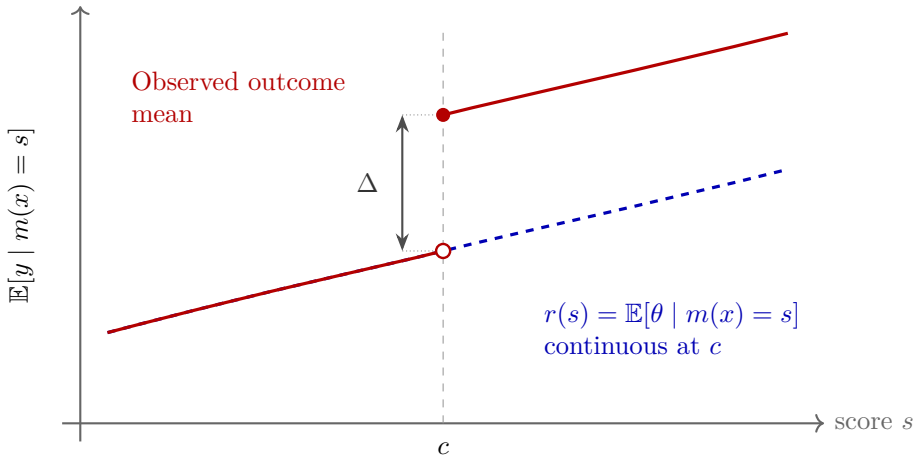


Figure 4: Identification at a threshold allocation. The figure normalizes the baseline response to $a = 0$. The dashed curve is the score-conditional fundamental mean $r(s)$. The solid curve is the observed outcome mean, which equals $r(s)$ below c and $r(s) + \Delta$ at and above c. The open point marks the left-hand limit of the observed mean, while the filled point gives its value at the cutoff, where the allocation is activated. Continuity of r at c ensures that the observed jump identifies Δ. A discontinuity in r would add a selection effect to the outcome contrast.

opposite sides. The units need not have the same posterior mean, but continuity of the score-conditional fundamental mean ensures that it contributes no jump. Section 3.1 develops another source of variation through model vintages evaluated at fixed feature values, under restrictions on archive disturbances. Each design supplies comparisons or restrictions that separate feedback from the fundamental component left unidentified by an unrestricted single deployment.

The next subsection uses an identified decomposition to evaluate a specified change in the allocation rule.

2.4 Counterfactual transport across deployment rules

Identification permits predictions under allocation rules that have not yet generated training data. Hold the joint distribution of fundamentals, features, and outcome shocks fixed, and suppose the posterior mean u has been identified after a level normalization. Consider a new allocation rule with feedback response $\widetilde{f}$. The counterfactual calculation requires this response to be known on the score values reached under the new rule. Identification of the original response on the observed score support does not by itself supply that knowledge, particularly when the new rule moves scores outside that support. The following proposition characterizes the stable predictor conditional on a specified counterfactual response.

Proposition 4 (Counterfactual transport)**.** *Suppose u is known and the specified counterfactual response $\widetilde{f}$ is differentiable with $\widetilde{f}'(s) \leq \widetilde{\kappa}^+ < 1$ for every s. Define $\widetilde{h}(s) = s - \widetilde{f}(s)$. Then the counterfactual regime has the unique stable predictor, up to almost-sure equality,*

$$\widetilde{m}^*(x) = \widetilde{h}^{-1}(u(x))\,. \tag{39}$$

If m^* is the stable predictor under feedback response f, the following comparative statements hold.

(i) If $\widetilde{f}(s) \geq f(s)$ for every s, then $\widetilde{m}^*(x) \geq m^*(x)$ for every x. The inequality reverses when $\widetilde{f}(s) \leq f(s)$ everywhere.

(ii) For every x, the change in the stable prediction satisfies $|\widetilde{m}^*(x) - m^*(x)| \leq |\widetilde{f}(m^*(x)) - f(m^*(x))|/(1-\widetilde{\kappa}^+)$. Consequently, if $\sup_s |\widetilde{f}(s) - f(s)| \leq d < \infty$, then

$$\sup_x |\widetilde{m}^*(x) - m^*(x)| \leq \frac{d}{1-\widetilde{\kappa}^+}. \tag{40}$$

A policy that raises the outcome response at every score also raises the stable prediction for each feature value. The bound in (40) controls the size of this change under the same one-sided slope restriction that ensures existence and uniqueness. As the upper slope bound $\widetilde{\kappa}^+$ approaches one from below, even a modest change in the response can permit a large change in stable predictions. The bound does not require contraction of the retraining operator.

Figure 5 illustrates a case in which the bound is attained. In panel (a), the posterior mean remains fixed at $u = 1$, while the response changes from $f(s) = 0.5s$ to $\widetilde{f}(s) = 0.5s + 0.5$. The stable prediction increases from 2 to 3. The increase exceeds the direct policy effect because the higher prediction induces an additional outcome response under the new rule. Panel (b) varies the common slope k. Since $\widetilde{\kappa}^+ = k$, the stable-score change $d/(1-k)$ attains the bound in Proposition 4(ii) and rises with positive feedback. For affine responses with $k < 0$, the same formula instead gives a multiplier below one, so feedback attenuates an additive policy change.

Applying the proposition also requires a consistent level normalization. A connected deployment graph and randomized holdouts identify the posterior mean and feedback response up to a common additive constant. Moving this constant between fundamentals and feedback changes their separate levels. In Figure 5, replacing u by $u - c$ and both feedback responses by their original values plus c leaves the conditional-mean schedules unchanged. Their intersections with the 45-degree line therefore remain fixed. The following corollary establishes this invariance for the general counterfactual calculation.

Corollary 6 (Normalization invariance of counterfactual transport)**.** *Under the conditions of Proposition 4, fix any constant $c \in \mathbb{R}$. Replace the identified decomposition by $u_c(x) = u(x) - c$ and $f_c(s) = f(s) + c$, and express the counterfactual response as $\widetilde{f}_c(s) = \widetilde{f}(s) + c$. Then the counterfactual stable predictor is unchanged. In particular, the common additive normalization left by Theorem 4(ii) or Proposition 2 has no effect on transported stable predictions when the factual and counterfactual responses use the same normalization.*

The identification results therefore support a policy calculation once the new feedback response has been specified. In a lending application, the lender combines the same posterior repayment capacity with the outcome response under new financing terms and solves the resulting fixed-point equation. This calculation can be performed before outcomes under the new policy are observed. Its substantive requirements are that posterior fundamentals remain unchanged and that the counterfactual response is known over the relevant score range.

We next turn from identification to the stochastic information contained in sequences of model vintages.

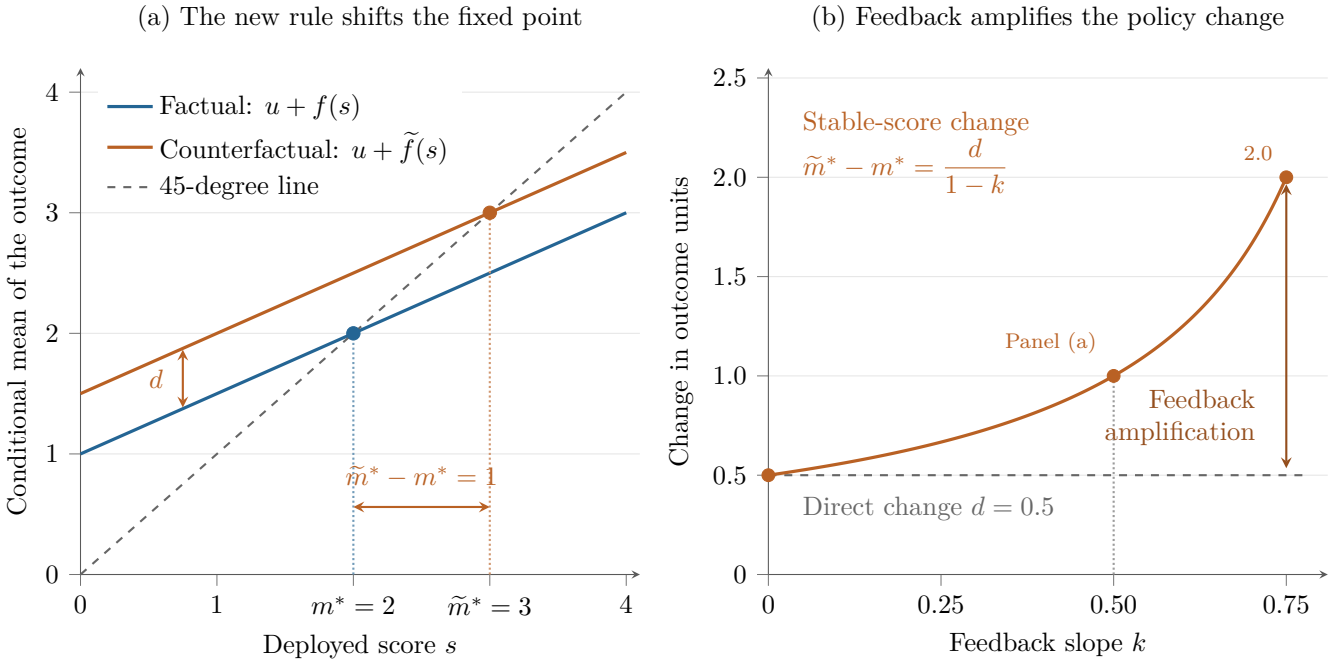


Figure 5: Counterfactual transport under a specified change in the allocation rule. The factual response is $f(s) = ks$, and the counterfactual response is $\widetilde{f}(s) = ks + d$, with $d = 0.5$. Posterior fundamentals remain fixed. Panel (a) sets $u = 1$ and $k = 0.5$. The intersections of the conditional-mean schedules with the 45-degree line determine the stable predictions. Raising the response by 0.5 at every score moves the stable prediction from 2 to 3. Panel (b) varies k and compares the stable-score change $\widetilde{m}^* - m^* = d/(1-k)$ with the direct response change d. Taking the upper slope bound to be $\widetilde{\kappa}^+ = k$, the bound in Proposition 4(ii) holds with equality. A common level normalization leaves both conditional-mean schedules, and hence their fixed points, unchanged.

3 Finite-Sample Retraining and Model Archives

Population retraining suppresses estimation noise. In practice, each model vintage is estimated from a finite sample, so estimation error changes the deployed rule and thereby changes the data used by later vintages. We first remain in the general feedback environment and characterize how small model-level perturbations propagate around the stable predictor. We then specialize to a scalar linear-Gaussian benchmark only where linearity buys an exact finite-sample stochastic law.

3.1 General archive dynamics

The identification results in Section 2 use variation in the score or allocation whose outcome effect can be separated from changes in fundamentals. Model archives provide a way to study this variation at fixed feature values. When the fundamental environment remains unchanged, evaluating successive model vintages at the same feature vector holds the posterior mean $u(x)$ fixed while predictions vary. Finite-sample estimation can generate these movements even when the underlying population retraining operator is unchanged. Once deployed, a prediction error can change the allocation and the subsequent outcome, entering the data used to estimate the next model. Sampling variation can therefore create identifying variation in deployment, provided restrictions on estimation disturbances allow its outcome effect to be

separated from conditional estimation bias.

This observation raises two related questions. First, how does a perturbation in one model vintage propagate into the next? Second, what features of the feedback response can be recovered from the resulting sequence of models? The first question concerns the population retraining map and does not require a particular sampling model. Because $(Tm)(x) = u(x) + f(m(x))$, a perturbation of the deployed prediction affects the next population update through the feedback response. We consider perturbations that keep the deployed outcome square integrable, so the population regression remains well defined. Around a stable predictor, the first-order effect of such a perturbation is governed by the local slope $f'(m^*(x))$. The subsequent archive results specify the additional conditions under which observed movements across vintages identify this response.

Proposition 5 (Local retraining persistence)**.** *Under Assumption 1, let m^* denote the stable predictor. Let $q\colon \mathbb{X} \to \mathbb{R}$ be measurable and suppose that $f(m^*(x) + \epsilon q(x))$ has a finite second moment for every ϵ in a neighborhood of zero. Then*

$$\lim_{\epsilon \to 0} \frac{T(m^* + \epsilon q)(x) - Tm^*(x)}{\epsilon} = f'(m^*(x))\, q(x). \tag{41}$$

If $f(t) = a + kt$ is affine, the same integrability condition ensures that $m^* + q$ is admissible and

$$T(m^* + q) - Tm^* = kq. \tag{42}$$

For affine feedback, every square-integrable perturbation q satisfies the integrability condition. If $\kappa_{\text{abs}} < 1$ and q is bounded, the condition also holds automatically, and the function on the right-hand side of (41) has a uniform norm at most $\kappa_{\text{abs}} \|q\|_\infty$.

Proposition 5 connects the population fixed point to propagation of errors across model vintages. A small prediction error at score level $m^*(x)$ enters the next population update in proportion to the local feedback gain $f'(m^*(x))$. This derivative measures how strongly the current prediction changes the outcome used for subsequent retraining. Interpreting observed archive persistence through this channel additionally requires restrictions on estimation disturbances, which the following results make explicit. In the affine economy, the same scalar k governs the fixed point, the propagation of prediction errors, and, under the sampling assumptions of Theorem 6, the autoregressive persistence of estimated coefficients.

The same-information logic of Section 2 has an even stronger archive implication. Local persistence uses only infinitesimal variation around a stable prediction. When archive movements have richer support and their estimation error is conditionally mean zero, feedback differences can be recovered on the archive's score support. We turn to that result next.

Theorem 5 (Exact identification from nonlinear model archives)**.** *Fix a feature value x and an archive date t. Let $\{\mathcal{F}_t\}$ be a filtration containing the archive history through date t. Suppose*

$$M_{t+1}(x) = u(x) + f(M_t(x)) + \xi_{t+1}(x), \qquad \mathbb{E}[\xi_{t+1}(x) \mid \mathcal{F}_t] = 0, \tag{43}$$

where $M_t(x)$ is $\mathcal{F}_t$-measurable and $\mathbb{E}|M_{t+1}(x)| + \mathbb{E}|\xi_{t+1}(x)| < \infty$. Suppose that the joint distribution of $(M_t(x), M_{t+1}(x))$ is identified. Let $\mathcal{S}_x$ be the support of $M_t(x)$, and let G_x be a version of $\mathbb{E}[M_{t+1}(x) \mid$

$M_t(x) = s]$. *Then:*

(i) For $P_{M_t(x)}$-almost every s,

$$G_x(s) = u(x) + f(s). \tag{44}$$

(ii) For $P_{M_t(x)} \otimes P_{M_t(x)}$-almost every pair (s, s_0),

$$G_x(s) - G_x(s_0) = f(s) - f(s_0). \tag{45}$$

If G_x and f are continuous on $\mathcal{S}_x$, the equality holds for every $s, s_0 \in \mathcal{S}_x$.

(iii) Under the continuity condition in part (ii), if s is an interior point of $\mathcal{S}_x$ and f is differentiable at s, then $G'_x(s) = f'(s)$.

(iv) Consider two feature values x and x' for which the transition means have continuous versions on their respective supports, and suppose that f is continuous on those supports. For any $s, s_0 \in \mathcal{S}_x \cap \mathcal{S}_{x'}$,

$$G_x(s) - G_x(s_0) = G_{x'}(s) - G_{x'}(s_0). \tag{46}$$

Therefore, a common scalar feedback response imposes an overidentifying restriction across feature values.

A fixed-x archive can be constructed even when the same economic unit never appears twice. An institution can retain model vintages $m_1, \ldots, m_T$, choose a frozen audit panel of feature vectors $x_1, \ldots, x_L$ in the relevant feature support, and evaluate every vintage on every audit vector. This yields the panel $M_t(x_\ell) = m_t(x_\ell)$. Changes in the fitted model rather than repeated observations of the same borrower, product, or patient result in variation at a fixed feature value.

The theorem identifies properties of the population transition mean. Estimating one transition mean from a sequence of saved vintages also requires a common transition rule across the dates being pooled, suitable time-series conditions such as stationarity and ergodicity, and variation in the archived scores over the support of interest. A sequence concentrated at one score cannot identify feedback differences elsewhere.

Figure 6 illustrates the identification argument in Theorem 5. Panel (a) compares archive transition means at two fixed audit feature values. The horizontal coordinate is the current archived prediction, and the vertical coordinate is the expected prediction in the next vintage. Under the conditional mean-zero assumption, this transition mean equals $u(x) + f(s)$. Posterior fundamentals shift the level of the curve, while feedback determines how it varies with the current score. Different audit feature values can have different posterior means, but a common feedback response requires their transition curves to differ by a constant wherever their archive supports overlap.

Panel (b) removes these level differences by subtracting each curve's value at the same reference score s_0. Both curves then coincide with $f(s) - f(s_0)$. Archive variation thus identifies changes in the feedback response without first identifying the posterior mean. Fixing one level $f(s_0)$ then identifies f on this common score interval and recovers each posterior mean from $u(x) = G_x(s) - f(s)$. The coincidence of the normalized curves also gives an overidentifying restriction. Systematic differences between them on

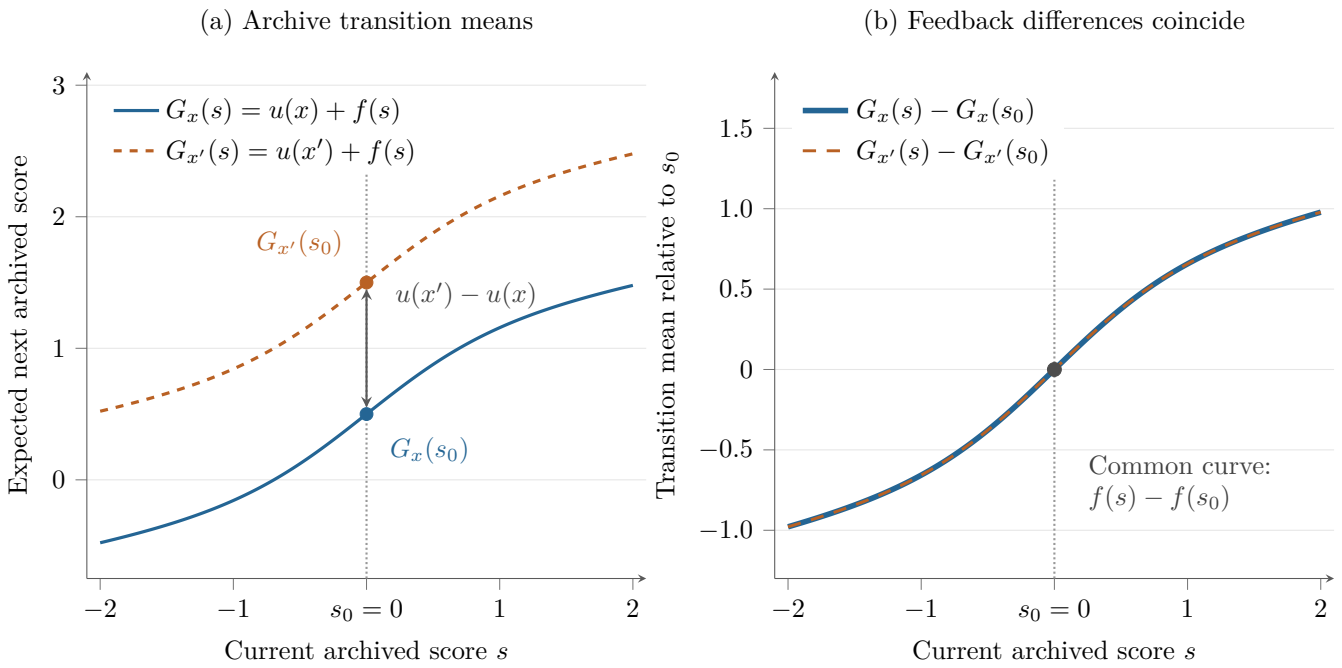


Figure 6: Identification from nonlinear model archives. Panel (a) plots the archive transition means $G_x(s) = \mathbb{E}[M_{t+1}(x) \mid M_t(x) = s]$ for two fixed audit feature values. Under conditionally mean-zero archive disturbances, each transition mean equals $u(x) + f(s)$, so the curves differ only by the constant $u(x') - u(x)$. Panel (b) subtracts each curve's value at the common reference score s_0. The curves coincide with $f(s) - f(s_0)$ on their common archive support. The illustration uses $f(s) = 0.2s + 0.6\tanh(s)$, $u(x) = 0.5$, $u(x') = 1.5$, and $s_0 = 0$. The displayed score interval is assumed to lie in both archive supports.

common support reject the joint benchmark of a common scalar response and conditionally mean-zero archive disturbances.

The conditional mean-zero restriction is significant. The next proposition allows the archive disturbance to have a score-dependent conditional mean and bounds the resulting error in identified feedback differences. Define the score-dependent conditional bias by $b_x(s) = \mathbb{E}[\xi_{t+1}(x) \mid M_t(x) = s]$, and let $B_x(s, s_0) \geq 0$ be a specified bound on $|b_x(s) - b_x(s_0)|$.

Proposition 6 (Archive identification with score-dependent bias)**.** *Suppose* $M_{t+1}(x) = u(x) + f(M_t(x)) + \xi_{t+1}(x)$ *and* $\mathbb{E}|M_{t+1}(x)| + \mathbb{E}|\xi_{t+1}(x)| < \infty$. *Then* $G_x(s) = u(x) + f(s) + b_x(s)$ *for* $P_{M_t(x)}$*-almost every* s. *Consequently, for* $P_{M_t(x)} \otimes P_{M_t(x)}$*-almost every pair* (s, s_0),

$$G_x(s) - G_x(s_0) = f(s) - f(s_0) + b_x(s) - b_x(s_0). \tag{47}$$

At any such pair for which $|b_x(s) - b_x(s_0)| \leq B_x(s, s_0)$,

$$G_x(s) - G_x(s_0) - B_x(s, s_0) \leq f(s) - f(s_0) \leq G_x(s) - G_x(s_0) + B_x(s, s_0). \tag{48}$$

If b_x *is constant* $P_{M_t(x)}$*-almost surely, then* $G_x(s) - G_x(s_0) = f(s) - f(s_0)$ *for almost every pair.*

If G_x *and* b_x *have continuous versions on* S_x *and* f *is continuous there, the equalities hold for every pair in* S_x^2. *The bounds then hold at every pair in* S_x^2 *for which the stated bound on conditional bias holds.*

The proposition turns the strongest assumption behind archive identification into a sensitivity object. Score-dependent estimation bias is the only term that contaminates the transition difference. Constant bias shifts the transition function in levels but leaves feedback differences identified, while a bound on changes in bias translates directly into a bound on the causal feedback difference.

The theorem does not require the archive noise to be small. It requires the stronger economic condition that the estimation disturbance be conditionally mean zero given the deployed prediction and the past. Under that condition, an auditor can run every stored model vintage on the same audit feature vector x, estimate the transition function G_x, and recover nonlinear feedback differences without observing θ. The archive therefore creates the same-information variation that is absent from a single deterministic deployment.

Corollary 7 (Exact recovery from overlapping archive supports)**.** *Fix an archive transition date. Suppose the conditions of Theorem 5 hold for each audit feature value considered, f is continuous on the relevant score supports, and each transition mean G_x has a continuous version on $\mathcal{S}_x$. Suppose $\mathcal{S}_{x_0}$ contains an interval I, and fix $s_0 \in I$. Then, for every $s \in \mathcal{S}_{x_0}$, $f(s) - f(s_0) = G_{x_0}(s) - G_{x_0}(s_0)$. One normalization $f(s_0)$ identifies f on $\mathcal{S}_{x_0}$, including I.*

If $\mathcal{S}_x$ intersects a score set on which f has been identified, then $u(x) = G_x(s) - f(s)$ at any point s of intersection. The same identity identifies f throughout $\mathcal{S}_x$. Repeating this argument propagates the normalization along any finite chain of audit feature values whose successive archive supports intersect.

A simpler diagnostic is available when the archive fluctuates only slightly around a stable prediction. In that case, persistence itself converges to the local feedback derivative.

Corollary 8 (Local persistence under small archive noise)**.** *Suppose $\overline{\kappa}_{\text{abs}} < 1$ and fix $x \in \mathbb{X}$. Let $m^*(x)$ be the stable score and define $a(x) = f'(m^*(x))$. Let $\{\xi_t\}_{t\in\mathbb{Z}}$ be i.i.d. with $\mathbb{E}[\xi_t] = 0$ and $0 < \text{Var}(\xi_t) = \sigma_\xi^2 < \infty$. For each $\tau > 0$, consider*

$$M_{t+1}^\tau(x) = u(x) + f(M_t^\tau(x)) + \tau\xi_{t+1}. \tag{49}$$

Then equation (49) has a unique causal stationary solution with finite second moment. With $Z_t^\tau(x) = [M_t^\tau(x) - m^*(x)]/\tau$, the stationary process satisfies $Z_t^\tau(x) \to Z_t(x)$ in L^2 for every fixed t as $\tau \downarrow 0$, where

$$Z_{t+1}(x) = a(x)Z_t(x) + \xi_{t+1} \tag{50}$$

is the unique stationary AR(1) solution. Consequently,

$$\lim_{\tau\downarrow 0} \text{Corr}(M_{t+1}^\tau(x), M_t^\tau(x)) = f'(m^*(x)). \tag{51}$$

The exact transition result and the small-noise result serve different purposes. Theorem 5 identifies nonlinear feedback differences from sufficiently rich exogenous archive movement. Corollary 8 shows why a simple persistence coefficient recovers the local derivative when the archive explores only a small neighborhood of the stable score.

At the level of the entire prediction function, the same local argument yields a covariance-operator restriction. The Internet Appendix IA.1 proves that the small-noise limit satisfies $C_1 = AC_0$, where $Aq(x) = f'(m^*(x))q(x)$. Because covariance-operator inversion is generally ill-conditioned in infinite

dimensions, the appendix also gives a finite-dimensional projection result and an explicit sensitivity bound showing that archive directions with little variance are weakly identified. The main text keeps the exact transition-function result as the operational nonlinear identification theorem.

Estimated retraining. The same contraction logic bounds the propagation of estimation error outside the linear-Gaussian benchmark. Let $\widehat{m}_t$ be the predictor deployed in period t and suppose its estimated update satisfies $\sup_{x\in\mathbb{X}} |\widehat{m}_{t+1}(x) - (T\widehat{m}_t)(x)| \le \overline{e}_t$. Assume

$$\overline{\kappa}_{\text{abs}} = \sup_{t\in\mathbb{R}} |f'(t)| < 1. \tag{52}$$

Proposition 7 (Amplification of estimation error)**.** *Suppose* (52) *holds and* $\sup_{x\in\mathbb{X}} |\widehat{m}_0(x) - m^*(x)| < \infty$. *Then, for every* $t \ge 1$,

$$\sup_{x\in\mathbb{X}} |\widehat{m}_t(x) - m^*(x)| \le \overline{\kappa}_{\text{abs}}^t \sup_{x\in\mathbb{X}} |\widehat{m}_0(x) - m^*(x)| + \sum_{j=0}^{t-1} \overline{\kappa}_{\text{abs}}^{t-1-j} \overline{e}_j. \tag{53}$$

If $\overline{e}_t \le \overline{e}$ *for every* t, *then*

$$\limsup_{t\to\infty} \sup_{x\in\mathbb{X}} |\widehat{m}_t(x) - m^*(x)| \le \frac{\overline{e}}{1 - \overline{\kappa}_{\text{abs}}}. \tag{54}$$

The proposition separates a general amplification mechanism from the distribution-specific results below. Feedback close to the contraction boundary makes any persistent estimation error decay slowly and can magnify its long-run worst-case effect by the factor $1/(1 - \overline{\kappa}_{\text{abs}})$. This bound concerns uniform sensitivity. It does not imply a particular variance or tail law. The exact linear-Gaussian archive analysis supplies the additional stochastic structure needed for those stronger conclusions.

3.2 Scalar benchmark for archive dynamics

The preceding results identify feedback from archive transitions and describe dynamics near a stable predictor. We now impose linearity and Gaussian sampling to obtain an exact law for the entire archive at every noise scale. This specialization connects the population retraining map to the stochastic sequence of estimated models. The companion paper develops the linear-Gaussian economic benchmark in detail. Here the specialization makes the propagation of feedback and the contribution of sampling error explicit.

Let $\theta = \mu + \widetilde{\theta}$, where $\widetilde{\theta} \sim N(0, \sigma_\theta^2)$, and let $x = \theta + \eta$, where $\eta \sim N(0, \sigma_\eta^2)$ is independent of $\widetilde{\theta}$. Define $V := \sigma_\theta^2 + \sigma_\eta^2 > 0$, $\rho := \sigma_\theta^2/V$, and $\widetilde{x} := x - \mu$. Let the feedback response be affine, $f(s) = ks + \beta a_0$, with $k < 1$. The posterior mean is $u(x) = \mu + \rho\widetilde{x}$, and population retraining transforms the slope according to $T(\lambda) = \rho + k\lambda$. The unique stable slope is $\lambda^* = \rho/(1-k)$.

This scalar benchmark is the affine specialization of the general structure theorem. Its stable score is

$$m^*(x) = \frac{\mu + \beta a_0 + \rho\widetilde{x}}{1-k}. \tag{55}$$

Its variance is

$$\text{Var}(m^*(x)) = \frac{\rho^2 V}{(1-k)^2}. \tag{56}$$

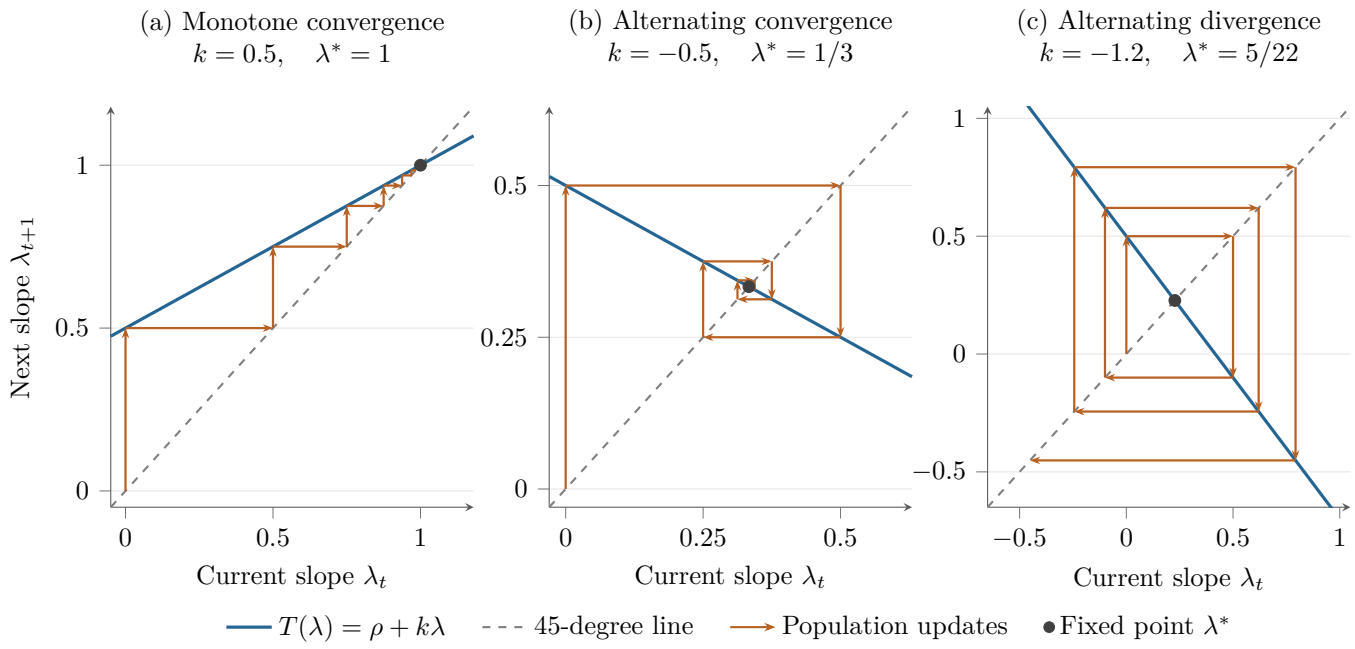


Figure 7: Stable predictors and population retraining. Each panel sets $\rho = 0.5$ and starts from $\lambda_0 = 0$. The blue line is the population slope map $T(\lambda) = \rho + k\lambda$, and its intersection with the dashed 45-degree line gives $\lambda^* = \rho/(1-k)$. The orange paths trace six updates. Each vertical segment applies the retraining map, and each horizontal segment carries the resulting slope to the next update. Positive feedback with $k = 0.5$ gives monotone convergence. Negative feedback with $k = -0.5$ gives alternating convergence. At $k = -1.2$, the stable slope still exists, but departures from it alternate and expand. Axis ranges differ across panels. Sampling error is absent from these population paths.

The score is a strictly increasing affine transformation of the posterior mean, so it preserves its rankings and information. For $\sigma_\theta^2 > 0$, it also satisfies $\mathrm{Corr}(m^*(x), \theta)^2 = \rho$ for every $k < 1$. These identities describe the population fixed point.

Convergence to this fixed point depends on the absolute feedback gain. Subtracting $\lambda^* = \rho + k\lambda^*$ from the population update gives $\lambda_{t+1} - \lambda^* = k(\lambda_t - \lambda^*)$. Each update therefore multiplies the previous deviation by k. Figure 7 illustrates the resulting paths with $\rho = 0.5$ and $\lambda_0 = 0$. At $k = 0.5$, the slope approaches $\lambda^* = 1$ monotonically. At $k = -0.5$, deviations alternate in sign and shrink toward $\lambda^* = 1/3$. At $k = -1.2$, the stable slope remains well defined at $\lambda^* = 5/22$, but every nonzero deviation alternates in sign and grows. Thus the maintained bound $k < 1$ guarantees a unique stable predictor, while convergence of population retraining from every initial slope requires $|k| < 1$.

Sampling error adds a new estimation disturbance to each update of this population map. Even when $|k| < 1$, fresh sampling can sustain fluctuations around the population fixed point. The finite-sample recursion derived below is algebraically well-defined for every real k. We use this extension to characterize both stationary archive dynamics and behavior outside the contraction region.

3.3 Setting

In period t, the modeler deploys $s_{i,t} = \bar{s}_t + \lambda_t \widetilde{x}_{i,t}$ to a fresh cross-section of $n \geq 2$ units. Across units and dates, $\widetilde{\theta}_{i,t} \sim N(0, \sigma_\theta^2)$, $\eta_{i,t} \sim N(0, \sigma_\eta^2)$, and $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2)$ are mutually independent, $\widetilde{x}_{i,t} = \widetilde{\theta}_{i,t} + \eta_{i,t}$,

$V = \sigma_\theta^2 + \sigma_\eta^2$, and $\rho = \sigma_\theta^2/V$. We allow $k \in \mathbb{R}$ and maintain $V > 0$, $\sigma_\theta^2 > 0$, and nondegenerate regression residual variance. Positive k describes reinforcing feedback and negative k describes offsetting feedback.

Under deployed coefficient λ_t, the realized outcome is

$$y_{i,t} = \overline{y}_t + (1 + k\lambda_t)\widetilde{\theta}_{i,t} + k\lambda_t\eta_{i,t} + \varepsilon_{i,t}. \tag{57}$$

The modeler estimates OLS with an intercept and deploys the estimated slope next period. With $\overline{x}_t = n^{-1}\sum_i \widetilde{x}_{i,t}$ and $\overline{y}_t^{\text{obs}} = n^{-1}\sum_i y_{i,t}$,

$$\lambda_{t+1} = \widehat{\lambda}_t^{\text{OLS}} = \frac{\sum_{i=1}^n (\widetilde{x}_{i,t} - \overline{x}_t)(y_{i,t} - \overline{y}_t^{\text{obs}})}{\sum_{i=1}^n (\widetilde{x}_{i,t} - \overline{x}_t)^2}. \tag{58}$$

The initial coefficient is nonanticipative, meaning it is independent of subsequent cross-sections.

Two identities drive the analysis. The population slope is $\rho + k\lambda$. The corresponding regression residual is

$$u_{i,t} = y_{i,t} - \mathbb{E}[y_{i,t}] - (\rho + k\lambda)\widetilde{x}_{i,t}, \tag{59}$$

and substitution gives

$$u_{i,t} = (1 - \rho)\widetilde{\theta}_{i,t} - \rho\eta_{i,t} + \varepsilon_{i,t}. \tag{60}$$

Thus $u_{i,t}$ is independent of the deployed coefficient. Joint Gaussianity strengthens $\text{Cov}(\widetilde{x}_{i,t}, u_{i,t}) = 0$ to independence. Write $\sigma_u^2 = \text{Var}(u_{i,t}) = \rho(1-\rho)V + \sigma_\varepsilon^2 > 0$. The deployed coefficient changes the conditional mean but not the law of the regression disturbance, which makes retraining noise exogenous to the preceding deployment.

3.4 Fresh-sample retraining as an exact autoregression

Finite-sample retraining turns the population fixed-point problem into a time-series problem. Each cross-section produces a new estimate of the slope, and that estimate determines the allocation applied to the next cross-section. The sequence of model vintages therefore records both the deterministic feedback in the population retraining operator and the sampling variation generated by estimation.

In the linear-Gaussian economy, this sequence has an exact autoregressive representation. The autoregressive coefficient is the economic loop gain, while the innovation is the ordinary least-squares estimation error. The result provides the foundation for the stability, diagnostic, and identification results that follow.

Theorem 6 (Exact autoregressive representation)**.** *For every t,*

$$\lambda_{t+1} = \rho + k\lambda_t + U_t, \tag{61}$$

where the innovations $\{U_t\}_{t\geq 0}$ are independent and identically distributed across periods and are independent of $(\lambda_0, U_0, \ldots, U_{t-1})$, with the exact distribution

$$U_t \overset{d}{=} \frac{\sigma_u}{\sqrt{V(n-1)}} T_{n-1}, \quad T_{n-1} \sim \text{Student-}t_{n-1}. \tag{62}$$

Consequently, $\mathbb{E}[U_t] = 0$ *for* $n \geq 3$*, and* $\mathrm{Var}(U_t) = \dfrac{\sigma_u^2}{V\,(n-3)}$ *for* $n \geq 4$*.*

The representation follows by substituting the residual decomposition into OLS:

$$\widehat{\lambda}_t^{\mathrm{OLS}} = \rho + k\lambda_t + \frac{\sum_{i=1}^n (\widetilde{x}_{i,t} - \overline{x}_t) u_{i,t}}{\sum_{i=1}^n (\widetilde{x}_{i,t} - \overline{x}_t)^2}. \tag{63}$$

Let $S_t = \sum_{i=1}^n (\widetilde{x}_{i,t} - \overline{x}_t)^2$. Conditional on the features, the numerator of the estimation error is Gaussian with mean zero and variance $\sigma_u^2 S_t$. Dividing that numerator by $\sigma_u\sqrt{S_t}$ therefore gives a standard normal variable Z_t whose conditional distribution does not depend on the features. Hence Z_t is independent of S_t. Since $S_t/V \sim \chi^2_{n-1}$, the estimation error $U_t = \sigma_u Z_t/\sqrt{S_t}$ has the scaled Student distribution in (62). Fresh independent cross-sections make these innovations i.i.d. across vintages and independent of the previously deployed coefficient.

The exact AR gives the archive a direct structural interpretation. Its persistence coefficient is the loop gain k, while its innovation scale depends on sample size, feature variance, and residual outcome uncertainty. Gaussianity delivers the exact Student law and zero finite-sample intercept bias. The Internet Appendix IA.4 shows that the archive slope coefficient and stability frontier survive under weaker sampling conditions even when the intercept is biased.

The autoregressive representation also converts the population frontier into a sharp stochastic phase transition.

Theorem 7 (Stochastic phase transition)**.** *Let* λ_t *satisfy equation* (61) *from any nonanticipative initial condition* λ_0*.*

(i) Suppose that $|k| < 1$*. The process is geometrically ergodic and has a unique stationary distribution represented by*

$$\lambda_\infty = \lambda^* + \sum_{j=0}^{\infty} k^j U_{-j}, \quad \lambda^* = \frac{\rho}{1-k}. \tag{64}$$

The stationary distribution is symmetric about λ^**. For* $n \geq 3$*,* $\mathbb{E}[\lambda_\infty] = \lambda^*$ *and for* $n \geq 4$*,* $\mathrm{Var}(\lambda_\infty) = \frac{\sigma_u^2}{V(n-3)}\frac{1}{1-k^2}$*. The Internet Appendix IA.4 records its exact tail index and moment implications under Gaussian sampling.*

(ii) Suppose that $k = 1$*. For* $n \geq 3$*,*

$$\lambda_t = \lambda_0 + \rho t + \sum_{j=0}^{t-1} U_j, \tag{65}$$

and $\lambda_t/t \to \rho$ *almost surely. Since* $\rho > 0$ *under the maintained assumptions, the coefficient diverges at the linear rate* ρt*. If* $\rho = 0$*, the coefficient instead follows a centered random walk.*

(iii) Suppose that $k = -1$ *and* $n \geq 3$*. Let* $\lambda^- = \rho/2$ *and define* $W_t = (-1)^t(\lambda_t - \lambda^-)$*. Then*

$$W_{t+1} = W_t + (-1)^{t+1} U_t. \tag{66}$$

Because the innovation distribution is symmetric, $\{(-1)^{t+1}U_t\}$ *is i.i.d. with the same law as* $\{U_t\}$*. Hence* W_t *is a centered nondegenerate random walk and satisfies* $\limsup_{t\to\infty} W_t = +\infty$

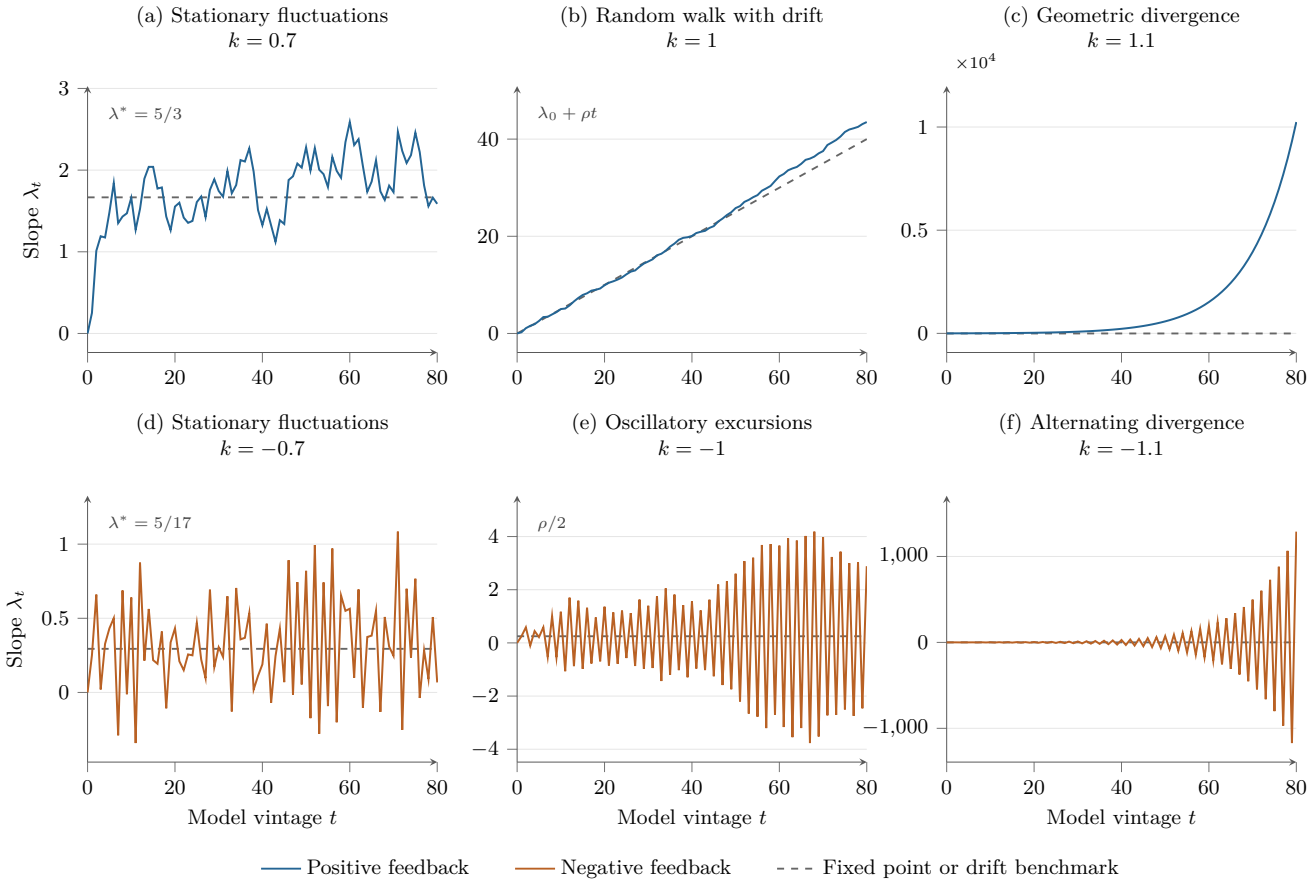


Figure 8: Fresh-sample archive dynamics across the stability frontier. Each panel plots 80 updates of the exact recursion $\lambda_{t+1} = \rho + k\lambda_t + U_t$, starting from $\lambda_0 = 0$. All panels use the same innovation realization and set $\rho = 0.5$, $n = 12$, $V = 1$, and $\sigma_\varepsilon^2 = 0.25$. These values imply $\sigma_u^2 = 0.5$ and independent innovations $U_t \overset{d}{=} T_{11}/\sqrt{22}$. Dashed lines mark the stationary mean λ^* in panels (a) and (d), the drift benchmark $\lambda_0 + \rho t$ in panel (b), the center $\rho/2$ in panel (e), and the repelling fixed point $\widetilde{\lambda}$ in panels (c) and (f). Vertical scales differ across panels. The paths are finite-horizon illustrations of the regimes characterized in Theorem 7.

and $\liminf_{t\to\infty} W_t = -\infty$ *almost surely. The original coefficient therefore has unbounded oscillatory excursions around* $\rho/2$.

(iv) Suppose that $|k| > 1$. *Define the repelling population fixed point* $\widetilde{\lambda} = \rho/(1-k)$. *Then*

$$\lambda_t - \widetilde{\lambda} = k^t \left[\lambda_0 - \widetilde{\lambda} + M_t\right], \qquad M_t = \sum_{j=0}^{t-1} k^{-j-1} U_j. \tag{67}$$

The sequence M_t *converges almost surely to a finite random variable* M_∞, *and* $\Pr(\lambda_0 - \widetilde{\lambda} + M_\infty = 0) = 0$. *Therefore* $|\lambda_t| \to \infty$ *almost surely at the geometric rate* $|k|^t$. *If* $k < -1$, *the sign of* $\lambda_t - \widetilde{\lambda}$ *eventually alternates every period.*

Figure 8 illustrates how the exact autoregression in Theorem 6 produces the regimes in Theorem 7. All panels use the same initial slope and the same realization of the Student innovations, so the differences between paths arise entirely from the loop gain k. In panels (a) and (d), $|k| < 1$ and the effect of the initial

condition decays, while fresh estimation errors sustain fluctuations around λ^*. Positive feedback makes deviations persistent, while negative feedback induces negative serial correlation. Population contraction therefore permits a stationary stochastic archive without requiring its realized coefficients to converge to a constant.

At the two boundaries, sampling disturbances accumulate. In panel (b), $k = 1$ and the informational component ρ add a positive drift, giving the linear divergence described by $\lambda_t/t \to \rho$. In panel (e), $k = -1$ and the transformed process $(-1)^t(\lambda_t - \rho/2)$ is a centered random walk. Its accumulated disturbances appear in the original coefficient as oscillatory excursions around $\rho/2$.

Beyond either boundary, the feedback magnifies the disturbances geometrically. Panel (c) illustrates eventual divergence on one side of the repelling fixed point when $k > 1$. Panel (f) illustrates the alternating divergence when $k < -1$. In both cases, the asymptotic magnitude grows at a rate $|k|^t$. Sampling noise thus changes the distribution of the archive paths while leaving the absolute-gain frontier at $|k| = 1$.

The Internet Appendix IA.4 gives a complementary stationary diagnostic. As $k \uparrow 1$, coefficient levels become more persistent and dispersed while the adjacent revisions become smoother.

Corollary 9 (Identification from model vintages)**.** *Suppose that $n \geq 4$ and that the coefficient process is observed in its subcritical stationary distribution. Then $\mathrm{Corr}\,(\lambda_{t+1}, \lambda_t) = k$ and $\mathbb{E}[\lambda_t] = \frac{\rho}{1-k}$. The loop gain and feature reliability are therefore identified by $k = \mathrm{Corr}\,(\lambda_{t+1}, \lambda_t)$ and $\rho = (1-k)\mathbb{E}[\lambda_t]$. The stationary variance provides the additional restriction*

$$\mathrm{Var}(\lambda_t) = \frac{\rho(1-\rho)V + \sigma_\varepsilon^2}{V(n-3)(1-k^2)}. \tag{68}$$

Corollary 9 turns ordinary estimation noise into identifying variation. A single deterministic deployment does not separate fundamental information from feedback, but fresh-sample estimation perturbs the deployed coefficient exogenously under Gaussianity. Archive persistence identifies k, its stationary mean identifies ρ, and its variance supplies an overidentifying restriction. The Internet Appendix IA.4 shows that persistence continues to identify k beyond Gaussian sampling, while the intercept may no longer identify ρ without correction.

3.5 Estimation and testing from model archives

Identification results above translate directly into estimators and specification tests. Suppose that an auditor observes $T + 1$ consecutive coefficient vintages $\lambda_1, \ldots, \lambda_{T+1}$ from the stationary subcritical regime. Let $(\widehat{\rho}_T, \widehat{k}_T)$ denote the intercept and slope from the time-series regression

$$\lambda_{t+1} = a + b\lambda_t + e_{t+1}, \quad t = 1, \ldots, T. \tag{69}$$

Under the exact autoregressive law in Theorem 6, the population coefficients in this regression are $a = \rho$ and $b = k$. The model archive can therefore be used to estimate the reliability of the feature and the strength of the allocation loop without observing the latent fundamental.

The same environment also permits an exact finite-sample test of whether feedback is present. The test uses the cross-sectional design sums stored with each model vintage and exploits the independence created

by retaining alternate observations from the archive. A randomized holdout provides a complementary estimator based on contemporaneous rather than intertemporal variation.

Proposition 8 (Inference from model archives and holdouts)**.** *Suppose that the coefficient process is in the stationary regime of Theorem 7(i), with $|k| < 1$ and $\sigma_u^2 > 0$. The following statements hold.*

(i) Suppose that $n \geq 4$. Then $(\widehat{\rho}_T, \widehat{k}_T) \to (\rho, k)$ almost surely and $\sqrt{T}\left(\widehat{k}_T - k\right) \xrightarrow{d} N(0, 1-k^2)$. The asymptotic variance of the slope estimator is free of nuisance parameters. A feasible asymptotic standard error is therefore $\operatorname{se}(\widehat{k}_T) = \sqrt{\frac{1-\widehat{k}_T^2}{T}}$.

(ii) Suppose that $n \geq 2$ and that the archive also records $S_t = \sum_{i=1}^{n} (\widetilde{x}_{i,t} - \overline{x}_t)^2$, the design sum of the cross-section used to estimate λ_{t+1}. Retain only the observations with odd indices $t \in \{1, 3, 5, \ldots\}$ and let $m = \lceil T/2 \rceil$ be their number. Estimate equation (69) by weighted least squares using weight S_t on observation t. Under $H_0 \colon k = 0$, the conventional t-statistic for the slope has the exact distribution

$$t_{\widehat{k}} \sim t_{m-2} \tag{70}$$

for every $m \geq 3$ and for all admissible values of the remaining parameters.

(iii) Consider a single cross-section of size n in which the deployed slope is λ and a fraction $h \in (0,1)$ of the units is randomly assigned to a holdout group. The holdout group receives the baseline allocation, while the remaining units receive the score-dependent allocation. If the treated and holdout slopes are estimated separately by OLS with intercepts, then

$$\sqrt{n}\left(\widehat{\beta}_{\text{treated}} - \widehat{\beta}_{\text{holdout}} - k\lambda\right) \xrightarrow{d} N\left(0, \frac{\sigma_u^2}{V}\left[\frac{1}{1-h} + \frac{1}{h}\right]\right). \tag{71}$$

Part (i) gives a nuisance-free asymptotic standard error because the innovation variance cancels against the stationary variance of the AR regressor. Part (ii) gives an exact finite-sample test of $H_0 : k = 0$. Adjacent archive observations share a coefficient, so the test retains alternate periods. Conditional on their design sums, the retained errors are independent Gaussian variables with variance proportional to S_t^{-1}. Weighting by S_t therefore gives a classical homoskedastic normal regression and the exact Student statistic in (70).

Part (iii) identifies the same loop gain through contemporaneous randomized variation. The treated-minus-holdout slope difference is $k\lambda$, and the variance in (71) is minimized at $h = 1/2$ when statistical precision is the only objective. The archive and holdout approaches therefore use different identifying variation. The archive requires a stable parametric retraining environment but no intervention. The holdout consumes experimental capacity but does not require a long stationary history. The Internet Appendix IA.4 gives the associated structural variance restriction.

3.6 Pooled retraining with decreasing gain

Fresh-sample retraining has constant gain. If the model instead re-estimates on the complete history, the influence of each new cross-section declines over time, and retraining becomes a decreasing-gain process.

Suppose that, at the end of period t, the modeler estimates the slope using all cross-sections observed through that date. Period fixed effects remove changes in the cross-sectional intercept, so the pooled estimator is

$$\lambda_{t+1} = \frac{\sum_{s=1}^{t}\sum_{i=1}^{n}(\widetilde{x}_{i,s} - \overline{x}_s)(y_{i,s} - \overline{y}_s)}{\sum_{s=1}^{t} S_s}, \quad S_s = \sum_{i=1}^{n}(\widetilde{x}_{i,s} - \overline{x}_s)^2. \tag{72}$$

Each observation remains in the estimation sample after it is collected, even though its outcome was generated by the coefficient deployed at that date. The historical sample therefore records a sequence of policy regimes rather than repeated observations from a fixed data-generating process.

Proposition 9 (Decreasing-gain retraining)**.** *Suppose the modeler deploys the pooled OLS coefficient in equation (72) after every period.*

(i) If $k < 1$, then $\lambda_t \to \lambda^*$ almost surely from every nonanticipative initial condition. This includes $k \leq -1$, where fresh-sample retraining is unstable. No auxiliary projection device or exogenous boundedness restriction is required.

(ii) If $k = 1$, then

$$\frac{\lambda_t}{\log t} \longrightarrow \rho \qquad \text{almost surely.} \tag{73}$$

Under the maintained assumption $\rho > 0$, the pooled coefficient therefore diverges to $+\infty$ at the logarithmic rate $\rho \log t$. If $\rho = 0$, the pooled coefficient converges almost surely to a finite random limit.

(iii) If $k > 1$, then $|\lambda_t| \to \infty$ almost surely from every nonanticipative initial condition, including the repelling rest point $\widetilde{\lambda} = \frac{\rho}{1-k}$. The divergence is polynomial and satisfies

$$|\lambda_t| \asymp t^{\,k-1} \qquad \text{almost surely.} \tag{74}$$

In particular, the probability of convergence to the repelling rest point is zero.

Figure 9 compares the two learning rules using the same realizations of fundamentals, features, and outcome shocks. Each rule generates its own outcome history through the coefficient deployed in each period. Panel (a) illustrates the distinction between a stationary archive and convergence to a constant. With $k = 0.7$, the fresh-sample coefficients continue to fluctuate around λ^*, while the pooled coefficients approach that fixed point. Under pooling, the weight on a new cross-section is $S_t / \sum_{s=1}^{t} S_s$, which tends to zero almost surely. The decreasing gain allows sampling fluctuations to dissipate in this stable region.

Panel (b) shows why the change in the stability region is asymmetric. At $k = -1.1$, fresh-sample retraining magnifies deviations with alternating signs. Pooling dampens these oscillations and restores convergence to the same population fixed point. Retaining past observations changes the adjustment process without changing the coefficient at which the population retraining map leaves the model unchanged.

Panels (c) and (d) show the limit of this stabilization. At $k = 1$, pooling slows linear divergence to logarithmic divergence. At $k > 1$, it slows down the geometric divergence to polynomial divergence at a

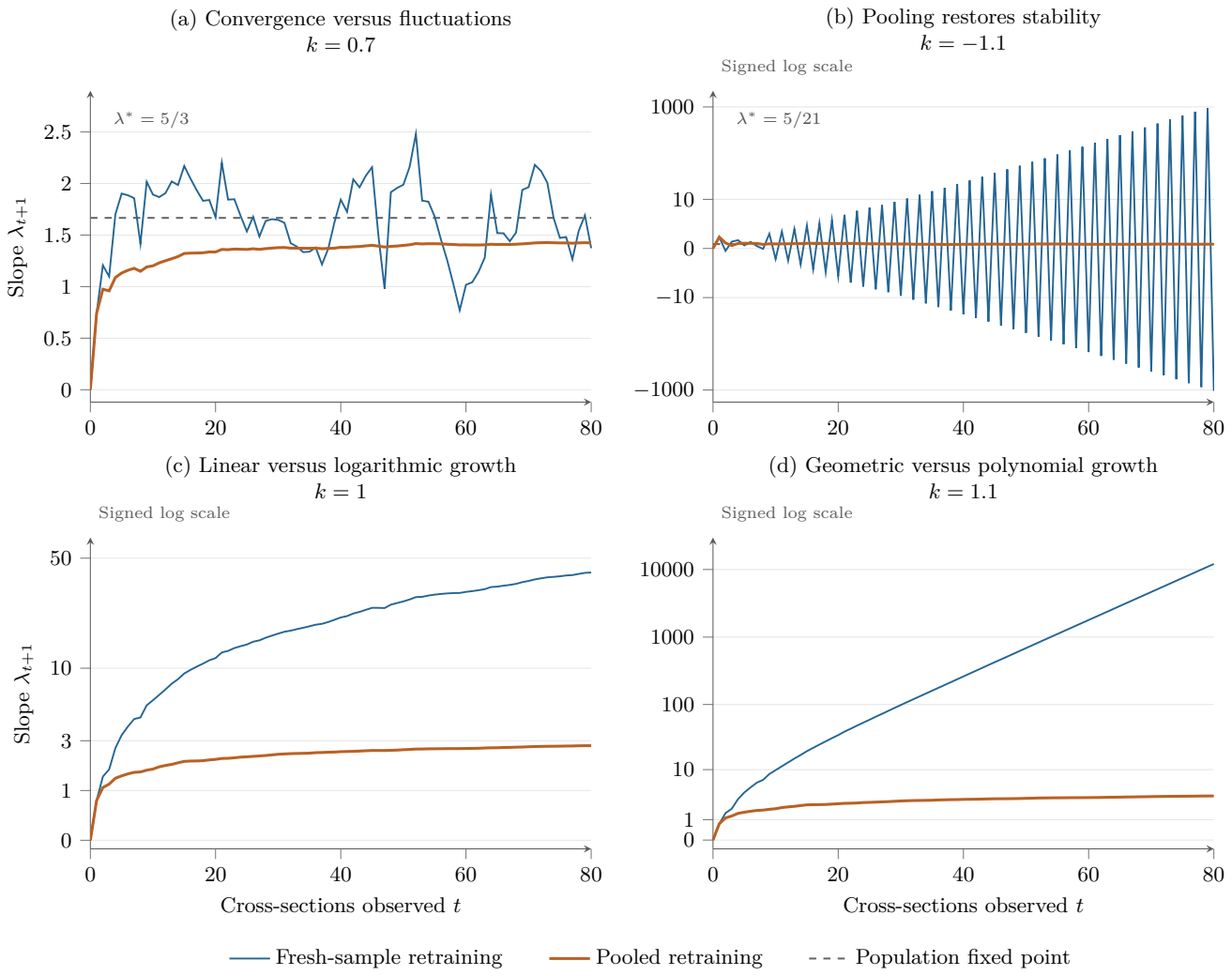


Figure 9: How pooling changes retraining dynamics. Each panel compares fresh-sample and pooled OLS retraining over 80 updates, starting from $\lambda_1 = 0$. All panels use the same Gaussian draws of fundamentals, feature noise, and outcome shocks, with $n = 12$, $\rho = 0.5$, $V = 1$, and $\sigma_\varepsilon^2 = 0.25$. Outcomes are generated separately under each rule's deployed coefficient. Pooled OLS retains all previous cross-sections and includes period fixed effects. Dashed lines mark $\lambda^* = \rho/(1-k)$ in panels (a) and (b). Panel (a) uses a linear vertical scale. Panels (b)–(d) use the transformation $\operatorname{sgn}(\lambda)\log(1+|\lambda|)$, with ticks labeled in the original coefficient units. The finite paths illustrate the convergence and divergence regimes established in Proposition 9 and Theorem 7.

rate t^{k-1}. Therefore, a growing estimation sample restores convergence for every $k < 1$, including strong negative feedback, but cannot restore convergence when the reinforcement feedback reaches or exceeds one. Table 1 summarizes the comparison.

The Internet Appendix IA.5 gives the stochastic-approximation representation, its connection to decreasing-gain learning, and rolling-window variants. Proposition 7 gives the corresponding worst-case amplification bound beyond linear estimation.

Table 1: Fresh-sample and pooled retraining

The table compares the long-run behavior of the estimated slope under affine feedback with slope k, Gaussian cross-sections with $n \geq 3$, positive residual variance, and $\rho > 0$. Fresh-sample retraining uses only the latest cross-section. Pooled retraining uses all available cross-sections with period-specific intercepts. For $k < 1$, the population fixed point is $\lambda^* = \rho/(1-k)$. Convergence and divergence statements hold almost surely, and divergence rates refer to coefficient magnitudes. Pooling reduces the influence of each new cross-section, extending convergence to all $k < 1$ and slowing divergence when $k \geq 1$.

Feedback region	Fresh-sample retraining	Pooled retraining
$k < -1$	Alternating geometric divergence	Converges to λ^*
$k = -1$	Oscillatory random walk	Converges to λ^*
$-1 < k < 1$	Stationary and geometrically ergodic	Converges to λ^*
$k = 1$	Linear divergence at rate ρt	Logarithmic divergence at rate $\rho \log t$
$k > 1$	Geometric divergence	Polynomial divergence at rate t^{k-1}

4 Stateful Retraining

The preceding sections study feedback that operates within one deployment period. Throughout this section we maintain $\beta \geq 0$ and $\gamma \geq 0$, so the contemporaneous gain $k = \beta\gamma$ is nonnegative. Section 3.4 already gives the complete scalar phase diagram for negative feedback. The durable-state analysis focuses on reinforcing allocations, which are also the case developed economically in the companion paper.

Many allocations create a state that survives across model vintages. Credit finances capital, recommendations can build a stock of demand, and treatment can change health beyond the period in which the score is assigned. Once these effects persist, current outcomes depend on both the current score and a state generated by earlier scores.

The status of the results is important. The population frontier and the finite-horizon large-panel limit are proved. For a fixed panel, a negative Lyapunov exponent is sufficient for a unique causal stationary and ergodic solution, while a positive exponent implies exponential expansion of homogeneous coefficient products. We prove that every population economy strictly below the frontier has a negative finite-panel exponent for all sufficiently large panels. Above the population frontier, we prove fixed-horizon expansion in expected log norm and with probability approaching one. Turning those expanding blocks into a positive infinite-horizon Lyapunov exponent requires a lower alignment argument that is not yet implied by the primitive Gaussian panel model. The Internet Appendix IA.3 isolates a sufficient condition. The exact critical case also remains open. Numerical results are used only to study these unresolved statements.

4.1 Population benchmark

Stateful performative models allow earlier deployments to affect later data through an inherited state (Brown et al., 2022; Izzo et al., 2022; Li and Wai, 2022; Ray et al., 2022). We use the durable-allocation specialization developed economically in Livdan et al. (2026) only to obtain the deterministic benchmark for the finite-panel theory below.

Let the centered fundamental follow $\widetilde{\theta}_{i,t} = \phi\widetilde{\theta}_{i,t-1} + w_{i,t}$, where $\phi \in [0,1)$ and $w_{i,t} \sim N(0,(1-\phi^2)\sigma_\theta^2)$. The observed feature is $\widetilde{x}_{i,t} = \widetilde{\theta}_{i,t} + \eta_{i,t}$ with $\eta_{i,t} \sim N(0,\sigma_\eta^2)$, while $\varepsilon_{i,t} \sim N(0,\sigma_\varepsilon^2)$. Shocks are mutually independent across units and dates. We take the fundamental process in its stationary Gaussian law.

Specifically, for $t \in \mathbb{Z}$, let $\widetilde{\theta}_{i,t} = \sum_{j=0}^{\infty} \phi^j w_{i,t-j}$. Then $\widetilde{\theta}_{i,t} \sim N(0, \sigma_\theta^2)$ at every date, and the fundamental processes are independent across units. Define $V = \sigma_\theta^2 + \sigma_\eta^2$ and $\rho = \sigma_\theta^2 / V$.

The model deploys slope λ_t on the demeaned feature. The induced allocation accumulates into capital according to

$$K_{i,t} = \delta K_{i,t-1} + \gamma \lambda_t \widetilde{x}_{i,t}^c, \qquad \delta \in [0, 1), \tag{75}$$

and the outcome is

$$y_{i,t} = \theta_{i,t} + \beta K_{i,t} + \varepsilon_{i,t}. \tag{76}$$

The contemporaneous loop gain is $k = \beta\gamma$. Capital durability and fundamental persistence enlarge this gain because an allocation made today remains in future outcomes and remains correlated with future features.

The within-period timing determines which objects are predetermined when period-t shocks arrive and which information the next estimate uses. At the beginning of period t, the inherited capital stock $K_{i,t-1}$ and the previously estimated coefficient λ_t are given. The fundamental $\theta_{i,t}$ and the observed feature $x_{i,t}$ are realized. The model assigns the score $\lambda_t \widetilde{x}_{i,t}^c$. The score induces current investment and determines $K_{i,t}$ through equation (75). The outcome $y_{i,t}$ is realized through equation (76). The modeler runs the period-t cross-sectional regression and obtains λ_{t+1}. Figure IA.1 displays this sequence within one period, together with the two channels through which the period-t deployment crosses the period boundary.

This timing explains how feedback persists across model vintages. A score assigned today changes the borrower's investment, and surviving capital affects future outcomes after the model has been updated. Persistent fundamentals keep inherited capital correlated with future features, so subsequent regressions attribute part of the effect of earlier allocations to the current score. We assume that the initial slope and capital cross-section depend only on the primitive history before the first retraining period and, if needed, auxiliary randomization independent of the entire primitive process. Let $\mathcal{F}_t$ be the sigma-algebra generated by $\{\theta_{i,s}, \eta_{i,s}, \varepsilon_{i,s}\}_{i \le n,\ s \le t}$ together with the initial slope and capital cross-section. Then $K_{i,t-1}$ and λ_t are $\mathcal{F}_{t-1}$-measurable, while the current innovations $w_{i,t}$, $\eta_{i,t}$, and $\varepsilon_{i,t}$ are independent of $\mathcal{F}_{t-1}$ and have conditional mean zero. All subsequent conditional-mean statements use this filtration.

To obtain the population benchmark, suppose that a fixed slope λ has been deployed long enough for the economy to be stationary. For this calculation, features are centered by their population mean, so the allocation uses $\widetilde{x}_{i,t} = x_{i,t} - \mathbb{E}[x_{i,t}]$. We return to cross-sectional sample demeaning in the finite-panel analysis below. Stationary capital is then

$$K_{i,t} = \gamma\lambda \sum_{j=0}^{\infty} \delta^j \widetilde{x}_{i,t-j}. \tag{77}$$

The current feature has variance V, while $\mathrm{Cov}(\widetilde{x}_{i,t}, \widetilde{x}_{i,t-j}) = \phi^j \rho V$ for $j \ge 1$. Consequently, $\mathrm{Cov}(\widetilde{x}_{i,t}, K_{i,t}) = \gamma\lambda V[1 + \rho \sum_{j=1}^{\infty} (\delta\phi)^j]$. Summing the geometric series gives

$$\frac{\mathrm{Cov}(\widetilde{x}_{i,t}, K_{i,t})}{V} = \gamma\lambda\psi, \qquad \psi = \frac{1 - \delta\phi(1-\rho)}{1 - \delta\phi} = 1 + \frac{\delta\phi\rho}{1 - \delta\phi}. \tag{78}$$

Population regression combines the fundamental contribution $\mathrm{Cov}(\widetilde{x}_{i,t}, \theta_{i,t})/V = \rho$ with the contribution of

capital. Since $k = \beta\gamma$, it maps the deployed slope into

$$\lambda' = \rho + k\psi\lambda = \rho + \kappa\lambda, \qquad \kappa = k\psi. \tag{79}$$

Durability therefore affects the population retraining slope through the correlation between past allocations and the current feature. The product $\delta\phi$ combines capital survival with fundamental persistence, while ρ measures the share of feature variance attributable to the fundamental. If either durability or fundamental persistence is absent, inherited allocations make no contribution to the population slope and the effective gain reduces to k.

Corollary 10 (Destabilization through durability and persistence)**.** *The population retraining map $\lambda \mapsto \rho + \kappa\lambda$ is stable if and only if $\kappa < 1$, equivalently $k < k_c$, where*

$$k_c = \frac{1}{\psi} = \frac{1 - \delta\phi}{1 - \delta\phi(1 - \rho)}. \tag{80}$$

The critical gain is weakly decreasing in δ, ϕ, and ρ. It is strictly decreasing in δ when $\rho\phi > 0$, in ϕ when $\rho\delta > 0$, and in ρ when $\delta\phi > 0$. If $\delta = 0$, $\phi = 0$, or $\rho = 0$, then $\kappa = k$ and $k_c = 1$. For fixed $\rho > 0$, $k_c \to 0$ as $\delta\phi \uparrow 1$.

The corollary provides the deterministic frontier against which the finite-panel system is evaluated. Its economic implications, welfare consequences, and policy interpretation are developed in the companion paper. Here the important object is the benchmark $\kappa = 1$ and whether finite-panel stochastic retraining inherits that frontier.

We now return to the finite panel. The modeler re-estimates the slope by cross-sectional OLS in every period using the same panel of $n \geq 2$ units. Let

$$s_t^{xx} = \frac{1}{n}\sum_{i=1}^{n}(\widetilde{x}_{i,t}^{c})^2, \quad s_t^{x\theta} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_{i,t}^{c}\widetilde{\theta}_{i,t}^{c}, \tag{81}$$

and define

$$c_t^{xK} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{x}_{i,t}^{c}K_{i,t-1}^{c}, \quad p_t = \frac{1}{n}\sum_{i=1}^{n}\widetilde{\theta}_{i,t}^{c}K_{i,t}^{c}. \tag{82}$$

The variable p_t measures the covariance between current fundamentals and the capital inherited from current and earlier allocations. In a credit application, it records whether borrowers with stronger latent repayment capacity also carry more capital because earlier scores gave them greater access to financing.

The identity

$$c_t^{xK} = \phi p_{t-1} + \zeta_t + \iota_t, \tag{83}$$

where

$$\zeta_t = \frac{1}{n}\sum_{i=1}^{n}w_{i,t}^{c}K_{i,t-1}^{c}, \quad \iota_t = \frac{1}{n}\sum_{i=1}^{n}\eta_{i,t}^{c}K_{i,t-1}^{c}, \tag{84}$$

gives the exact finite-panel recursion

$$\lambda_{t+1} = k\lambda_t + \frac{\beta\delta\phi}{s_t^{xx}}p_{t-1} + \nu_t. \tag{85}$$

The inherited covariance evolves according to

$$p_t = \delta\phi p_{t-1} + \gamma s_t^{\theta x}\lambda_t + \delta\zeta_t. \tag{86}$$

where $s_t^{\theta x} = s_t^{x\theta}$ and

$$\nu_t = \frac{s_t^{x\theta} + s_t^{x\varepsilon} + \beta\delta(\zeta_t + \iota_t)}{s_t^{xx}}, \quad s_t^{x\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_{i,t}^{c}\varepsilon_{i,t}^{c}. \tag{87}$$

Conditional on the past, ζ_t and ι_t have mean zero. Their conditional variances depend on the inherited dispersion of capital, which makes the noise in the two-state system state-dependent.

Let $Z_t = (\lambda_t, p_{t-1})'$. Equations (85) and (86) can be written as

$$Z_{t+1} = A_t\, Z_t + \xi_t, \tag{88}$$

where

$$A_t = \begin{pmatrix} k & \beta\delta\phi/s_t^{xx} \\ \gamma s_t^{\theta x} & \delta\phi \end{pmatrix}, \quad \xi_t = \begin{pmatrix} \nu_t \\ \delta\zeta_t \end{pmatrix}. \tag{89}$$

This representation separates the contemporaneous score channel from the inherited-capital channel. The upper-left entry is the static loop gain, while the off-diagonal entries transmit past allocations into current retraining and current scores into future capital.

Proposition 10 (Large-panel dynamics and the stability frontier)**.** *As $n \to \infty$, the random coefficient matrix in equation (89) converges in probability to*

$$\overline{A} = \begin{pmatrix} k & \beta\delta\phi/V \\ \gamma\sigma_\theta^2 & \delta\phi \end{pmatrix}. \tag{90}$$

Its characteristic polynomial is

$$\det(\mu I - \overline{A}) = \mu^2 - (k + \delta\phi)\mu + k\delta\phi(1-\rho). \tag{91}$$

Writing $d = \delta\phi$, the eigenvalues

$$\mu_{1,2} = \frac{k + d \pm \sqrt{(k-d)^2 + 4kd\rho}}{2} \tag{92}$$

are real and nonnegative. When $kd\rho > 0$ they satisfy $\mu_1 > \max\{k, d\}$ and $\mu_2 < \min\{k, d\}$, and $\mu_2 > 0$ exactly when $kd(1-\rho) > 0$, since $\mu_1\mu_2 = kd(1-\rho)$. The spectral radius of $\overline{A}$ is below one if and only if $\kappa < 1$, with $\kappa = k\psi$ the effective gain in equation (79). In that case, the large-panel deterministic system has the stationary state

$$\overline{\lambda}^* = \frac{\rho}{1-\kappa}, \quad \overline{p}^* = \frac{\gamma\sigma_\theta^2}{1-\delta\phi}\,\overline{\lambda}^*. \tag{93}$$

The eigenvalue ordering gives the two-state system a two-mode reading. The dominant mode combines current feedback with inherited capital and exceeds both the contemporaneous gain and the capital-covariance decay rate, while the smaller mode describes the faster adjustment of the remaining linear combination of the two states. The stationary covariance state $\overline{p}^*$ gives economic content to the second coordinate. Borrowers with stronger fundamentals carry more inherited capital in proportion to the deployed

slope, and this covariance is what makes the stationary slope rise steeply as the frontier approaches.

The frontier follows by evaluating equation (91) at one. In particular,

$$\det(I - \overline{A}) = (1-k)(1-\delta\phi) - k\delta\phi\rho = (1-\delta\phi)(1-\kappa). \tag{94}$$

Since the two eigenvalues are nonnegative and $\delta\phi < 1$, this determinant is positive exactly when both eigenvalues lie below one. The population stability frontier is therefore the spectral-radius condition for the large-panel limit of the exact finite-panel dynamics.

Two qualifications are useful but secondary to the stability argument. The probability limit $\overline{A}$ need not be equal to $\mathbb{E}[A_t]$, so finite-panel stationary means need not be equal to the population fixed point. In addition, convergence of A_t can be strengthened to convergence of the finite-panel path over every fixed horizon. The Internet Appendix IA.3 states both results, while the next subsection turns to unbounded-horizon stability.

The two eigenvalues describe how the effects of a model vintage decay in the large-panel deterministic system. Whether they also govern the autocorrelation of the stationary finite-panel process is not established. The forcing that drives the finite-panel fluctuations is serially correlated because the fundamental follows an AR(1), so the long-lag autocorrelation of the limiting fluctuation system could be governed by primitive persistence as well as by the dominant eigenvalue. In the calibrations reported below, the two candidates are numerically close, with $\phi = 0.875$ and $\mu_1 \approx 0.8808$ at $k = 0.30$, so a fitted decay of approximately 0.89 cannot sharply distinguish the coupled mode from primitive persistence. The autocorrelation evidence in Section 4.2.1 is therefore presented as a calibration-specific numerical finding rather than as an identification result, and the finite-panel analysis below replaces the spectrum of the limiting matrix with the growth rate of random matrix products as the relevant stability object.

4.2 Finite-panel dynamics and Lyapunov stability

The two-state system is useful for interpretation, although its disturbances inherit state-dependent variances from lagged capital. A higher-dimensional representation removes this complication. The resulting recurrence is linear, its additive term is exogenous, and both the coefficient matrix and the additive term are generated entirely by primitive shocks.

Let $K^c_{t-1} = (K^c_{1,t-1}, \ldots, K^c_{n,t-1})'$. Cross-sectional demeaning implies $\mathbb{1}'K^c_{t-1} = 0$. Fix an $n \times (n-1)$ matrix Q with orthonormal columns spanning $\mathbb{1}^\perp$, and define the reduced state

$$\widetilde{W}_t = \begin{pmatrix} \lambda_t \\ Q'K^c_{t-1} \end{pmatrix} \in \mathbb{R}^n. \tag{95}$$

Lemma 2 (Exact linear representation)**.** *The reduced state satisfies the exact stochastic recurrence*

$$\widetilde{W}_{t+1} = \widetilde{B}_t \widetilde{W}_t + \widetilde{E}_t, \tag{96}$$

where

$$\widetilde{B}_t = \begin{pmatrix} k & \dfrac{\beta\delta}{S_t}(\widetilde{x}^c_t)'Q \\ \gamma Q'\widetilde{x}^c_t & \delta I_{n-1} \end{pmatrix}, \quad S_t = \sum_{i=1}^{n} (\widetilde{x}^c_{i,t})^2, \tag{97}$$

and

$$\widetilde{E}_t = \begin{pmatrix} g_t & 0 & \cdots & 0 \end{pmatrix}', \quad g_t = \frac{\sum_{i=1}^n \widetilde{x}_{i,t}^c (\widetilde{\theta}_{i,t}^c + \varepsilon_{i,t}^c)}{\sum_{i=1}^n (\widetilde{x}_{i,t}^c)^2}. \tag{98}$$

The additive term has the marginal representation

$$g_t - \rho = \frac{\sigma_r}{\sqrt{V(n-1)}} T_{n-1}, \quad \sigma_r^2 = \sigma_\theta^2(1-\rho) + \sigma_\varepsilon^2, \tag{99}$$

where T_{n-1} is a Student t variable with $n-1$ degrees of freedom. The mean of g_t is therefore undefined for $n = 2$, equals ρ with infinite variance for $n = 3$, and for $n \geq 4$

$$\mathrm{Var}(g_t) = \frac{\sigma_r^2}{V(n-3)}. \tag{100}$$

The logarithmic moment required below exists for every $n \geq 2$. The sequences $\{\widetilde{B}_t\}$ and $\{\widetilde{E}_t\}$ are measurable functions of the primitive panel process and form a stationary, geometrically β-mixing sequence. They do not depend functionally on deployed scores or inherited capital, although they are dependent on one another contemporaneously and serially dependent through the persistent fundamentals.

Lemma 2 replaces state-dependent regression noise with an exogenous additive forcing term. We call $\widetilde{E}_t$ a forcing term rather than an innovation because its first component has mean ρ for $n \geq 3$ and it is statistically dependent on the contemporaneous coefficient matrix. Serial dependence remains because the fundamental is persistent, so the current coefficient matrix is correlated with earlier matrices and with the capital inherited from earlier periods. This dependence is generated by the primitive AR(1) process rather than by the endogenous state. The distinction permits the use of stationary random-recurrence results.

The Internet Appendix IA.3 gives the one-period spectral decomposition of the coefficient matrix. It separates a feature-aligned score-capital mode from capital directions that decay at rate δ and shows why rotating feature-aligned subspaces make the Lyapunov exponent, rather than one-period eigenvalues, the relevant long-run object.

Define the top Lyapunov exponent on the reduced state space by

$$\gamma_{L,n} = \lim_{t\to\infty} \frac{1}{t} \mathbb{E}\left[\log \left\| \widetilde{B}_t \cdots \widetilde{B}_1 \right\|\right], \tag{101}$$

where $\|\cdot\|$ is the Euclidean operator norm. The subscript indicates that the exponent depends on the panel size through the dimension and the law of the coefficient matrices. For fixed n, every matrix norm yields the same exponent, and fixing the operator norm provides a common convention for comparisons across panel sizes. The limit exists under the logarithmic moment conditions

$$\mathbb{E}\log^+\|\widetilde{B}_1\| < \infty, \quad \mathbb{E}\log^+\|\widetilde{E}_1\| < \infty, \tag{102}$$

which are verified in the appendix. Kingman's subadditive ergodic theorem implies that the normalized logarithm of the matrix-product norm converges almost surely to the same constant.

The Lyapunov exponent governs propagation of the full reduced state. It need not equal a slope-specific persistence rate because dominant directions can lie mainly in capital coordinates. The Internet

Appendix IA.3 discusses this distinction together with the corresponding moment and critical-case qualifications.

Theorem 8 (Contraction under a negative Lyapunov exponent)**.** *Suppose that the logarithmic moment conditions* (102) *hold and* $\gamma_{L,n} < 0$*. Then equation* (96) *has a unique causal stationary and ergodic solution, given by the almost surely convergent series*

$$\widetilde{W}_t^* = \widetilde{E}_{t-1} + \sum_{j=1}^{\infty} \widetilde{B}_{t-1} \cdots \widetilde{B}_{t-j}\, \widetilde{E}_{t-j-1}. \tag{103}$$

Any two solutions driven by the same primitive sequence converge to one another almost surely from arbitrary initial conditions. The vintage-slope process initialized in the invariant law is therefore stationary and ergodic.

The theorem holds under the actual dependence structure of the model, because the random-recurrence results it invokes require only a stationary ergodic sequence of pairs $\{(\widetilde{B}_t, \widetilde{E}_t)\}$ together with the logarithmic moments in equation (102). Serial dependence of the coefficients and contemporaneous dependence between the coefficient matrix and the forcing term are permitted. The series solution has a direct economic reading. The current model vintage aggregates all past regression disturbances after each has been propagated through the subsequent sequence of allocation rules. A negative exponent implies that the effect of every initial condition eventually disappears, even though individual coefficient matrices may be locally expansive. In a lending application, temporary overreaction in one model vintage is reversed by later vintages, and the system remains stable when these expansions are dominated by contraction over long histories.

Conversely, if $\gamma_{L,n} > 0$, the homogeneous coefficient products expand exponentially at rate $\gamma_{L,n}$ almost surely. The Internet Appendix Proposition IA.3 states this implication formally. Divergence of the affine state additionally requires that the forcing term repeatedly excites the expanding directions, so the sign of the Lyapunov exponent alone does not complete the state-divergence argument.

The finite-panel results are stated in terms of the Lyapunov exponent, while the population frontier is stated in terms of the effective gain κ. The next result provides a one-sided bridge between these objects. It shows that every population economy strictly below the frontier eventually generates a contracting finite-panel recurrence as the cross-section grows.

Theorem 9 (Stable side of the large-panel Lyapunov bridge)**.** *Fix the primitives of the dynamic economy and let* $\gamma_{L,n}$ *denote the top Lyapunov exponent of the exact reduced coefficient sequence for a panel of size* n*. Let* $\mu_1(\kappa)$ *be the dominant eigenvalue of the large-panel matrix* $\overline{A}$ *in Proposition 10. Then*

$$\limsup_{n\to\infty} \gamma_{L,n} \le \log \max\left\{\mu_1(\kappa), \delta\right\}. \tag{104}$$

Consequently, if $\kappa < 1$*, there is an* $N < \infty$ *such that* $\gamma_{L,n} < 0$ *for every* $n \ge N$*. Dynamic retraining therefore has the unique causal stationary and ergodic solution characterized in Theorem 8 for every sufficiently large panel whenever the population economy lies strictly below the stability frontier.*

The theorem turns the stable side of the population frontier into a finite-panel stability result. Durability creates many capital directions that decay at the rate δ, while the coupled score-capital direction evolves

asymptotically at the rate $\mu_1(\kappa)$. Since both rates are below one when $\kappa < 1$, sufficiently large panels have a negative Lyapunov exponent. The conclusion is uniform over all panel sizes beyond a finite threshold, although the theorem does not provide a matching lower bound on the exponent.

A direct attempt at the positive side yields a partial result. For each n, let D_n be diagonal with first entry one and all remaining entries $n^{-1/2}$. We use the equivalent vector norm $\|w\|_n = \|D_n w\|_2$ and the induced operator norm. This deterministic change of norm leaves $\gamma_{L,n}$ unchanged.

Proposition 11 (Finite-block expansion above the population frontier)**.** *Fix primitives with $\kappa > 1$. Then there exist an integer $m \geq 1$, a number $\epsilon > 0$, and $N < \infty$ such that for every $n \geq N$,*

$$\frac{1}{m}\mathbb{E}\log\left\|\widetilde{B}_m \cdots \widetilde{B}_1\right\|_n > \epsilon. \tag{105}$$

Moreover, for the same m and some possibly smaller $\epsilon > 0$,

$$\Pr\left(\left\|\widetilde{B}_m \cdots \widetilde{B}_1\right\|_n > e^{m\epsilon}\right) \longrightarrow 1 \qquad (n \to \infty). \tag{106}$$

Proposition 11 establishes expansion over a fixed horizon, while the Lyapunov exponent measures growth over arbitrarily long histories. Kingman's formula expresses $\gamma_{L,n}$ as the infimum of expected log growth per period over all block lengths. Positive expected growth at one fixed block length therefore does not provide a positive lower bound on the exponent. Successive blocks can lose growth when their expanding directions change.

Figure 10 illustrates this distinction with two deterministic matrices, $B^{(1)} = \mathrm{diag}(6/5, 1/2)$ and $B^{(2)} = \mathrm{diag}(1/2, 6/5)$. Both have eigenvalues $6/5$ and $1/2$, so each expands one direction and contracts the other. In panel (a), every update uses $B^{(1)}$, and the product norm grows as $(6/5)^t$. In panel (b), the matrices alternate. Since $B^{(2)}B^{(1)} = (3/5)I$, every pair of updates contracts all directions. The asymptotic log growth rate is consequently $\log(6/5) > 0$ in the first case and $\frac{1}{2}\log(3/5) < 0$ in the second. Identical eigenvalues at each update can thus produce opposite signs of long-run growth.

In the stateful economy, earlier scores create a capital allocation across units. Its contribution to later regressions depends on its alignment with subsequent feature vectors. Changes in that alignment can move earlier perturbations into directions that later updates contract. Theorem 9 establishes stability below the population frontier for sufficiently large panels. Extending the positive finite-block result to a positive Lyapunov exponent requires a lower bound that controls losses from changing directions. Lemma IA.2 gives one sufficient condition. Whether or not the primitive Gaussian panel satisfies that condition remains open. The following conjecture states the stronger full large-panel limit.

Conjecture 1 (The large-panel Lyapunov frontier)**.** *As $n \to \infty$ with the primitives fixed,*

$$\gamma_{L,n} \longrightarrow \log\max\left\{\mu_1(\kappa), \delta\right\}. \tag{107}$$

The capital mode governs when $\mu_1(\kappa) < \delta$, while the coupled score-capital mode governs when $\mu_1(\kappa) > \delta$. Since $\delta < 1$ and $\mu_1(\kappa) > 1$ if and only if $\kappa > 1$, the limiting Lyapunov exponent changes sign at the population frontier $\kappa = 1$.

Theorem 9 establishes the upper bound in Conjecture 1. Lemma IA.2 shows that sufficient block

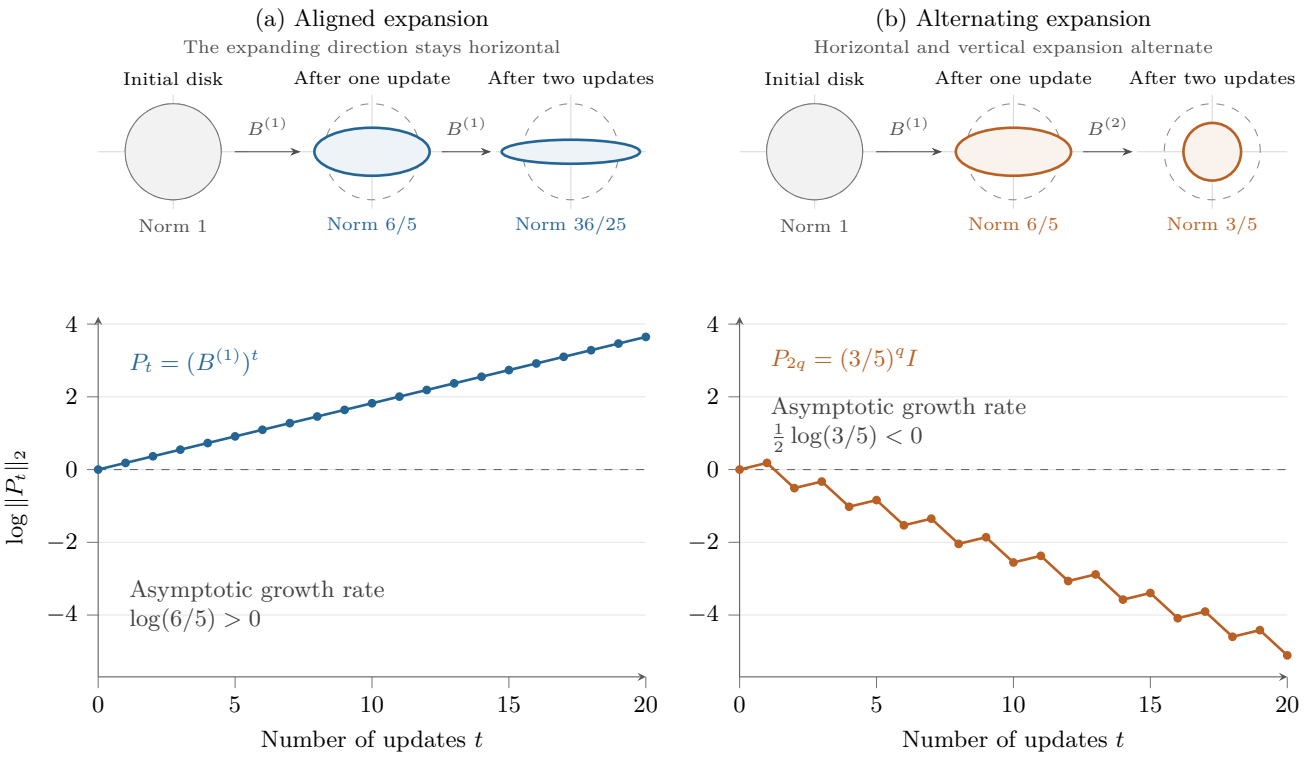


Figure 10: Alignment and the growth of matrix products. The schematic matrices $B^{(1)} = \mathrm{diag}(6/5, 1/2)$ and $B^{(2)} = \mathrm{diag}(1/2, 6/5)$ have the same eigenvalues and Euclidean operator norm 6/5. Panel (a) repeats $B^{(1)}$, while panel (b) alternates $B^{(1)}$ and $B^{(2)}$. The upper diagrams show images of the unit disk after zero, one, and two updates. Dashed circles mark the original disk. The lower plots show $\log \|P_t\|_2$, where $P_t = B_t \cdots B_1$ and $P_0 = I$. Under aligned expansion, $\|P_t\|_2 = (6/5)^t$. Under alternating expansion, $P_{2q} = (3/5)^q I$, so the product contracts despite each matrix having an expansive direction. This is a deterministic geometric illustration, separate from the Gaussian panel specification.

alignment would deliver the missing lower-bound logic. The numerical evidence below measures the remaining primitive gap between expanding finite blocks and long-run alignment.

4.2.1 Numerical evidence

We use numerical analysis to examine dynamic behavior that the theory does not yet fully characterize. The replication code estimates the top Lyapunov exponent on the reduced state space and compares it with the conjectured large-panel limit $\log \max\{\mu_1(\kappa), \delta\}$. The baseline calibration sets $\rho = 0.5$, $\delta = 0.8$, and $\phi = 0.875$, so $\delta\phi = 0.7$ and the population critical gain is $k_c = 6/13 \approx 0.46154$. Panel (a) of Figure 11 compares the estimated exponents for $n \in \{25, 100, 200, 1000\}$. The conjectured limit is flat where the capital mode governs and rises once the coupled score-capital mode becomes dominant. The finite-panel estimates move toward this benchmark as panel size increases.

Panel (b) isolates the zero crossings, whose differences are difficult to distinguish in the full-range plot. We estimate the crossings on a gain grid with step 0.0005, and the approximate 95% Monte Carlo intervals use joint resampling of the log-gain batches to preserve dependence across gains. The estimated crossing falls from 0.47628 at $n = 25$ to 0.46179 at $n = 1000$. Over these panel sizes, the estimated finite-panel

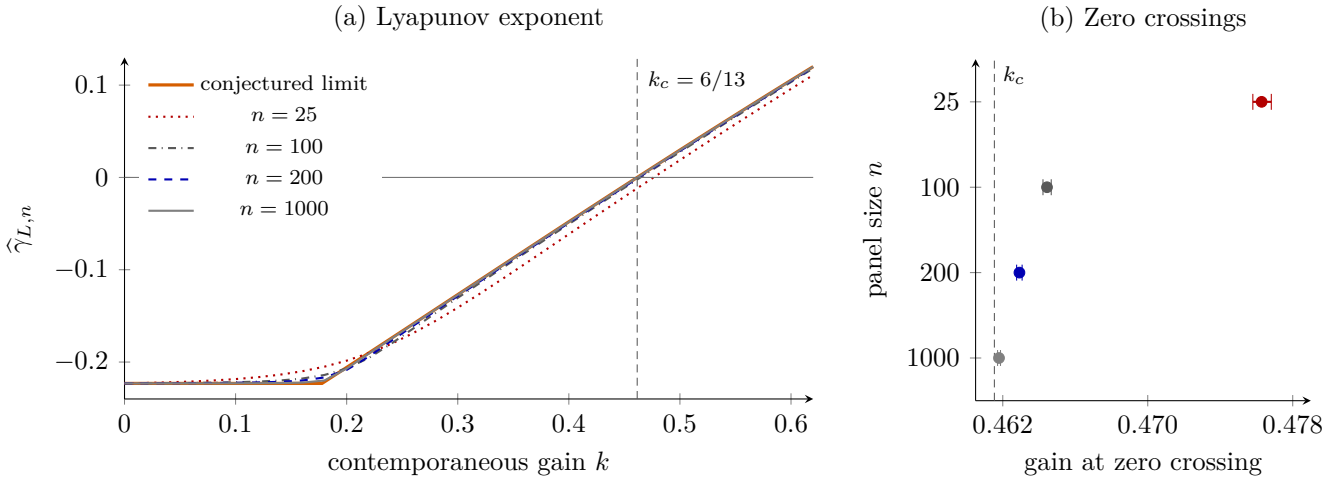


Figure 11: Lyapunov estimates and the population frontier under $\rho = 0.5$, $\delta = 0.8$, and $\phi = 0.875$. Panel (a) plots estimated exponents and the conjectured limit $\log \max\{\mu_1(\kappa), \delta\}$. Each estimate uses four runs of 200,000 retained updates after 5,000 direction updates. Panel (b) shows zero crossings estimated on a 0.0005 gain grid, with approximate 95% Monte Carlo intervals from joint resampling of the log-gain batches. The dashed vertical line marks the exact population frontier $k_c = 6/13$. The estimated crossings approach that frontier from above over the panel sizes studied. These results provide numerical evidence for the conjecture, without establishing the missing long-run lower bound.

frontier therefore approaches the population frontier from above.

At the population frontier, the running exponent estimates for $n = 200$ remain slightly negative over the simulated horizons. Separate path experiments at $k = 0.50$ show rapid expansion, while the experiments at $k = k_c$ reveal substantial dispersion and persistent differences across panel sizes. Appendix B reports the full calibration, simulation settings, Monte Carlo errors, hitting times, and medians and upper quantiles of coefficient magnitudes. These findings support the conjecture but do not establish the matching long-run lower bound. The limiting behavior at the population frontier and the alignment needed for a full positive-side theorem remain open.

4.3 Falsifiable restrictions of the benchmark models

The theory produces several restrictions that can fail in data. A rejection should be interpreted as evidence against the maintained benchmark, not as evidence that performative feedback is absent. Table 2 summarizes the most direct diagnostics.

These restrictions distinguish the theory from a generic claim that coefficients or fit may drift over time. The population model restricts ordering. The deterministic-regime and nonlinear-archive results restrict path and transition differences without requiring the latent fundamental to be observed. The static archive model restricts the entire one-step law of model vintages. The stateful model predicts that departures from that law should be organized by an inherited state rather than by arbitrary coefficient instability. The Internet Appendix IA.2 records an additional no-rotation restriction for multivariate Gaussian and elliptical linear scores.

Table 2: Falsifiable restrictions

The table summarizes testable implications of the model's identification and retraining results under their respective maintained assumptions. A systematic violation provides evidence against the corresponding joint benchmark, but need not identify which assumption fails. Rank comparisons require an independent benchmark for posterior fundamentals, while comparisons of archive transition differences across feature values require overlapping score support. The final row describes a diagnostic implication of stateful feedback. Persistent dependence on older model vintages can motivate this extension, but additional evidence is needed to attribute that persistence to an inherited allocation state.

Benchmark	**Restriction**	**Interpretation of a systematic violation**
Stable scalar feedback	The stable score is a strictly increasing transformation of the posterior mean.	With an independent fundamental benchmark or holdout, rank reversals reject the one-index stable-feedback representation.
Multiple deterministic regimes	Edge labels satisfy cycle restrictions: cross-regime outcome differences must sum to zero around every cycle of the deployment graph.	Nonzero cycle sums reject a common feedback function or a regime-invariant fundamental environment.
Nonlinear model archive	At fixed x, $G_x(s) - G_x(s_0) = f(s) - f(s_0)$, and the same difference must agree across feature values with overlapping archive support.	Transition differences that vary with x reject a common scalar feedback response or conditionally mean-zero archive disturbances.
Static fresh-sample archive	$\lambda_{t+1} = \rho + k\lambda_t + U_t$, with the innovation independent of λ_t and design-scaled as in Theorem 6.	Systematic older-vintage dependence or state-dependent innovations reject the static linear-Gaussian archive model.
Stateful retraining	Older deployments affect current retraining through an inherited state, and stability is governed by the Lyapunov exponent of the joint recurrence.	Rejection of the static AR(1) together with persistent lag dependence is consistent with stateful feedback, but does not by itself identify the durable-state mechanism.

5 Conclusion

A model can fit deployment outcomes more closely without learning more about the underlying fundamentals. The structure theorem explains how this can happen. Under the maintained feedback bound, the stable score is a strictly increasing transformation of the posterior mean. It preserves rankings and fundamental information and remains calibrated to the outcomes generated under its own deployment. Feedback can nevertheless change measured fit by changing the dispersion of the predictable component of those outcomes. Measured accuracy is therefore partly a consequence of how the prediction is used.

Model archives can help separate these effects. Evaluating successive vintages on fixed audit features holds the informational state constant while predictions change. Under conditionally mean-zero archive disturbances, differences in the transition mean identify nonlinear feedback differences exactly. When estimation bias depends on the current score, the comparison also includes the change in bias, and bounds on that change give bounds on feedback. Around a stable predictor, the local propagation of a perturbation is governed by the feedback slope. The affine benchmark makes this relationship exact at every scale.

Policy variation provides a complementary route to identification. Comparisons across deterministic regimes recover feedback differences when the same feature values receive different scores and the fundamental environment is unchanged. A connected deployment graph identifies the decomposition up to a common level normalization. Randomized holdouts provide a direct benchmark, while threshold rules identify feedback jumps under continuity of the score-conditional fundamental mean. Once the decomposition is identified, a new allocation rule can be evaluated before outcomes under that rule are observed. This calculation

requires unchanged posterior fundamentals and a specified counterfactual response over the relevant score range. A consistent change in the level normalization leaves the resulting predictions unchanged.

The retraining results show why the learning rule also matters. In the linear-Gaussian benchmark, fresh-sample retraining produces an exact autoregression whose persistence coefficient is the feedback gain. For $|k| < 1$, it has a unique stationary distribution symmetric about the population fixed point. Pooling past observations instead yields convergence for every $k < 1$, including arbitrarily strong negative feedback. At $k = 1$, pooled coefficients grow logarithmically under positive informational drift, and for $k > 1$ they diverge polynomially. Pooling can therefore stabilize retraining under strong negative feedback while only slowing divergence under strong positive feedback.

Durable allocations carry the effects of earlier predictions into later training outcomes. The stability then depends on the products of random matrices. A negative top Lyapunov exponent yields a unique causal stationary and ergodic solution, and this condition holds for sufficiently large panels whenever the population economy lies strictly below the stability frontier. Above the frontier, suitable finite blocks expand in expected log norm and with probability approaching one as the panel grows. Establishing a positive long-run exponent requires further control of how the expanding directions align across successive blocks. Deriving that control from the primitive Gaussian model remains open, as does the exact critical case.

For applications, fit and calibration should be interpreted together with the allocation rule and the retraining process. Policy comparisons and model archives can distinguish what a prediction reveals about a fundamental from what its deployment causes. That distinction is essential both for evaluating measured accuracy and for predicting outcomes under a different allocation rule.

References

Ball, Ian. 2025. Scoring strategic agents. *American Economic Journal: Microeconomics* 17(1).

Bougerol, Philippe, and Nico Picard. 1992. Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20(4): 1714–1730.

Brandt, Andreas. 1986. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18(1): 211–220.

Brown, Gavin, Shlomi Hod, and Iden Kalemaj. 2022. Performative prediction in a stateful world. *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, PMLR 151: 6045–6061.

Duarte, Victor, and Julia Fonseca. 2026. Self-fulfilling credit scores. NBER Working Paper 35508.

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

Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96(1): 187–199.

D'Amour, Alexander, Peng Ding, Avi Feller, Lihua Lei, and Jasjeet Sekhon. 2021. Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics* 221(2): 644–654.

Drusvyatskiy, Dmitriy, and Lin Xiao. 2023. Stochastic optimization with decision-dependent distributions. *Mathematics of Operations Research* 48(2): 954–998.

Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Evans, George W., and Seppo Honkapohja. 2001. *Learning and Expectations in Macroeconomics*. Princeton, NJ: Princeton University Press.

Farina, Gabriele, and Juan Carlos Perdomo. 2026a. The stability of online algorithms in performative prediction. arXiv:2602.24207.

Farina, Gabriele, and Juan Carlos Perdomo. 2026b. Stabilizing performative feedback loops with minimal model deployments. arXiv:2609.14065.

Frankel, Alex, and Navin Kartik. 2022. Improving information from manipulable data. *Journal of the European Economic Association* 20(1): 79–115.

Furstenberg, Harry, and Harry Kesten. 1960. Products of random matrices. *Annals of Mathematical Statistics* 31(2): 457–469.

Hahn, Jinyong. 1998. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* 66(2): 315–332.

Haidar-Wehbe, Sami, Samuel R. Emerson, Louis J. M. Aslett, and James Liley. 2025. Optimal sizing of a holdout set for safe predictive model updating. *Annals of Applied Statistics* 19(2).

Hardt, Moritz, and Celestine Mendler-Dünner. 2025. Performative prediction: Past and future. *Statistical Science* 40(3): 417–436.

Hardt, Moritz. 2026. Retraining seeks stable signals. arXiv:2607.15623.

Hennessy, Christopher A., and Charles A. E. Goodhart. 2023. Goodhart's law and machine learning: A structural perspective. *International Economic Review* 64(3): 1075–1086.

Izzo, Zachary, Lexing Ying, and James Zou. 2021. How to learn when data reacts to your model: Performative gradient descent. *Proceedings of the 38th International Conference on Machine Learning*, PMLR 139: 4641–4650.

Izzo, Zachary, James Zou, and Lexing Ying. 2022. How to learn when data gradually reacts to your model. *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, PMLR 151: 3998–4035.

Kim, Michael P., and Juan C. Perdomo. 2023. Making decisions under outcome performativity. *Proceedings of the 14th Innovations in Theoretical Computer Science Conference*.

Kingman, J. F. C. 1973. Subadditive ergodic theory. *Annals of Probability* 1(6): 883–899.

Liley, James, Samuel R. Emerson, Bilal A. Mateen, Catalina A. Vallejos, Louis J. M. Aslett, and Sebastian J. Vollmer. 2021. Model updating after interventions paradoxically introduces bias. *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, PMLR 130: 3916–3924.

Lee, Jaewook, and Tijana Zrnic. 2026. Partially performative prediction. arXiv:2606.07890.

Li, Qiang, and Hoi-To Wai. 2022. State dependent performative prediction with stochastic approximation. *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, PMLR 151: 3164–3186.

Livdan, Dmitry, Bernardo Maciel, and Norman Schürhoff. 2026. Manufactured fit: AI allocation and the Lucas critique. Working paper.

Marcet, Albert, and Thomas J. Sargent. 1989. Convergence of least squares learning mechanisms in self-referential linear stochastic models. *Journal of Economic Theory* 48(2): 337–368.

Mendler-Dünner, Celestine, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. 2020. Stochastic optimization for performative prediction. *Advances in Neural Information Processing Systems* 33: 4929–4939.

Mendler-Dünner, Celestine, Frances Ding, and Yixin Wang. 2022. Anticipating performativity by predicting from predictions. *Advances in Neural Information Processing Systems* 35: 31171–31185.

Miller, John P., Juan C. Perdomo, and Tijana Zrnic. 2021. Outside the echo chamber: Optimizing the performative risk. *Proceedings of the 38th International Conference on Machine Learning*, PMLR 139: 7710–7720.

Munro, Evan. 2025. Treatment allocation with strategic agents. *Management Science* 71(1).

Pemantle, Robin. 1990. Nonconvergence to unstable points in urn models and stochastic approximations. *Annals of Probability* 18(2): 698–712.

Perdomo, Juan C. 2025. Revisiting the predictability of performative, social events. *Proceedings of the 42nd International Conference on Machine Learning*.

Perdomo, Juan, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. 2020. Performative prediction. *Proceedings of the 37th International Conference on Machine Learning*: 7599–7609.

Ray, Mitas, Lillian J. Ratliff, Dmitriy Drusvyatskiy, and Maryam Fazel. 2022. Decision-dependent risk minimization in geometrically decaying dynamic environments. *Proceedings of the AAAI Conference on Artificial Intelligence* 36(7): 8081–8088.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41–55.

Sargent, Thomas J. 1999. *The Conquest of American Inflation*. Princeton, NJ: Princeton University Press.

Scarone, Bruno, and Ricardo Baeza-Yates. 2026. The illusion of improvement: Reject inference strategies in credit scoring. arXiv:2606.18479.

Appendix

Appendix A Proofs

Proof of Theorem 1. Part (i). Let $c = 1 - \kappa^+ > 0$. Since $h(t) = t - f(t)$, Assumption 1 gives $h'(t) = 1 - f'(t) \geq c$ for every t. For any $b > a$, the mean-value theorem therefore gives $h(b) - h(a) \geq c(b - a) > 0$. Thus h is strictly increasing.

Taking $a = 0$ shows that $h(t) \geq h(0) + ct$ for $t > 0$. Taking $b = 0$ shows that $h(t) \leq h(0) + ct$ for $t < 0$. Consequently, $h(t)$ tends to $+\infty$ as t tends to $+\infty$ and to $-\infty$ as t tends to $-\infty$. Because h is continuous and strictly increasing, it is a bijection from $\mathbb{R}$ onto $\mathbb{R}$.

Let $\Phi = h^{-1}$. It is strictly increasing. Since h is differentiable with $h' > 0$ everywhere, the inverse derivative rule gives $\Phi'(u) = 1/h'(\Phi(u)) = 1/[1 - f'(\Phi(u))] = 1/[1 - k(\Phi(u))] > 0$, proving (9).

For $u_2 > u_1$, set $t_j = \Phi(u_j)$. Then $t_2 > t_1$, and the mean-value bound established above gives $u_2 - u_1 = h(t_2) - h(t_1) \geq c(t_2 - t_1)$. Hence $\Phi(u_2) - \Phi(u_1) \leq (u_2 - u_1)/c$. Reversing the order of the arguments proves (10).

Part (ii). Stability requires $m(x) = (Tm)(x) = u(x) + f(m(x))$ for almost every x. Equivalently, $h(m(x)) = u(x)$. Because h is bijective, this equation has the unique solution $m(x) = \Phi(u(x))$. Conversely, $h(\Phi(u(x))) = u(x)$ implies $\Phi(u(x)) = u(x) + f(\Phi(u(x)))$, so this solution is stable.

The solution is admissible for the retraining operator. Indeed, (10) implies $|\Phi(u(x))| \leq |\Phi(0)| + |u(x)|/c$. The posterior mean $u(x)$ has a finite second moment by Jensen's inequality. Thus $m^* = \Phi(u)$ has a finite second moment, as does $f(m^*) = m^* - u$. Uniqueness holds up to almost-sure equality, as required by the definition of stability.

Part (iii). Suppose $\kappa_{\text{abs}} < 1$. For any $a, b \in \mathbb{R}$, the mean-value theorem gives $|f(a) - f(b)| \leq \kappa_{\text{abs}}|a - b|$. Thus, for each fixed x, the map $t \mapsto u(x) + f(t)$ is a contraction with fixed point $m^*(x)$. Since $m_{n+1}(x) = u(x) + f(m_n(x))$, it follows that $|m_{n+1}(x) - m^*(x)| \leq \kappa_{\text{abs}}|m_n(x) - m^*(x)|$. Iterating proves (12) and hence pointwise convergence.

The recursion remains within the domain of T. In fact, $|f(t)| \leq |f(0)| + \kappa_{\text{abs}}|t|$, so a score with a finite second moment is mapped to another score with a finite second moment. □

Proof of Corollary 1. Fix $u_2 > u_1$ and set $t_j = \Phi(u_j)$. Theorem 1 implies $t_2 > t_1$. By the mean-value theorem and $0 \leq f' \leq \kappa^+$, we have $0 \leq f(t_2) - f(t_1) \leq \kappa^+(t_2 - t_1)$. Therefore

$$(1 - \kappa^+)(t_2 - t_1) \leq h(t_2) - h(t_1) \leq t_2 - t_1.$$

Since $h(t_j) = u_j$ and $t_j = \Phi(u_j)$, rearranging these inequalities gives (13). □

Proof of Corollary 2. Theorem 1 gives $m^*(x) = \Phi(u(x))$, where Φ is a continuous strictly increasing bijection of $\mathbb{R}$. Because Φ is Borel measurable, $m^*(x)$ is measurable with respect to $\sigma(u(x))$. Therefore $\sigma(m^*(x)) \subseteq \sigma(u(x))$. Since $\Phi^{-1} = h$ is also continuous and Borel measurable and $u(x) = h(m^*(x))$, $u(x)$ is measurable with respect to $\sigma(m^*(x))$. Therefore $\sigma(u(x)) \subseteq \sigma(m^*(x))$. Combining the two inclusions gives equation (14).

For part (i), conditional distributions with respect to two equal sigma-algebras are versions of the same regular conditional distribution. Since $\sigma(m^*(x)) = \sigma(u(x))$, the conditional law of θ given $m^*(x)$ equals the conditional law of θ given $u(x)$ almost surely.

For part (ii), the equality of sigma-algebras already implies that the two random variables contain the same information about θ. To express this through mutual information, note first that $m^* = \Phi(u)$ is a deterministic function of u, so the data-processing inequality gives $I(m^*; \theta) \leq I(u; \theta)$. Conversely,

$u = h(m^*)$ is a deterministic function of m^*, so the data-processing inequality gives $I(u;\theta) \le I(m^*;\theta)$. The two inequalities imply equality, which is equation (15).

For part (iii), for any pair of realizations x_1, x_2, strict monotonicity of Φ gives $m^*(x_1) < m^*(x_2)$ if and only if $u(x_1) < u(x_2)$, with equality preserved as well. Thus every pairwise ordering generated by m^* is exactly the ordering generated by u. Any statistic that is invariant to a strictly increasing transformation of the score depends on the score only through this preserved ordering and its ties. Therefore, every such statistic is unchanged. Spearman rank correlation and Kendall's tau are particular examples. □

Proof of Corollary 3. Performative stability under nonparametric squared-loss retraining means $m^*(x) = (Tm^*)(x) = \mathbb{E}[y \mid x]$ almost surely. Since $m^*(x)$ is measurable with respect to x, the sigma-algebra $\sigma(m^*(x))$ is contained in $\sigma(x)$. The tower property therefore gives $\mathbb{E}[y \mid m^*(x)] = \mathbb{E}[\mathbb{E}[y \mid x] \mid m^*(x)]$. Substituting $\mathbb{E}[y \mid x] = m^*(x)$ yields $\mathbb{E}[y \mid m^*(x)] = \mathbb{E}[m^*(x) \mid m^*(x)]$. A random variable is measurable with respect to its own sigma-algebra, so the final conditional expectation equals $m^*(x)$. This proves equation (16). □

Proof of Lemma 1. Under equation (3), the realized outcome is $y = \theta + f(m(x)) + \varepsilon$. Conditional on x, the quantity $f(m(x))$ is deterministic because $m(x)$ is measurable in x. Also $\mathbb{E}[\varepsilon \mid x] = \mathbb{E}[\mathbb{E}[\varepsilon \mid \theta, x] \mid x] = 0$ by the maintained condition $\mathbb{E}[\varepsilon \mid \theta, x] = 0$. Therefore $\mathbb{E}[y \mid x] = \mathbb{E}[\theta \mid x] + f(m(x)) + \mathbb{E}[\varepsilon \mid x] = u(x) + f(m(x))$.

Subtracting this conditional mean from y gives $y - \mathbb{E}[y \mid x] = \theta + f(m(x)) + \varepsilon - u(x) - f(m(x)) = \theta - u(x) + \varepsilon$. This is an identity, so conditional on x the two random variables have exactly the same law. The right-hand side contains neither f nor m, which proves the law-invariance claim.

For the variance statement, conditional on x we have $\mathrm{Var}(y \mid x) = \mathrm{Var}(\theta - u(x) + \varepsilon \mid x)$. Since $u(x)$ is constant conditional on x, $\mathrm{Var}(\theta - u(x) + \varepsilon \mid x) = \mathrm{Var}(\theta + \varepsilon \mid x)$. Expanding the variance gives $\mathrm{Var}(\theta + \varepsilon \mid x) = \mathrm{Var}(\theta \mid x) + \mathrm{Var}(\varepsilon \mid x) + 2\mathrm{Cov}(\theta, \varepsilon \mid x)$. We now show that the conditional covariance is zero. First, $\mathbb{E}[\theta\varepsilon \mid x] = \mathbb{E}[\mathbb{E}[\theta\varepsilon \mid \theta, x] \mid x] = \mathbb{E}[\theta\mathbb{E}[\varepsilon \mid \theta, x] \mid x] = 0$. Second, $\mathbb{E}[\varepsilon \mid x] = 0$. Hence $\mathrm{Cov}(\theta, \varepsilon \mid x) = \mathbb{E}[\theta\varepsilon \mid x] - \mathbb{E}[\theta \mid x]\mathbb{E}[\varepsilon \mid x] = 0 - u(x) \cdot 0 = 0$. Therefore $\mathrm{Var}(y \mid x) = \mathrm{Var}(\theta \mid x) + \mathrm{Var}(\varepsilon \mid x)$. Taking unconditional expectations gives the expression for σ^2_{res}. □

Proof of Theorem 2. Write $U = u(x)$. Conditional Jensen's inequality gives $\mathbb{E}[U^2] = \mathbb{E}[\mathbb{E}[\theta \mid x]^2] \le \mathbb{E}[\mathbb{E}[\theta^2 \mid x]] = \mathbb{E}[\theta^2] < \infty$. Because f is differentiable and nondecreasing, $f'(t) \ge 0$ for every t. The strength bound implies $1 - \tau f'(t) \ge 1 - \overline{\tau} \sup_{r \in \mathbb{R}} f'(r) > 0$ for every $\tau \in [0, \overline{\tau}]$ and $t \in \mathbb{R}$. Consequently, Φ_τ exists and is Lipschitz with constant at most $[1 - \overline{\tau} \sup_r f'(r)]^{-1}$. In particular, $|m^*_\tau(x)| = |\Phi_\tau(U)| \le |\Phi_\tau(0)| + |U|/[1 - \overline{\tau} \sup_r f'(r)]$. Each stable score therefore has a finite second moment. Its fixed-point equation gives $\tau f(m^*_\tau(x)) = m^*_\tau(x) - U$, so the outcome in each regime also has a finite second moment.

Fix $0 \le \tau_1 < \tau_2 \le \overline{\tau}$. Differentiating $\Phi_\tau(v) - \tau f(\Phi_\tau(v)) = v$ with respect to v gives $[1 - \tau f'(\Phi_\tau(v))]\Phi'_\tau(v) = 1$. The first factor is positive, so $\Phi'_\tau(v) = [1 - \tau f'(\Phi_\tau(v))]^{-1}$. Monotone amplification says that $\tau_2 f'(\Phi_{\tau_2}(v)) \ge \tau_1 f'(\Phi_{\tau_1}(v))$ for every v. The positive denominators in the derivative formula therefore satisfy $\Phi'_{\tau_2}(v) \ge \Phi'_{\tau_1}(v)$.

Define $d(v) = \Phi_{\tau_2}(v) - \Phi_{\tau_1}(v)$. Since $d'(v) \ge 0$, the mean value theorem implies that d is nondecreasing. If $v_2 > v_1$, then $d(v_2) \ge d(v_1)$, which gives $\Phi_{\tau_2}(v_2) - \Phi_{\tau_2}(v_1) \ge \Phi_{\tau_1}(v_2) - \Phi_{\tau_1}(v_1)$. Both inverse transformations are increasing. Exchanging v_1 and v_2 therefore shows that $|\Phi_{\tau_2}(v_2) - \Phi_{\tau_2}(v_1)| \ge |\Phi_{\tau_1}(v_2) - \Phi_{\tau_1}(v_1)|$ for every pair v_1, v_2.

Let U' be an independent copy of U. For any square-integrable Z and an independent copy Z', expansion of the square gives $\mathbb{E}[(Z - Z')^2] = 2\mathbb{E}[Z^2] - 2\mathbb{E}[Z]^2 = 2\mathrm{Var}(Z)$. Apply this identity to $Z = \Phi_{\tau_j}(U)$ for $j = 1, 2$. The pairwise inequality remains true after squaring and evaluating it at (U, U'). Taking expectations gives $2\mathrm{Var}(m^*_{\tau_2}(x)) = \mathbb{E}[(\Phi_{\tau_2}(U) - \Phi_{\tau_2}(U'))^2] \ge \mathbb{E}[(\Phi_{\tau_1}(U) - \Phi_{\tau_1}(U'))^2] = 2\mathrm{Var}(m^*_{\tau_1}(x))$. Hence, stable-score variance is nondecreasing in τ.

Lemma 1 shows that σ^2_{res} is constant across regimes. For $v + \sigma^2_{\mathrm{res}} > 0$, the derivative of $v/(v + \sigma^2_{\mathrm{res}})$ with respect to v is $\sigma^2_{\mathrm{res}}/(v + \sigma^2_{\mathrm{res}})^2 \ge 0$. At $\tau = 0$, $m^*_0(x) = U$. The condition $\mathrm{Var}(U) + \sigma^2_{\mathrm{res}} > 0$, together with

the variance inequality, makes the denominator defining $R^2(\tau)$ positive in every regime. It follows that $R^2(\tau)$ is nondecreasing.

Suppose first that $f(t) = a+kt$ with $k \geq 0$. Then $f'(t) = k$, so the effective local gain is $\tau f'(\Phi_\tau(v)) = \tau k$. This is nondecreasing in τ, which verifies monotone amplification. The fixed-point equation is $m^*_\tau(x) = U + \tau(a + km^*_\tau(x))$. Moving the term involving the stable score to the left gives $(1-\tau k)m^*_\tau(x) = U + \tau a$. The strength bound implies $1 - \tau k > 0$, so $m^*_\tau(x) = (U + \tau a)/(1 - \tau k)$. The term $\tau a/(1 - \tau k)$ is constant across units. Taking the variance therefore gives $\mathrm{Var}(m^*_\tau(x)) = \mathrm{Var}(U)/(1 - \tau k)^2$. If $k > 0$ and $\mathrm{Var}(U) > 0$, its derivative with respect to τ is $2k\mathrm{Var}(U)/(1 - \tau k)^3 > 0$. If also $\sigma^2_{\mathrm{res}} > 0$, the map $v \mapsto v/(v + \sigma^2_{\mathrm{res}})$ is strictly increasing. Thus $R^2(\tau)$ is strictly increasing.

Suppose next that f is convex, nonnegative, and nondecreasing. Fix v and write $t_j = \Phi_{\tau_j}(v)$ for $j = 1, 2$. Since $t_1 - \tau_1 f(t_1) = v$, we have $t_1 - \tau_2 f(t_1) = v - (\tau_2 - \tau_1)f(t_1) \leq v$. The function $t \mapsto t - \tau_2 f(t)$ is strictly increasing and takes the value v at t_2. Hence $t_2 \geq t_1$. Convexity implies that f' is nondecreasing, while monotonicity of f implies that $f' \geq 0$. Consequently, $\tau_2 f'(t_2) \geq \tau_2 f'(t_1) \geq \tau_1 f'(t_1)$. This verifies monotone amplification.

Finally, Φ_τ is an increasing bijection and $m^*_\tau(x) = \Phi_\tau(U)$ in every regime. Since $U = \Phi^{-1}_\tau(m^*_\tau(x))$, the stable score and posterior mean generate the same sigma-algebra. This proves the information claim. Stability also gives $\mathbb{E}[y \mid x] = m^*_\tau(x)$ in each regime. Iterated expectations then give $\mathbb{E}[y \mid m^*_\tau(x)] = \mathbb{E}[\mathbb{E}[y \mid x] \mid m^*_\tau(x)] = \mathbb{E}[m^*_\tau(x) \mid m^*_\tau(x)] = m^*_\tau(x)$, which proves calibration. □

Proof of Corollary 4. Under affine feedback $f(s) = a + ks$, the stable-point equation is $m^*(x) = u(x) + a + km^*(x)$. Move the term $km^*(x)$ to the left $(1-k)m^*(x) = u(x) + a$. Because $k < 1$, the coefficient $1 - k$ is strictly positive. Dividing gives $m^*(x) = \frac{u(x)+a}{1-k}$. The term $a/(1 - k)$ is constant across units and therefore has zero variance. Hence

$$\mathrm{Var}(m^*(x)) = \mathrm{Var}\left(\frac{u(x)}{1-k}\right) = \frac{\mathrm{Var}(u(x))}{(1-k)^2},$$

which proves equation (24).

At $k = 0$, the variance equals $\mathrm{Var}(u)$. If $0 < k < 1$, then $0 < 1 - k < 1$, so $(1 - k)^2 < 1$ and therefore $\mathrm{Var}(m^*) > \mathrm{Var}(u)$ whenever $\mathrm{Var}(u) > 0$. If $k < 0$, then $1 - k > 1$, so $(1 - k)^2 > 1$ and $\mathrm{Var}(m^*) < \mathrm{Var}(u)$ whenever $\mathrm{Var}(u) > 0$.

By Lemma 1, the residual variance σ^2_{res} is independent of k. Deployment fit is $R^2 = V_m/(V_m + \sigma^2_{\mathrm{res}})$ with $V_m = \mathrm{Var}(m^*)$. The map $v \mapsto v/(v + \sigma^2_{\mathrm{res}})$ has derivative $\sigma^2_{\mathrm{res}}/(v + \sigma^2_{\mathrm{res}})^2$, which is strictly positive when $\sigma^2_{\mathrm{res}} > 0$. Thus, the ordering of R^2 follows the ordering of $\mathrm{Var}(m^*)$. Corollary 2 and Corollary 3 apply for every $k < 1$ satisfying the maintained assumptions, so the information and calibration claims follow. □

Proof of Theorem 3. Let $S = (\mu, f)$ be any structure consistent with the observed deployment law P, and let c satisfy the conditions in the theorem. Define $\theta_c = \theta - c(m(x))$ and $f_c = f + c$. We verify each required property.

First, the transformed structure generates the same outcome pathwise. Under S_c, the outcome is $y_c = \theta_c + f_c(m(x)) + \varepsilon$. Substituting the definitions gives $y_c = [\theta - c(m(x))] + [f(m(x)) + c(m(x))] + \varepsilon = \theta + f(m(x)) + \varepsilon = y$. Thus, for every realization of the primitive variables, the transformed outcome equals the original outcome. Since x is unchanged, the joint law of (x, y) is unchanged. Therefore, S_c is observationally equivalent to S.

Second, the transformed structure satisfies the setup moment conditions. By the inequality $(a - b)^2 \leq 2a^2 + 2b^2$, $\mathbb{E}[\theta_c^2] = \mathbb{E}[(\theta - c(m(x)))^2] \leq 2\mathbb{E}[\theta^2] + 2\mathbb{E}[c(m(x))^2] < \infty$. The law of x is unchanged. To verify the conditional-mean restriction on the outcome shock, note that $\theta = \theta_c + c(m(x))$. Hence, conditional on (θ_c, x), the original fundamental θ is known as a measurable function of (θ_c, x). Conversely, $\theta_c = \theta - c(m(x))$ is measurable in (θ, x). Therefore $\sigma(\theta_c, x) = \sigma(\theta, x)$. It follows that $\mathbb{E}[\varepsilon \mid \theta_c, x] = \mathbb{E}[\varepsilon \mid \theta, x] = 0$.

Third, compute the posterior mean under the transformed structure. Since $c(m(x))$ is measurable with respect to x, $u_c(x) = \mathbb{E}[\theta_c \mid x] = \mathbb{E}[\theta - c(m(x)) \mid x] = \mathbb{E}[\theta \mid x] - c(m(x)) = u(x) - c(m(x))$. This proves the

transformation in equation (26).

Now suppose that m is injective on the support of x. The support is a Borel subset of the Polish space $\mathbb{X}$. The restriction of the Borel map m to this Borel set is Borel and injective. By the Lusin–Souslin theorem, its image $m(\operatorname{supp} x)$ is Borel and the inverse map m^{-1} on that image is Borel. Hence for any measurable function $g(x)$ on the support, define $c_g(s) = g(m^{-1}(s))$ for s in the image. Then c_g is measurable and $c_g(m(x)) = g(x)$ on the support. Therefore, any measurable component of the conditional mean can be shifted between u and $f \circ m$, subject only to the maintained integrability restriction. This proves the unrestricted decomposition claim in part (i).

Part (ii) follows directly from observational equivalence. A functional computed solely from the joint law P has the same value under any two structures that generate the same P. Since S and S_c generate exactly the same law of (x, y), all calibration statistics, R^2 values, proper scoring rules, and other deployment-data functionals are identical across the transformation orbit.

For part (iii), define the orbit $\mathcal{O}(S) = \{S_c : c \text{ admissible}\}$. The first part of the proof shows that $\mathcal{O}(S)$ is contained in the identified set. Let $G(S)$ be a structural functional. If there exist two admissible functions c_1, c_2 for which $G(S_{c_1}) \neq G(S_{c_2})$, then the observed law is compatible with two different values of G. Hence G is not identified. Invariance under every admissible shift is therefore necessary for identification.

We now show explicitly that fundamental-information quantities need not be invariant. Consider the linear-Gaussian benchmark with $0 < \rho < 1$ and deploy $m = u$. Write the fundamental as $\theta = u + r$, where $r = \theta - u$ satisfies $\mathbb{E}[r \mid x] = 0$. Under joint Gaussianity, u and r are independent. Let $\sigma_u^2 = \mathrm{Var}(u) > 0$ and $\sigma_r^2 = \mathrm{Var}(r) > 0$. Choose $c(m) = m/2 = u/2$. Then $\theta_c = \theta - u/2 = u/2 + r$. Since (u, θ) is jointly Gaussian, $I(u;\theta) = \frac{1}{2}\log[\mathrm{Var}(\theta)/\mathrm{Var}(r)] = \frac{1}{2}\log[1 + \sigma_u^2/\sigma_r^2]$. Likewise $I(u;\theta_c) = \frac{1}{2}\log[\mathrm{Var}(\theta_c)/\mathrm{Var}(r)] = \frac{1}{2}\log[1 + (\sigma_u^2/4)/\sigma_r^2]$. Since $\sigma_u^2 > 0$, $1 + \sigma_u^2/(4\sigma_r^2) < 1 + \sigma_u^2/\sigma_r^2$, so $I(u;\theta_c) < I(u;\theta)$. Hence, mutual information with the latent fundamental is not identified in general.

Finally, the shift orbit need not exhaust the identified set. Let x be uniform on $\{-1, 1\}$, let $z \sim N(0, 1)$ be independent of x, and let the observed outcome be $y = x + z$. Take m and f to be constant. Structure A sets $\theta_A = x + z$ and $\varepsilon_A = 0$. Structure B sets $\theta_B = x$ and $\varepsilon_B = z$. In both structures, $y = \theta + f(m) + \varepsilon$ after absorbing the common constant into f, and both satisfy $\mathbb{E}[\varepsilon \mid \theta, x] = 0$. Both have posterior mean $u(x) = x$. Yet $\theta_A - \theta_B = z$ is random conditional on x, while every shift $c(m(x))$ is constant because m is constant. Hence no deterministic score-dependent shift maps A into B. The identified set can therefore be strictly larger than the shift orbit. This establishes every claim in part (iii). □

Proof of Theorem 4. For every $x \in \mathbb{X}_0$ and regime r, the maintained conditional-mean relation gives $g_r(x) = u(x) + f(m_r(x))$. Subtracting the relations for regimes r and q yields $g_r(x) - g_q(x) = f(m_r(x)) - f(m_q(x))$. Thus, an observed edge label equals the difference in feedback values at its endpoints.

Part (i). Fix a connected component $\mathcal{C}$ and scores $s_0, s \in \mathcal{C}$. There is a finite path $s_0, s_1, \ldots, s_L = s$. For each edge from $s_{\ell-1}$ to s_ℓ, choose $x_\ell \in \mathbb{X}_0$ and regimes r_ℓ, q_ℓ such that $s_\ell = m_{r_\ell}(x_\ell)$ and $s_{\ell-1} = m_{q_\ell}(x_\ell)$. The observed label on this edge is $d_\ell = g_{r_\ell}(x_\ell) - g_{q_\ell}(x_\ell) = f(s_\ell) - f(s_{\ell-1})$. Summing along the path gives

$$\sum_{\ell=1}^{L} d_\ell = \sum_{\ell=1}^{L} [f(s_\ell) - f(s_{\ell-1})] = f(s) - f(s_0).$$

The intermediate values cancel, so the data determine every feedback difference within $\mathcal{C}$.

Now let $(\widetilde{u}, \widetilde{f})$ be another decomposition that generates the same conditional means on $\mathbb{X}_0$, and set $d(s) = \widetilde{f}(s) - f(s)$ for $s \in \mathcal{S}$. If $s = m_r(x)$ and $s' = m_q(x)$ form an edge, equality of the two observed conditional-mean differences gives $\widetilde{f}(s) - \widetilde{f}(s') = f(s) - f(s')$. Hence $d(s) = d(s')$. This equality holds along every path, so d is constant on each connected component.

Part (ii). If $\mathcal{G}$ is connected, the path calculation identifies $f(s) - f(s_0)$ for every $s \in \mathcal{S}$. Fixing $f(s_0)$ therefore determines f on $\mathcal{S}$. For $x \in \mathbb{X}_0$, rearranging the conditional-mean relation gives $u(x) = g_r(x) - f(m_r(x))$, which proves (31). The expression is independent of r: for any two regimes r, q, their recovered values differ

by $[g_r(x) - g_q(x)] - [f(m_r(x)) - f(m_q(x))] = 0$.

Part (iii). Let the components be $\mathcal{C}_1, \ldots, \mathcal{C}_C$. For each $x \in \mathbb{X}_0$, all scores $m_1(x), \ldots, m_R(x)$ belong to one component, because every pair of these scores is joined by an edge. Write $j(x)$ for its index.

Choose any constants $c_1, \ldots, c_C \in \mathbb{R}$. Since C is finite and each component is Borel, the function

$$\widetilde{f}(s) = f(s) + \sum_{j=1}^{C} c_j \mathbb{1}_{\mathcal{C}_j}(s),$$

is measurable on $\mathbb{R}$. Define $b(x) = c_{j(x)}$ on $\mathbb{X}_0$ and set $b(x) = 0$ outside $\mathbb{X}_0$. This is measurable: on $\mathbb{X}_0$, the event $j(x) = j$ is the inverse image of the Borel set $\mathcal{C}_j$ under the known measurable score m_1. Finiteness of C also makes b bounded.

Set $\widetilde{\theta} = \theta - b(x)$ and retain the original outcome shock ε. Because θ has a finite second moment and b is bounded, $\widetilde{\theta}$ has a finite second moment. Moreover, b is measurable with respect to x, so $\sigma(\widetilde{\theta}, x) = \sigma(\theta, x)$ and $\mathbb{E}[\varepsilon \mid \widetilde{\theta}, x] = \mathbb{E}[\varepsilon \mid \theta, x] = 0$. The transformed posterior mean is $\widetilde{u}(x) = \mathbb{E}[\widetilde{\theta} \mid x] = u(x) - b(x)$. For every $x \in \mathbb{X}_0$ and regime r, we have $m_r(x) \in \mathcal{C}_{j(x)}$, and therefore

$$\widetilde{\theta} + \widetilde{f}(m_r(x)) + \varepsilon = \theta - b(x) + f(m_r(x)) + c_{j(x)} + \varepsilon = \theta + f(m_r(x)) + \varepsilon.$$

The two structures generate the same outcome for every regime on the common full-measure set $\mathbb{X}_0$. They consequently generate the same observed deployment law. Thus every choice of the C constants is admissible in the unrestricted measurable class.

Conversely, part (i) shows that the feedback responses in any two decompositions consistent with the conditional means differ by a constant c_j on each $\mathcal{C}_j$. Equality of $g_r(x)$ then requires $\widetilde{u}(x) - u(x) = -c_{j(x)}$ for $x \in \mathbb{X}_0$. Componentwise shifts therefore exhaust the ambiguity in the pair $(u, f|_{\mathcal{S}})$. If differentiability or derivative bounds are imposed on f, a shifted response need not satisfy them, and the constants may no longer vary freely. □

Proof of Corollary 5. For an edge from s' to s, the observed label is $\ell(s, s') = f(s) - f(s')$. Reversing the edge changes the sign of its label. Thus, for any finite cycle $s_0, s_1, \ldots, s_L = s_0$,

$$\sum_{j=1}^{L} \ell(s_j, s_{j-1}) = \sum_{j=1}^{L} [f(s_j) - f(s_{j-1})] = f(s_L) - f(s_0) = 0. \tag{A.1}$$

This proves (32).

The sum along any path from a to b equals $f(b) - f(a)$. It therefore depends only on the endpoints, proving part (ii). For part (iii), suppose the graph is finite and connected. A spanning tree has $J - 1$ edges and a unique path from a chosen root s_0 to every vertex. Once $f(s_0)$ is fixed, the labels along these paths determine f at every vertex. A graph with fewer than $J - 1$ distinct edges is disconnected. Its edge labels determine feedback differences within each component but provide no difference between components. One reference value per component suffices for recovery from the edge labels. Under the unrestricted measurable class in Theorem 4(iii), these component levels can vary independently. Additional restrictions on f may constrain them. □

Proof of Proposition 1. At a stable deployment, equation (7) gives $u(x) = m^*(x) - f(m^*(x))$. Let $m_j = m^*(x_j)$ and suppose $m_2 \geq m_1$. Then $u(x_2) - u(x_1) = (m_2 - m_1) - [f(m_2) - f(m_1)]$. If $m_2 = m_1$, both sides of (28) are zero. If $m_2 > m_1$, the mean value theorem gives a point $\xi \in (m_1, m_2)$ such that $f(m_2) - f(m_1) = f'(\xi)(m_2 - m_1)$. The point ξ lies in the convex hull of the range of m^*, so the proposition's assumption implies $0 \leq f(m_2) - f(m_1) \leq \kappa^+(m_2 - m_1)$. Substituting these bounds into the preceding expression for $u(x_2) - u(x_1)$ proves (28).

For the variance bounds, let M and M' be independent copies of $m^*(x)$, and set $U = M - f(M)$ and $U' = M' - f(M')$. Stability implies that U and U' are independent copies of $u(x)$. Apply (28) with the larger of M and M' as m_2. In either ordering, $(1-\kappa^+)|M-M'| \le |U-U'| \le |M-M'|$. All terms are nonnegative, so squaring and taking expectations gives

$$(1-\kappa^+)^2\mathbb{E}[(M-M')^2] \le \mathbb{E}[(U-U')^2] \le \mathbb{E}[(M-M')^2].$$

For any square-integrable random variable Z and an independent copy Z', independence yields $\mathbb{E}[(Z-Z')^2] = 2\mathrm{Var}(Z)$. Applying this identity to M and U, then dividing by two, proves (29). □

Proof of Proposition 2. Because D is independent of (θ, x, ε) and both values of D have positive probability, conditioning on D does not change conditional expectations given x. Moreover, $\mathbb{E}[\varepsilon \mid x] = \mathbb{E}[\mathbb{E}[\varepsilon \mid \theta, x] \mid x] = 0$. Thus, for $d \in \{0,1\}$, $\mathbb{E}[\theta \mid x, D=d] = u(x)$ and $\mathbb{E}[\varepsilon \mid x, D=d] = 0$ for P_X-almost every x.

Part (i). In the holdout group, $y = \theta + f(s_0) + \varepsilon$. Taking the conditional expectation given $(x, D=1)$ yields $\mathbb{E}[y \mid x, D=1] = u(x) + f(s_0)$, proving (34). The joint law identifies this conditional mean P_X-almost surely. Fixing $f(s_0)$ therefore determines $u(x)$ almost surely.

Part (ii). In the deployed group, $y = \theta + f(m(x)) + \varepsilon$, so $\mathbb{E}[y \mid x, D=0] = u(x) + f(m(x))$ almost surely. Subtracting the holdout conditional mean proves (35). Because m is known, the contrast identifies $f(s) - f(s_0)$ for $P_{m(x)}$-almost every s.

If f is continuous, this identification extends to all of $\mathcal{S}_m$. To see why, suppose two continuous functions on $\mathcal{S}_m$ agree $P_{m(x)}$-almost surely but differ at some $s \in \mathcal{S}_m$. By continuity, they differ on a neighborhood of s. Every neighborhood of a support point has positive $P_{m(x)}$ probability, a contradiction. Thus the identified almost-sure values determine the continuous contrast throughout $\mathcal{S}_m$. Fixing $f(s_0)$ then determines f on $\mathcal{S}_m$ and, by part (i), $u(x)$ almost surely.

Part (iii). The identified contrast on $\mathcal{S}_m$ is $s \mapsto f(s) - f(s_0)$. On the interior of I, its derivative is $f'(s)$, so the local gain $k(s)$ is identified there. If $\mathbb{P}(m(x) \in \operatorname{int} I) = 1$, this identifies $f'(m(x))$ almost surely. The distribution of $m(x)$ is identified because m is known and the distribution of x is observed. Under $\mathbb{E}|f'(m(x))| < \infty$, integration with respect to that distribution identifies $\overline{k} = \mathbb{E}[f'(m(x))]$.

Finally, an absolutely continuous score distribution supported on I assigns probability one to $\operatorname{int} I$. If its density is positive almost everywhere on I, then I is contained in the score support, as required. □

Proof of Proposition 3. Let $s = m(x)$. Under the restricted feedback class, $y = \theta + a + \Delta \mathbb{1}\{s \ge c\} + \varepsilon$. Taking conditional expectations given s gives $\mathbb{E}[y \mid s] = \mathbb{E}[\theta \mid s] + a + \Delta \mathbb{1}\{s \ge c\} + \mathbb{E}[\varepsilon \mid s]$. The maintained condition $\mathbb{E}[\varepsilon \mid \theta, x] = 0$ implies $\mathbb{E}[\varepsilon \mid s] = 0$ because s is measurable in x. By definition, $r(s) = \mathbb{E}[\theta \mid s]$. Therefore $\mathbb{E}[y \mid s] = r(s) + a + \Delta \mathbb{1}\{s \ge c\}$.

Consider the one-sided limit from above. For $s > c$ sufficiently close to c, the indicator equals one, so $\lim_{s \downarrow c} \mathbb{E}[y \mid s] = \lim_{s \downarrow c} r(s) + a + \Delta = r(c) + a + \Delta$, where continuity of r at c is used in the last equality. From below, the indicator equals zero, so $\lim_{s \uparrow c} \mathbb{E}[y \mid s] = \lim_{s \uparrow c} r(s) + a = r(c) + a$. Subtracting gives Δ, proving equation (38).

Once Δ is known, equation (3) gives $\mathbb{E}[y \mid x] = u(x) + a + \Delta \mathbb{1}\{m(x) \ge c\}$. Therefore $\mathbb{E}[y \mid x] - \Delta \mathbb{1}\{m(x) \ge c\} = u(x) + a$. The left-hand side is identified from the deployment law, so $u(x)$ is identified up to the common additive constant a. □

Proof of Proposition 4. Define $\widetilde{h}(s) = s - \widetilde{f}(s)$. Since $\widetilde{f}'(s) \le \widetilde{\kappa}^+ < 1$, we have $\widetilde{h}'(s) = 1 - \widetilde{f}'(s) \ge 1 - \widetilde{\kappa}^+ > 0$ for every s. For any $s_2 > s_1$, the mean value theorem therefore gives $\widetilde{h}(s_2) - \widetilde{h}(s_1) \ge (1 - \widetilde{\kappa}^+)(s_2 - s_1) > 0$. Thus $\widetilde{h}$ is strictly increasing.

We next verify that $\widetilde{h}$ is onto. For $s \ge 0$, the preceding inequality gives $\widetilde{h}(s) \ge \widetilde{h}(0) + (1 - \widetilde{\kappa}^+)s$, so $\widetilde{h}(s) \to +\infty$ as $s \to +\infty$. For $s \le 0$, it gives $\widetilde{h}(0) - \widetilde{h}(s) \ge (1 - \widetilde{\kappa}^+)(-s)$, or $\widetilde{h}(s) \le \widetilde{h}(0) + (1 - \widetilde{\kappa}^+)s$.

Hence $\widetilde{h}(s) \to -\infty$ as $s \to -\infty$. Since $\widetilde{h}$ is continuous and strictly increasing, these limits imply that it is a bijection from $\mathbb{R}$ onto $\mathbb{R}$.

The same inequality implies that its inverse is globally Lipschitz. Indeed, for any $v_1, v_2 \in \mathbb{R}$, let $s_j = \widetilde{h}^{-1}(v_j)$ for $j = 1, 2$. Interchanging the two arguments when necessary gives $(1 - \widetilde{\kappa}^+)|s_2 - s_1| \leq |\widetilde{h}(s_2) - \widetilde{h}(s_1)| = |v_2 - v_1|$. Dividing by $1 - \widetilde{\kappa}^+ > 0$ yields $|\widetilde{h}^{-1}(v_2) - \widetilde{h}^{-1}(v_1)| \leq |v_2 - v_1|/(1 - \widetilde{\kappa}^+)$.

A counterfactual stable score must satisfy $\widetilde{m}^*(x) = u(x) + \widetilde{f}(\widetilde{m}^*(x))$. Rearranging gives $\widetilde{h}(\widetilde{m}^*(x)) = u(x)$. Applying the inverse gives $\widetilde{m}^*(x) = \widetilde{h}^{-1}(u(x))$, which is equation (39). Conversely, the score defined by this expression satisfies $\widetilde{h}(\widetilde{m}^*(x)) = u(x)$ and therefore satisfies the counterfactual stability equation. Uniqueness follows because $\widetilde{h}$ is one-to-one.

The score is admissible under the maintained second-moment assumptions. The inverse Lipschitz bound gives $|\widetilde{m}^*(x)| \leq |\widetilde{h}^{-1}(0)| + |u(x)|/(1 - \widetilde{\kappa}^+)$. Since $u(x)$ has a finite second moment, so does $\widetilde{m}^*(x)$. The stability equation then gives $\widetilde{f}(\widetilde{m}^*(x)) = \widetilde{m}^*(x) - u(x)$, which also has a finite second moment.

Part (i). Suppose $\widetilde{f}(s) \geq f(s)$ for every s. Then $\widetilde{h}(s) = s - \widetilde{f}(s) \leq s - f(s) = h(s)$ for every s. Fix x. Factual stability gives $h(m^*(x)) = u(x)$, so $\widetilde{h}(m^*(x)) \leq u(x)$. Counterfactual stability gives $\widetilde{h}(\widetilde{m}^*(x)) = u(x)$. Combining these relations yields $\widetilde{h}(m^*(x)) \leq \widetilde{h}(\widetilde{m}^*(x))$. Since $\widetilde{h}$ is strictly increasing, $m^*(x) \leq \widetilde{m}^*(x)$.

If $\widetilde{f}(s) \leq f(s)$ for every s, then $\widetilde{h}(s) \geq h(s)$. Evaluating at the factual stable score gives $\widetilde{h}(m^*(x)) \geq h(m^*(x)) = u(x) = \widetilde{h}(\widetilde{m}^*(x))$. Strict monotonicity of $\widetilde{h}$ now implies $\widetilde{m}^*(x) \leq m^*(x)$.

Part (ii). Fix x and subtract the two stability equations. The common posterior mean cancels, giving $\widetilde{m}^*(x) - m^*(x) = \widetilde{f}(\widetilde{m}^*(x)) - f(m^*(x))$. Adding and subtracting $\widetilde{f}(m^*(x))$ on the right gives $\widetilde{m}^*(x) - m^*(x) = [\widetilde{f}(\widetilde{m}^*(x)) - \widetilde{f}(m^*(x))] + [\widetilde{f}(m^*(x)) - f(m^*(x))]$.

Move the first bracket to the left and collect terms. This yields $[\widetilde{m}^*(x) - \widetilde{f}(\widetilde{m}^*(x))] - [m^*(x) - \widetilde{f}(m^*(x))] = \widetilde{f}(m^*(x)) - f(m^*(x))$. By the definition of $\widetilde{h}$, the identity is $\widetilde{h}(\widetilde{m}^*(x)) - \widetilde{h}(m^*(x)) = \widetilde{f}(m^*(x)) - f(m^*(x))$.

Taking absolute values gives $|\widetilde{h}(\widetilde{m}^*(x)) - \widetilde{h}(m^*(x))| = |\widetilde{f}(m^*(x)) - f(m^*(x))|$. The lower slope bound for $\widetilde{h}$ established above implies $(1 - \widetilde{\kappa}^+)|\widetilde{m}^*(x) - m^*(x)| \leq |\widetilde{h}(\widetilde{m}^*(x)) - \widetilde{h}(m^*(x))|$. Combining these expressions and dividing by $1 - \widetilde{\kappa}^+ > 0$ gives the pointwise bound $|\widetilde{m}^*(x) - m^*(x)| \leq |\widetilde{f}(m^*(x)) - f(m^*(x))|/(1 - \widetilde{\kappa}^+)$.

Finally, if $\sup_s |\widetilde{f}(s) - f(s)| \leq d$, then $|\widetilde{f}(m^*(x)) - f(m^*(x))| \leq d$ for every x. Substitution gives $|\widetilde{m}^*(x) - m^*(x)| \leq d/(1 - \widetilde{\kappa}^+)$. Taking the supremum over x proves equation (40). □

Proof of Corollary 6. Under the shifted normalization, the counterfactual stable-point equation is $\widetilde{m}_c^*(x) = u_c(x) + \widetilde{f}_c(\widetilde{m}_c^*(x))$. Substituting $u_c(x) = u(x) - c$ and $\widetilde{f}_c(s) = \widetilde{f}(s) + c$ gives

$$\widetilde{m}_c^*(x) = u(x) - c + \widetilde{f}(\widetilde{m}_c^*(x)) + c = u(x) + \widetilde{f}(\widetilde{m}_c^*(x)).$$

This is exactly the original counterfactual fixed-point equation. Proposition 4 gives a unique solution under the one-sided gain bound, so $\widetilde{m}_c^*(x) = \widetilde{m}^*(x)$ for every x. The same cancellation applies to any common additive normalization inherited from the factual decomposition. □

Proof of Proposition 5. The assumed integrability condition ensures that $m^* + \epsilon q$ is admissible for every ϵ in a neighborhood of zero. Thus the population retraining operator is well defined at each of these perturbations.

Fix a point x at which $q(x)$ is finite. Equation (5) gives $T(m^* + \epsilon q)(x) - Tm^*(x) = f(m^*(x) + \epsilon q(x)) - f(m^*(x))$. If $q(x) = 0$, this difference is zero for every ϵ, so (41) follows. If $q(x) \neq 0$, then for $\epsilon \neq 0$ the difference quotient equals $q(x)[f(m^*(x) + \epsilon q(x)) - f(m^*(x))]/[\epsilon q(x)]$. Differentiability of f at $m^*(x)$ implies that this expression converges to $f'(m^*(x))q(x)$ as $\epsilon \to 0$, proving (41).

Now suppose $f(t) = a + kt$. Choose a nonzero ϵ_0 in the neighborhood on which the integrability condition holds. Both $f(m^* + \epsilon_0 q)$ and $f(m^*)$ have finite second moments. Since $kq = [f(m^* + \epsilon_0 q) - f(m^*)]/\epsilon_0$, the function kq also has a finite second moment. The identity $f(m^* + q) = f(m^*) + kq$ therefore shows that $m^* + q$ is admissible. Applying (5) gives $T(m^* + q) - Tm^* = f(m^* + q) - f(m^*) = kq$, proving (42). This

argument includes $k = 0$. Moreover, if q itself has a finite second moment, then $f(m^* + \epsilon q) = f(m^*) + \epsilon k q$ has a finite second moment for every finite ϵ. Thus every square-integrable perturbation satisfies the required condition under affine feedback.

Finally, suppose $\kappa_{\text{abs}} < 1$ and q is bounded. The mean-value theorem gives $|f(m^*(x) + \epsilon q(x)) - f(m^*(x))| \le \kappa_{\text{abs}}|\epsilon|\|q\|_\infty$. Consequently, $|f(m^*(x) + \epsilon q(x))|^2 \le 2|f(m^*(x))|^2 + 2\kappa_{\text{abs}}^2\epsilon^2\|q\|_\infty^2$. Taking expectations shows that every finite perturbation is admissible because $f(m^*)$ has a finite second moment.

The pointwise derivative already established satisfies $|f'(m^*(x))q(x)| \le \kappa_{\text{abs}}|q(x)| \le \kappa_{\text{abs}}\|q\|_\infty$. Taking the supremum over x proves the claimed uniform norm bound. □

Proof of Theorem 5. Fix x and suppress it from the notation. Since M_t is $\mathcal{F}_t$-measurable, the tower property and the martingale-difference condition give

$$\mathbb{E}[\xi_{t+1} \mid M_t] = \mathbb{E}[\mathbb{E}[\xi_{t+1} \mid \mathcal{F}_t] \mid M_t] = 0.$$

Taking conditional expectations in equation (43) therefore yields

$$\mathbb{E}[M_{t+1} \mid M_t] = u + f(M_t).$$

By definition, G_x is a version of the left-hand conditional expectation. Hence $G_x(M_t) = u + f(M_t)$ almost surely. Equivalently, $G_x(s) = u + f(s)$ for P_{M_t}-almost every s in the support. This proves part (i).

For part (ii), evaluate the identity at two score values s and s_0 in the full-probability set on which the identity in part (i) holds. Subtracting gives

$$G_x(s) - G_x(s_0) = [u + f(s)] - [u + f(s_0)] = f(s) - f(s_0),$$

which is equation (45). If both sides are continuous on the score support, an equality holding almost everywhere with respect to a measure whose support is $\mathcal{S}_x$ extends to every point of the support. To see this, suppose the continuous difference between the two sides were nonzero at some support point. It would remain bounded away from zero on a neighborhood of that point, and every neighborhood of a support point has positive probability, contradicting almost-sure equality.

For part (iii), under the continuity condition, $G_x(s) = u(x) + f(s)$ throughout the score support. At an interior point where f is differentiable, this identity implies G_x is differentiable, with $G_x'(s) = f'(s)$.

For part (iv), apply part (ii) at feature value x and again at x'. The common feedback function gives

$$G_x(s) - G_x(s_0) = f(s) - f(s_0) = G_{x'}(s) - G_{x'}(s_0),$$

which is equation (46). The posterior means cancel separately within each difference, so the equality does not require $u(x) = u(x')$. This proves the theorem. □

Proof of Proposition 6. Fix x and t, and write $S = M_t(x)$. The integrability conditions ensure that the conditional expectations below exist. Since $u(x)$ is constant and $f(S)$ is measurable with respect to S, the archive recursion gives $\mathbb{E}[M_{t+1}(x) \mid S] = u(x) + f(S) + \mathbb{E}[\xi_{t+1}(x) \mid S] = u(x) + f(S) + b_x(S)$. By the definition of G_x, it follows that $G_x(S) = u(x) + f(S) + b_x(S)$ almost surely.

Thus there is a set of score values with P_S-probability one on which $G_x(s) = u(x) + f(s) + b_x(s)$. Evaluating this equality at any two values s and s_0 in that set and subtracting cancels $u(x)$. Hence $G_x(s) - G_x(s_0) = f(s) - f(s_0) + b_x(s) - b_x(s_0)$ for $P_S \otimes P_S$-almost every pair.

Rearranging gives $f(s) - f(s_0) = G_x(s) - G_x(s_0) - [b_x(s) - b_x(s_0)]$. If $|b_x(s) - b_x(s_0)| \le B_x(s, s_0)$, then $-B_x(s, s_0) \le b_x(s) - b_x(s_0) \le B_x(s, s_0)$. Substituting the two endpoints of this interval into the preceding equality gives (48). If b_x is constant P_S-almost surely, its difference is zero for almost every pair, and (47) reduces to $G_x(s) - G_x(s_0) = f(s) - f(s_0)$.

Finally, suppose the stated continuous versions exist on S_x. The function $d(s) = G_x(s) - u(x) - f(s) - b_x(s)$ is continuous on S_x and satisfies $d(S) = 0$ almost surely. If $d(s^*) \neq 0$ at some $s^* \in S_x$, continuity would give a neighborhood of s^* in S_x on which $|d|$ is bounded away from zero. Every neighborhood of a support point has positive P_S-probability, contradicting $d(S) = 0$ almost surely. Thus $d(s) = 0$ for every $s \in S_x$. Subtracting the identity at any two support points proves the pointwise equality. The same rearrangement proves the bounds wherever the stated bias bound holds. □

Proof of Corollary 7. Fix an audit feature value x. Part (i) of Theorem 5 gives $G_x(s) = u(x) + f(s)$ for $P_{M_t(x)}$-almost every $s \in \mathcal{S}_x$. The difference $G_x(s) - u(x) - f(s)$ is continuous on $\mathcal{S}_x$. If it were nonzero at a point of this support, continuity would make it nonzero throughout a neighborhood of that point. Every such neighborhood has positive $P_{M_t(x)}$ probability, contradicting the almost-sure equality. Hence

$$G_x(s) = u(x) + f(s) \qquad \text{for every } s \in \mathcal{S}_x.$$

Apply this identity at x_0 and subtract its values at s and s_0. For every $s \in \mathcal{S}_{x_0}$, we obtain $f(s) - f(s_0) = G_{x_0}(s) - G_{x_0}(s_0)$. Fixing $f(s_0)$ therefore identifies f throughout $\mathcal{S}_{x_0}$, including I. Now suppose $\mathcal{S}_x$ intersects a score set on which f has already been identified, and choose a score s in the intersection. The pointwise identity gives $u(x) = G_x(s) - f(s)$, so $u(x)$ is identified. For any $t \in \mathcal{S}_x$, the same identity then gives $f(t) = G_x(t) - u(x)$. Thus the normalization extends to all of $\mathcal{S}_x$. Applying this step successively proves recovery along every finite chain of audit feature values whose successive supports intersect. □

Proof of Corollary 8. Fix x and abbreviate $m^* = m^*(x)$, $u = u(x)$, $a = f'(m^*)$, and $\kappa = \overline{\kappa}_{\text{abs}} < 1$. Stability gives $m^* = u + f(m^*)$.

For $\tau > 0$, define the rescaled deviation $Z_t^\tau = (M_t^\tau - m^*)/\tau$. Subtract the stable-point identity from equation (49):

$$M_{t+1}^\tau - m^* = f(M_t^\tau) - f(m^*) + \tau\xi_{t+1}.$$

Since $M_t^\tau = m^* + \tau Z_t^\tau$, division by τ gives

$$Z_{t+1}^\tau = g_\tau(Z_t^\tau) + \xi_{t+1}, \qquad g_\tau(z) = \frac{f(m^* + \tau z) - f(m^*)}{\tau}. \tag{A.2}$$

The function g_τ satisfies $g_\tau(0) = 0$. For any z, z', the mean-value theorem gives a point c between $m^* + \tau z$ and $m^* + \tau z'$ such that

$$g_\tau(z) - g_\tau(z') = \frac{f'(c)\tau(z - z')}{\tau} = f'(c)(z - z').$$

Therefore $|g_\tau(z) - g_\tau(z')| \leq \kappa|z - z'|$ for every z, z'. Thus the rescaled recursion is a contraction with the same modulus κ for every τ.

Part (i). We construct the stationary solution directly. For an integer $J \geq 1$, start the recursion at date $t - J$ from zero and use the innovations $\xi_{t-J+1}, \ldots, \xi_t$. Denote the resulting value at date t by $Z_{t,J}^\tau$. Since $g_\tau(0) = 0$ and $|g_\tau(z)| \leq \kappa|z|$, repeated use of the triangle inequality gives

$$|Z_{t,J}^\tau| \leq |\xi_t| + \kappa|\xi_{t-1}| + \cdots + \kappa^{J-1}|\xi_{t-J+1}|.$$

Taking L^2 norms and using Minkowski's inequality yields

$$\|Z_{t,J}^\tau\|_2 \leq \sum_{j=0}^{J-1} \kappa^j \|\xi_{t-j}\|_2 = \sigma_\xi \sum_{j=0}^{J-1} \kappa^j \leq \frac{\sigma_\xi}{1 - \kappa}.$$

Hence the second moments are bounded uniformly in both J and τ.

We next show that the backward sequence is Cauchy. Compare constructions of lengths $J + 1$ and J. The longer construction first maps the additional oldest innovation ξ_{t-J} into the state at date $t - J + 1$,

while the shorter construction starts from zero at that date. The difference at date $t-J+1$ is therefore $|\xi_{t-J}|$. Each of the remaining J updates contracts differences by at most κ. Hence

$$|Z^\tau_{t,J+1} - Z^\tau_{t,J}| \le \kappa^J |\xi_{t-J}|.$$

For $K > J$, telescoping gives

$$|Z^\tau_{t,K} - Z^\tau_{t,J}| \le \sum_{r=J}^{K-1} \kappa^r |\xi_{t-r}|.$$

Taking L^2 norms and applying Minkowski again gives

$$\|Z^\tau_{t,K} - Z^\tau_{t,J}\|_2 \le \sigma_\xi \sum_{r=J}^{K-1} \kappa^r \le \frac{\sigma_\xi \kappa^J}{1-\kappa}.$$

The right-hand side tends to zero as $J \to \infty$, uniformly in $K > J$ and τ. Thus $Z^\tau_{t,J}$ converges in L^2 to a limit Z^τ_t. Passing to the limit in the recursion gives equation (A.2) for the limit. The construction uses the infinite past innovation sequence in a shift-invariant way, so $\{Z^\tau_t\}$ is causal and stationary.

Uniqueness follows from contraction. If Z^τ_t and $\widetilde{Z}^\tau_t$ are two stationary solutions driven by the same innovations, then

$$|Z^\tau_{t+1} - \widetilde{Z}^\tau_{t+1}| = |g_\tau(Z^\tau_t) - g_\tau(\widetilde{Z}^\tau_t)| \le \kappa |Z^\tau_t - \widetilde{Z}^\tau_t|.$$

Iterating backward J periods gives $|Z^\tau_t - \widetilde{Z}^\tau_t| \le \kappa^J |Z^\tau_{t-J} - \widetilde{Z}^\tau_{t-J}|$. Taking L^2 norms and using stationarity gives

$$\|Z^\tau_t - \widetilde{Z}^\tau_t\|_2 \le \kappa^J \|Z^\tau_t - \widetilde{Z}^\tau_t\|_2.$$

Letting $J \to \infty$ forces the norm to be zero. Therefore, the stationary solution is unique. Since $M^\tau_t = m^* + \tau Z^\tau_t$, it also has finite second moment.

Part (ii). Differentiability of f at m^* implies that for every fixed z,

$$\lim_{\tau \downarrow 0} g_\tau(z) = \lim_{\tau \downarrow 0} \frac{f(m^* + \tau z) - f(m^*)}{\tau} = az.$$

Let $Z_{t,J}$ denote the J-step backward construction for the linear recursion $Z_{t+1} = aZ_t + \xi_{t+1}$, again starting from zero at date $t-J$. We first prove convergence for fixed J. For $J = 1$, both constructions equal ξ_t, so convergence is immediate. Suppose $Z^\tau_{t-1,J-1} \to Z_{t-1,J-1}$ almost surely. Then

$$Z^\tau_{t,J} = g_\tau(Z^\tau_{t-1,J-1}) + \xi_t.$$

Add and subtract $g_\tau(Z_{t-1,J-1})$:

$$|g_\tau(Z^\tau_{t-1,J-1}) - aZ_{t-1,J-1}| \le |g_\tau(Z^\tau_{t-1,J-1}) - g_\tau(Z_{t-1,J-1})| + |g_\tau(Z_{t-1,J-1}) - aZ_{t-1,J-1}|.$$

The first term is at most $\kappa |Z^\tau_{t-1,J-1} - Z_{t-1,J-1}|$ and therefore tends to zero almost surely by the induction hypothesis. The second tends to zero almost surely because $Z_{t-1,J-1}$ is finite almost surely and $g_\tau(z) \to az$ pointwise. Thus $Z^\tau_{t,J} \to Z_{t,J}$ almost surely for every fixed J.

The same geometric envelope used above gives

$$|Z^\tau_{t,J}| + |Z_{t,J}| \le 2 \sum_{j=0}^{J-1} \kappa^j |\xi_{t-j}|,$$

because $|a| = |f'(m^*)| \le \kappa$. The right-hand side has finite second moment. Dominated convergence therefore

strengthens the fixed-J convergence to L^2 convergence:

$$\lim_{\tau \downarrow 0} \|Z^\tau_{t,J} - Z_{t,J}\|_2 = 0.$$

Now compare the finite backward constructions with their stationary limits. The Cauchy estimate above gives

$$\|Z^\tau_t - Z^\tau_{t,J}\|_2 \le \frac{\sigma_\xi \kappa^J}{1-\kappa}$$

uniformly in τ. The same bound applies to the stationary linear solution because $|a| \le \kappa$:

$$\|Z_t - Z_{t,J}\|_2 \le \frac{\sigma_\xi \kappa^J}{1-\kappa}.$$

Hence

$$\|Z^\tau_t - Z_t\|_2 \le \frac{2\sigma_\xi \kappa^J}{1-\kappa} + \|Z^\tau_{t,J} - Z_{t,J}\|_2.$$

Given $\varepsilon > 0$, first choose J so large that the first term is below $\varepsilon/2$. For that fixed J, choose τ small enough that the second term is below $\varepsilon/2$. This proves $Z^\tau_t \to Z_t$ in L^2.

The limiting linear recursion has the causal representation $Z_t = \sum_{j=0}^\infty a^j \xi_{t-j}$. Since $|a| < 1$ and ξ has finite variance, the series converges in L^2 and is the unique stationary solution of equation (50).

Part (iii). Correlation is unchanged by adding a constant and multiplying both variables by the same positive scalar. Therefore

$$\mathrm{Corr}(M^\tau_{t+1}, M^\tau_t) = \mathrm{Corr}(Z^\tau_{t+1}, Z^\tau_t).$$

Part (ii) gives L^2 convergence of both coordinates of the pair (Z^τ_{t+1}, Z^τ_t) to (Z_{t+1}, Z_t). Consequently, their means, second moments, variances, and covariance converge. The limiting AR(1) process has mean zero because $\mathbb{E}\xi_t = 0$. Its variance satisfies

$$\mathrm{Var}(Z_t) = a^2 \mathrm{Var}(Z_t) + \sigma^2_\xi,$$

so $\mathrm{Var}(Z_t) = \sigma^2_\xi/(1-a^2) > 0$. Also

$$\mathrm{Cov}(Z_{t+1}, Z_t) = \mathrm{Cov}(aZ_t + \xi_{t+1}, Z_t) = a\mathrm{Var}(Z_t) + \mathrm{Cov}(\xi_{t+1}, Z_t).$$

The innovation ξ_{t+1} is independent of the past and therefore independent of Z_t, so the final covariance is zero. Thus $\mathrm{Cov}(Z_{t+1}, Z_t) = a\mathrm{Var}(Z_t)$. Stationarity gives equal variances at dates t and $t+1$, so

$$\mathrm{Corr}(Z_{t+1}, Z_t) = \frac{a\mathrm{Var}(Z_t)}{\mathrm{Var}(Z_t)} = a = f'(m^*).$$

Taking the limit in the correlations proves equation (51). □

Proof of Proposition 7. Define $d_t = \sup_{x \in \mathbb{X}} |\widehat{m}_t(x) - m^*(x)|$. The initial-distance assumption gives $d_0 < \infty$. By the mean-value theorem and (52), $|f(a) - f(b)| \le \kappa_{\mathrm{abs}}|a-b|$ for all $a, b \in \mathbb{R}$.

Since $Tm^* = m^*$, the triangle inequality gives $|\widehat{m}_{t+1}(x) - m^*(x)| \le |\widehat{m}_{t+1}(x) - (T\widehat{m}_t)(x)| + |(T\widehat{m}_t)(x) - (Tm^*)(x)|$. The first term is at most e_t. Equation (5) and the Lipschitz bound imply that the second is at most $\kappa_{\mathrm{abs}}|\widehat{m}_t(x) - m^*(x)|$. Taking the supremum over x yields

$$d_{t+1} \le \kappa_{\mathrm{abs}} d_t + e_t. \tag{A.3}$$

In particular, finite initial distance and finite estimation-error bounds imply that d_t is finite at every date.

We prove the finite-horizon bound by induction. For $t = 1$, (A.3) gives $d_1 \le \kappa_{\mathrm{abs}} d_0 + e_0$. If $d_t \le$

$\kappa_{\text{abs}}^t d_0 + \sum_{j=0}^{t-1} \kappa_{\text{abs}}^{t-1-j} e_j$, substitution into (A.3) gives $d_{t+1} \le \kappa_{\text{abs}}^{t+1} d_0 + \sum_{j=0}^{t-1} \kappa_{\text{abs}}^{t-j} e_j + e_t = \kappa_{\text{abs}}^{t+1} d_0 + \sum_{j=0}^{t} \kappa_{\text{abs}}^{t-j} e_j$. Thus (53) holds for every $t \ge 1$.

If $e_j \le e$ for every j, summing the geometric series gives $d_t \le \kappa_{\text{abs}}^t d_0 + e(1 - \kappa_{\text{abs}}^t)/(1 - \kappa_{\text{abs}})$. Because $0 \le \kappa_{\text{abs}} < 1$ and $d_0 < \infty$, taking the limit superior proves (54).

The initial-distance condition matters for a uniform conclusion. For example, let $\mathbb{X} = \mathbb{R}$, $u \equiv 0$, and $f(s) = \kappa s$ with $0 < \kappa < 1$. Then $Tm = \kappa m$ and $m^* \equiv 0$. Exact retraining from $\widehat{m}_0(x) = x$ gives $\widehat{m}_t(x) = \kappa^t x$. This converges to zero for each fixed x, but $\sup_{x \in \mathbb{R}} |\kappa^t x| = \infty$ at every finite date. Choosing a feature distribution with full support and finite second moment makes these deployments admissible. Hence pointwise convergence alone does not imply convergence in the uniform norm. □

Proof of Theorem 6. Fix a period t and suppress the time subscript on the cross-sectional variables. The population regression decomposition in equations (59)–(60) is $y_i = \alpha_t + (\rho + k\lambda_t)\widetilde{x}_i + u_i$, where α_t is the cross-sectional mean component and $u_i = (1 - \rho)\widetilde{\theta}_i - \rho\eta_i + \varepsilon_i$.

We first verify the orthogonality and independence properties of u_i. Since $\widetilde{x}_i = \widetilde{\theta}_i + \eta_i$ and the primitive shocks are mutually independent, $\operatorname{Cov}(\widetilde{x}_i, u_i) = \operatorname{Cov}(\widetilde{\theta}_i + \eta_i, (1 - \rho)\widetilde{\theta}_i - \rho\eta_i + \varepsilon_i) = (1 - \rho)\sigma_\theta^2 - \rho\sigma_\eta^2$. Using $\rho = \sigma_\theta^2/V$ and $1 - \rho = \sigma_\eta^2/V$, the first term is $(\sigma_\eta^2/V)\sigma_\theta^2$ and the second term is $(\sigma_\theta^2/V)\sigma_\eta^2$, so the difference is zero. The pair $(\widetilde{x}_i, u_i)$ is jointly Gaussian because both are linear combinations of jointly Gaussian primitive shocks. Zero covariance therefore implies independence. The variance of u_i is σ_u^2 by definition.

The OLS slope with an intercept is $\lambda_{t+1} = \sum_i (\widetilde{x}_i - \bar{x})(y_i - \bar{y})/S$, where $S = \sum_i (\widetilde{x}_i - \bar{x})^2$. Substituting $y_i = \alpha_t + (\rho + k\lambda_t)\widetilde{x}_i + u_i$ and subtracting the sample mean gives $y_i - \bar{y} = (\rho + k\lambda_t)(\widetilde{x}_i - \bar{x}) + (u_i - \bar{u})$. Hence, the OLS numerator is

$$\sum_i (\widetilde{x}_i - \bar{x})(y_i - \bar{y}) = (\rho + k\lambda_t) \sum_i (\widetilde{x}_i - \bar{x})^2 + \sum_i (\widetilde{x}_i - \bar{x})(u_i - \bar{u}).$$

Since $\sum_i (\widetilde{x}_i - \bar{x}) = 0$, the term involving $\bar{u}$ vanishes. Thus $\sum_i (\widetilde{x}_i - \bar{x})(u_i - \bar{u}) = \sum_i (\widetilde{x}_i - \bar{x})u_i$. Dividing by S gives

$$\lambda_{t+1} = \rho + k\lambda_t + \frac{\sum_i (\widetilde{x}_i - \bar{x})u_i}{S}.$$

Define the last ratio as U_t. This proves equation (61).

We now derive its exact distribution. Conditional on the full design vector $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$, the weights $a_i = \widetilde{x}_i - \bar{x}$ are fixed. The residuals u_i are independent $N(0, \sigma_u^2)$ variables and are independent of the design. Therefore, the conditional numerator $N = \sum_i a_i u_i$ is Gaussian with mean zero and variance $\sigma_u^2 \sum_i a_i^2 = \sigma_u^2 S$. Hence there exists a standard normal $Z \sim N(0, 1)$, independent of S, such that $N = \sigma_u \sqrt{S}\, Z$. Consequently $U_t = N/S = \sigma_u Z/\sqrt{S}$.

Because the design observations are i.i.d. $N(0, V)$, the centered sum of squares satisfies $S/V \sim \chi_{n-1}^2$. Let $Q = S/V$, so $Q \sim \chi_{n-1}^2$ and Q is independent of Z. Then

$$U_t = \frac{\sigma_u}{\sqrt{VQ}} Z = \frac{\sigma_u}{\sqrt{V(n-1)}} \frac{Z}{\sqrt{Q/(n-1)}}.$$

By definition, $Z/\sqrt{Q/(n-1)} \sim T_{n-1}$. This is equation (62).

The law of U_t contains no λ_t because both the design and the regime-invariant residual law are independent of the deployed coefficient. Each period uses a fresh independent cross-section, so the pairs (S_t, Z_t) are i.i.d. across t. Therefore the innovations U_t are i.i.d. across periods and independent of the sigma-algebra generated by $\lambda_0, U_0, \ldots, U_{t-1}$.

For $n \ge 3$, the Student variable has degrees of freedom $n - 1 > 1$, so its mean exists and equals zero.

Hence $\mathbb{E}[U_t] = 0$. For $n \geq 4$, $n-1 > 2$ and $\mathrm{Var}(T_{n-1}) = (n-1)/(n-3)$. Therefore

$$\mathrm{Var}(U_t) = \frac{\sigma_u^2}{V(n-1)} \frac{n-1}{n-3} = \frac{\sigma_u^2}{V(n-3)}.$$

This completes the proof. □

Proof of Theorem 7. Write the recursion as

$$\lambda_{t+1} = \rho + k\lambda_t + U_t. \tag{A.4}$$

By Theorem 6, the innovations are i.i.d., symmetric, nondegenerate, and have an everywhere-positive continuous density. Their scaled Student distribution satisfies $\mathbb{E}|U_t|^p < \infty$ for every $0 < p < n-1$. The initial coefficient is finite almost surely and independent of the innovation sequence.

Part (i). Suppose $|k| < 1$ and define $\lambda^* = \rho/(1-k)$. Subtracting $\lambda^* = \rho + k\lambda^*$ from (A.4) gives

$$X_{t+1} = kX_t + U_t, \qquad X_t = \lambda_t - \lambda^*. \tag{A.5}$$

Repeated substitution yields $X_t = k^t X_0 + \sum_{j=0}^{t-1} k^j U_{t-1-j}$. The initial-condition term $k^t X_0$ tends to zero almost surely.

To construct a stationary solution, extend the innovations to an i.i.d. sequence indexed by $\mathbb{Z}$ and consider $X_t^{\mathrm{st}} = \sum_{j=0}^{\infty} k^j U_{t-1-j}$. Choose $p \in (0, \min\{1, n-1\})$. Tonelli's theorem gives $\mathbb{E}[\sum_{j=0}^{\infty} |k|^{jp} |U_{t-1-j}|^p] = \mathbb{E}|U_0|^p/(1-|k|^p) < \infty$. Therefore, the sum of these nonnegative terms is finite almost surely. In particular, $|k|^j |U_{t-1-j}|$ is eventually at most one. Since $a \leq a^p$ for $a \in [0,1]$, the series defining X_t^{st} converges absolutely almost surely.

The construction depends only on innovations dated before t and is invariant under a common shift of all dates. It therefore defines a causal stationary process. Absolute convergence permits reindexing, which gives $kX_t^{\mathrm{st}} + U_t = U_t + \sum_{j=1}^{\infty} k^j U_{t-j} = \sum_{j=0}^{\infty} k^j U_{t-j} = X_{t+1}^{\mathrm{st}}$. Thus it satisfies (A.5). At each date, its marginal distribution is that of $\sum_{j=0}^{\infty} k^j U_{-j}$. Adding λ^* proves (64).

We establish geometric ergodicity and uniqueness without imposing a moment condition on the initial coefficient. For the same p, define $V_p(x) = 1 + |x|^p$. The inequality $|a+b|^p \leq |a|^p + |b|^p$ gives $\mathbb{E}[V_p(X_{t+1}) \mid X_t = x] \leq |k|^p V_p(x) + B$, where $B = 1 + \mathbb{E}|U_0|^p - |k|^p < \infty$. Choose $r \in (|k|^p, 1)$ and a compact interval C large enough that $B \leq (r - |k|^p) V_p(x)$ outside C. There is then a finite b such that $\mathbb{E}[V_p(X_{t+1}) \mid X_t = x] \leq r V_p(x) + b\mathbf{1}_C(x)$ for every x.

If g denotes the innovation density, the transition density is $g(y - kx)$. Its positivity makes the chain irreducible and aperiodic. Continuity and positivity also imply that its infimum is strictly positive on any compact rectangle, so every compact interval is a small set. The drift-and-minorization theorem therefore gives geometric ergodicity and a unique invariant probability distribution. The stationary distribution constructed above is unique. Moreover, any two trajectories driven by the same innovations satisfy $|X_t - \widetilde{X}_t| = |k|^t |X_0 - \widetilde{X}_0|$, so their difference tends to zero almost surely.

Each innovation is symmetric about zero, and multiplication by k^j preserves symmetry even when $k < 0$. Independence implies that every finite partial sum is symmetric, and this property passes to the almost-sure limit. Hence the stationary coefficient is symmetric about λ^*.

For $n \geq 3$, $\mathbb{E}|U_0| < \infty$ and $\mathbb{E}U_0 = 0$. Since $\sum_{j=0}^{\infty} |k|^j \mathbb{E}|U_0| < \infty$, Fubini's theorem gives $\mathbb{E}X_t^{\mathrm{st}} = \sum_{j=0}^{\infty} k^j \mathbb{E}U_0 = 0$. Thus $\mathbb{E}\lambda_\infty = \lambda^*$. For $n \geq 4$, the stationary series also converges in L^2, and independence gives $\mathrm{Var}(\lambda_\infty) = \sum_{j=0}^{\infty} k^{2j} \mathrm{Var}(U_0) = \mathrm{Var}(U_0)/(1-k^2)$. Substituting $\mathrm{Var}(U_0) = \sigma_u^2/[V(n-3)]$ proves the stated variance formula.

Part (ii). Suppose $k = 1$. Summing $\lambda_{t+1} - \lambda_t = \rho + U_t$ from date zero through date $t-1$ gives $\lambda_t = \lambda_0 + \rho t + \sum_{j=0}^{t-1} U_j$, proving (65). Dividing by t yields $\lambda_t/t = \lambda_0/t + \rho + t^{-1} \sum_{j=0}^{t-1} U_j$.

For $n \geq 3$, the innovations are integrable with mean zero, so the strong law gives $t^{-1} \sum_{j=0}^{t-1} U_j \to 0$

almost surely. Since λ_0 is finite almost surely, $\lambda_0/t \to 0$ as well. Hence $\lambda_t/t \to \rho$ almost surely. Under the maintained assumption $\rho > 0$, the coefficient diverges to $+\infty$ at linear rate ρt. If $\rho = 0$, the recursion is a centered random walk.

Part (iii). Suppose $k = -1$. The fixed point is $\lambda^- = \rho/2$. Setting $X_t = \lambda_t - \lambda^-$ gives $X_{t+1} = -X_t + U_t$. Define $W_t = (-1)^t X_t$. Then $W_{t+1} = (-1)^{t+1}(-X_t + U_t) = W_t + (-1)^{t+1}U_t$, proving (66).

Symmetry implies that $(-1)^{t+1}U_t$ has the same distribution as U_t, and deterministic sign changes preserve independence. For $n \geq 3$, the transformed increments are therefore i.i.d., integrable, nondegenerate, and centered. The oscillation theorem for centered random walks gives $\limsup_{t\to\infty} W_t = +\infty$ and $\liminf_{t\to\infty} W_t = -\infty$ almost surely.

The even subsequence satisfies $W_{2j+2} - W_{2j} = -U_{2j} + U_{2j+1}$. These increments are also i.i.d., integrable, nondegenerate, and centered, so the even subsequence oscillates without bound in both directions. Since $\lambda_{2j} - \lambda^- = W_{2j}$, the original coefficient has arbitrarily large positive and negative excursions around $\rho/2$.

Part (iv). Suppose $|k| > 1$ and define $\widetilde{\lambda} = \rho/(1-k)$. Subtracting $\widetilde{\lambda} = \rho + k\widetilde{\lambda}$ from (A.4) gives $\lambda_{t+1} - \widetilde{\lambda} = k(\lambda_t - \widetilde{\lambda}) + U_t$. Iterating yields $\lambda_t - \widetilde{\lambda} = k^t(\lambda_0 - \widetilde{\lambda}) + \sum_{j=0}^{t-1} k^{t-1-j}U_j$. Factoring out k^t proves (67), with $M_t = \sum_{j=0}^{t-1} k^{-j-1}U_j$.

Choose $p \in (0, \min\{1, n-1\})$. Tonelli's theorem gives $\mathbb{E}[\sum_{j=0}^{\infty} |k|^{-(j+1)p}|U_j|^p] = \mathbb{E}|U_0|^p|k|^{-p}/(1-|k|^{-p}) < \infty$. The sum of the pth powers is therefore finite almost surely. As in part (i), this implies absolute convergence of $\sum_{j=0}^{\infty} k^{-j-1}U_j$. Denote its finite limit by M_∞.

Let $C = \lambda_0 - \widetilde{\lambda} + M_\infty$. To show that C is nonzero almost surely, write $C = \lambda_0 - \widetilde{\lambda} + k^{-1}U_0 + \sum_{j=1}^{\infty} k^{-j-1}U_j$. Conditional on $\lambda_0, U_1, U_2, \ldots$, every term except $k^{-1}U_0$ is fixed. Nonanticipativity and independence of the innovations leave U_0 independent of these conditioning variables. Since U_0 has a continuous density and $k^{-1} \neq 0$, the conditional probability that $C = 0$ is zero. Taking expectations gives $\Pr(C = 0) = 0$.

Equation (67) now implies $(\lambda_t - \widetilde{\lambda})/k^t \to C$ almost surely. Consequently, $|\lambda_t - \widetilde{\lambda}|/|k|^t \to |C| > 0$. Since $\widetilde{\lambda}/|k|^t \to 0$, we also have $|\lambda_t|/|k|^t \to |C| > 0$. Thus $|\lambda_t| \to \infty$ at geometric rate $|k|^t$.

If $k < -1$, then $k^t = (-1)^t|k|^t$. The factor $\lambda_0 - \widetilde{\lambda} + M_t$ converges to the nonzero limit C, so its sign is eventually constant. The sign of $\lambda_t - \widetilde{\lambda}$ therefore eventually alternates every period. □

Proof of Corollary 9. In the stationary subcritical regime, equation (61) is $\lambda_{t+1} = \rho + k\lambda_t + U_t$, where U_t is independent of λ_t and has mean zero. Taking covariance with λ_t gives $\mathrm{Cov}(\lambda_{t+1}, \lambda_t) = \mathrm{Cov}(\rho, \lambda_t) + k\mathrm{Cov}(\lambda_t, \lambda_t) + \mathrm{Cov}(U_t, \lambda_t)$. The first covariance is zero because ρ is constant. The last is zero by independence. Therefore $\mathrm{Cov}(\lambda_{t+1}, \lambda_t) = k\mathrm{Var}(\lambda_t)$. Stationarity gives $\mathrm{Var}(\lambda_{t+1}) = \mathrm{Var}(\lambda_t)$, so the lag-one correlation is $\mathrm{Cov}(\lambda_{t+1}, \lambda_t)/\mathrm{Var}(\lambda_t) = k$.

Taking expectations in the stationary recursion gives $\mathbb{E}[\lambda_t] = \rho + k\mathbb{E}[\lambda_t] + \mathbb{E}[U_t] = \rho + k\mathbb{E}[\lambda_t]$. Rearranging yields $(1-k)\mathbb{E}[\lambda_t] = \rho$, so $\mathbb{E}[\lambda_t] = \rho/(1-k)$. Therefore the archive identifies k from the lag-one correlation and then identifies $\rho = (1-k)\mathbb{E}[\lambda_t]$.

For the variance restriction, subtract the stationary mean to get $\lambda_{t+1} - \mathbb{E}\lambda = k(\lambda_t - \mathbb{E}\lambda) + U_t$. Taking variances and using independence gives $\mathrm{Var}(\lambda) = k^2\mathrm{Var}(\lambda) + \mathrm{Var}(U)$. Therefore $(1-k^2)\mathrm{Var}(\lambda) = \mathrm{Var}(U)$. By Theorem 6, $\mathrm{Var}(U) = \sigma_u^2/[V(n-3)]$. By equation (60), $\sigma_u^2 = \rho(1-\rho)V + \sigma_\varepsilon^2$. Substituting these expressions gives equation (68). □

Proof of Proposition 8. Part (i). In the stationary subcritical regime, Theorem 6 gives $\lambda_{t+1} = \rho + k\lambda_t + U_t$, where the innovations are i.i.d., have mean zero and finite positive variance for $n \geq 4$, and are independent of the past and the currently deployed coefficient. The stationary process is geometrically ergodic by Theorem 7, hence ergodic. Write $\mu_\lambda = \mathbb{E}[\lambda_t] = \rho/(1-k)$.

Let $\bar{\lambda}_T = T^{-1}\sum_{t=1}^{T} \lambda_t$ and $\bar{\lambda}_{+,T} = T^{-1}\sum_{t=1}^{T} \lambda_{t+1}$. The OLS slope is $\widehat{k}_T = \sum_{t=1}^{T}(\lambda_t - \bar{\lambda}_T)(\lambda_{t+1} - \bar{\lambda}_{+,T})/\sum_{t=1}^{T}(\lambda_t - \bar{\lambda}_T)^2$. Substituting $\lambda_{t+1} = \rho + k\lambda_t + U_t$ and using $\sum_{t=1}^{T}(\lambda_t - \bar{\lambda}_T) = 0$ removes the constant and the sample mean of the dependent variable. Thus $\widehat{k}_T - k = \sum_{t=1}^{T}(\lambda_t - \bar{\lambda}_T)U_t/\sum_{t=1}^{T}(\lambda_t - \bar{\lambda}_T)^2$.

By the ergodic theorem, $\bar{\lambda}_T \to \mu_\lambda$ and $T^{-1}\sum_{t=1}^T \lambda_t^2 \to \mathbb{E}[\lambda_t^2]$ almost surely. Therefore, $T^{-1}\sum_{t=1}^T(\lambda_t - \bar{\lambda}_T)^2 = T^{-1}\sum_{t=1}^T \lambda_t^2 - \bar{\lambda}_T^2 \to \mathrm{Var}(\lambda_\infty) > 0$ almost surely.

For the numerator, write $T^{-1}\sum_{t=1}^T(\lambda_t - \bar{\lambda}_T)U_t = T^{-1}\sum_{t=1}^T(\lambda_t - \mu_\lambda)U_t - (\bar{\lambda}_T - \mu_\lambda)T^{-1}\sum_{t=1}^T U_t$. The joint process (λ_t, U_t) is stationary and ergodic, and the product $(\lambda_t - \mu_\lambda)U_t$ is integrable by the finite second moments. Independence and $\mathbb{E}[U_t] = 0$ give $\mathbb{E}[(\lambda_t - \mu_\lambda)U_t] = 0$, so the first term converges almost surely to zero by the ergodic theorem. The second term also converges almost surely to zero because $\bar{\lambda}_T \to \mu_\lambda$ and $T^{-1}\sum_{t=1}^T U_t \to 0$. Dividing by the positive denominator limit proves $\widehat{k}_T \to k$ almost surely.

The intercept estimator is $\widehat{\rho}_T = \bar{\lambda}_{+,T} - \widehat{k}_T\bar{\lambda}_T$. Both sample means converge almost surely to μ_λ. Hence $\widehat{\rho}_T \to \mu_\lambda - k\mu_\lambda = (1-k)\mu_\lambda = \rho$ almost surely.

For asymptotic normality, denote the numerator by N_T and write $N_T = \sum_{t=1}^T(\lambda_t - \mu_\lambda)U_t - (\bar{\lambda}_T - \mu_\lambda)\sum_{t=1}^T U_t$. Let $\mathcal{F}_t = \sigma(\lambda_1, U_1, \ldots, U_{t-1})$ and define $M_t = (\lambda_t - \mu_\lambda)U_t$. The coefficient λ_t is $\mathcal{F}_t$-measurable, while U_t is independent of $\mathcal{F}_t$ with mean zero. Therefore, M_t is $\mathcal{F}_{t+1}$-measurable and satisfies $\mathbb{E}[M_t \mid \mathcal{F}_t] = 0$.

Its conditional variance is $\mathbb{E}[M_t^2 \mid \mathcal{F}_t] = (\lambda_t - \mu_\lambda)^2\mathrm{Var}(U)$. The ergodic theorem consequently gives $T^{-1}\sum_{t=1}^T \mathbb{E}[M_t^2 \mid \mathcal{F}_t] \to \mathrm{Var}(\lambda_\infty)\mathrm{Var}(U)$ almost surely. Independence also gives $\mathbb{E}[M_t^2] = \mathbb{E}[(\lambda_t - \mu_\lambda)^2]\mathbb{E}[U_t^2] = \mathrm{Var}(\lambda_\infty)\mathrm{Var}(U) < \infty$.

To verify the conditional Lindeberg condition, fix $\epsilon > 0$. Stationarity and iterated expectations imply $\mathbb{E}[T^{-1}\sum_{t=1}^T \mathbb{E}[M_t^2 \mathbb{1}\{|M_t| > \epsilon\sqrt{T}\} \mid \mathcal{F}_t]] = \mathbb{E}[M_1^2 \mathbb{1}\{|M_1| > \epsilon\sqrt{T}\}]$. The right-hand side tends to zero by dominated convergence because $\mathbb{E}[M_1^2] < \infty$. Since the conditional Lindeberg sum is nonnegative, Markov's inequality implies that it converges to zero in probability. The martingale central limit theorem therefore yields $T^{-1/2}\sum_{t=1}^T(\lambda_t - \mu_\lambda)U_t \Rightarrow N(0, \mathrm{Var}(\lambda_\infty)\mathrm{Var}(U))$.

Next, we show that the sample-centering correction is negligible at the $\sqrt{T}$ scale. Subtracting the stationary mean from the archive recursion gives $\lambda_{t+1} - \mu_\lambda = k(\lambda_t - \mu_\lambda) + U_t$. Summing over $t = 1, \ldots, T$ and using $\sum_{t=1}^T(\lambda_{t+1} - \mu_\lambda) = \sum_{t=1}^T(\lambda_t - \mu_\lambda) + \lambda_{T+1} - \lambda_1$ yields $(1-k)T(\bar{\lambda}_T - \mu_\lambda) = \sum_{t=1}^T U_t + \lambda_1 - \lambda_{T+1}$.

The i.i.d. central limit theorem gives $\sum_{t=1}^T U_t = O_p(T^{1/2})$, and stationarity gives $\lambda_1 - \lambda_{T+1} = O_p(1)$. Since $1 - k > 0$, division by $(1-k)T$ gives $\bar{\lambda}_T - \mu_\lambda = O_p(T^{-1/2})$. Consequently, $(\bar{\lambda}_T - \mu_\lambda)\sum_{t=1}^T U_t = O_p(1)$, which is $o_p(\sqrt{T})$. Thus, $T^{-1/2}N_T$ has the same limiting distribution as the martingale sum.

The denominator divided by T converges almost surely to $\mathrm{Var}(\lambda_\infty)$. Slutsky's theorem therefore gives $\sqrt{T}(\widehat{k}_T - k) \Rightarrow N(0, \mathrm{Var}(U)/\mathrm{Var}(\lambda_\infty))$. Theorem 7(i) gives $\mathrm{Var}(\lambda_\infty) = \mathrm{Var}(U)/(1-k^2)$, so the limiting variance is $1 - k^2$. Since $|k| < 1$ and $\widehat{k}_T \to k$ almost surely, eventually $|\widehat{k}_T| < 1$ almost surely. The continuous mapping then gives the feasible asymptotic standard error $\sqrt{(1-\widehat{k}_T^2)/T}$.

Part (ii). Under $H_0 : k = 0$, the archive recursion becomes $\lambda_{t+1} = \rho + U_t$. Conditional on the period-t cross-sectional design, Theorem 6 gives $U_t \mid S_t \sim N(0, \sigma_u^2/S_t)$. Different periods use independent cross-sections, so conditional on the full collection of design sums, the errors remain independent.

Retain $m \geq 3$ observations with odd indices $t \in \mathcal{I}$. In the regression observation indexed by t, the dependent variable is $\lambda_{t+1} = \rho + U_t$ and the regressor is $\lambda_t = \rho + U_{t-1}$. The regression errors therefore use odd innovation indices, whereas the retained regressors depend on even innovation indices. These sets are disjoint. Conditional on the design sums and the retained regressors, the errors U_t, $t \in \mathcal{I}$, consequently remain independent centered Gaussian variables with variances σ_u^2/S_t.

Weighted least squares with weight S_t is equivalent to multiplying the dependent variable, the intercept regressor, and the slope regressor in observation t by $\sqrt{S_t}$. The transformed design therefore has columns $\sqrt{S_t}$ and $\sqrt{S_t}\lambda_t$. Its regression error is $\sqrt{S_t}U_t$, which is conditionally $N(0, \sigma_u^2)$ in every retained observation. The transformed errors are independent, so the retained sample is a normal linear regression with two coefficients, fixed conditional design, and homoskedastic errors.

The design has rank 2 almost surely because $S_t > 0$ almost surely and the retained regressors are not all equal almost surely under the nondegenerate continuous innovation law. Conditional normal regression theory therefore implies that the slope error divided by its conditional standard deviation is standard normal. The weighted residual sum of squares divided by σ_u^2 has a χ^2_{m-2} distribution and is independent of

the slope estimator. Replacing σ_u^2 by the weighted residual sum of squares divided by $m-2$ consequently produces a Studentized slope statistic with a t_{m-2} distribution under the null.

This conditional distribution does not depend on the realized design sums or retained regressors. Integrating over those conditioning variables therefore gives the same unconditional t_{m-2} distribution. This proves equation (70).

Part (iii). Fix the deployed coefficient λ in a single randomized cross-section. The treated units face the score-dependent allocation and therefore have a population slope $\beta_T = \rho + k\lambda$. Holdout units receive the baseline allocation and have a population slope $\beta_H = \rho$. Subtracting gives $\beta_T - \beta_H = k\lambda$.

Let n_T and n_H denote the treated and holdout sample sizes. Since assignment is independent of the primitives and the holdout probability is $h \in (0,1)$, the law of large numbers gives $n_T/n \to 1-h$ and $n_H/n \to h$ in probability. Both group sizes therefore diverge in probability. Random assignment also preserves the joint distribution of the feature and regression residual within each group.

Under the maintained Gaussian assumptions, the centered feature $\widetilde{x}_i$ has mean zero and variance V, and the regression residual u_i has mean zero and variance σ_u^2 and is independent of $\widetilde{x}_i$. Thus, $\mathbb{E}[\widetilde{x}_i u_i] = 0$ and $\mathbb{E}[\widetilde{x}_i^2 u_i^2] = \mathbb{E}[\widetilde{x}_i^2]\mathbb{E}[u_i^2] = V\sigma_u^2$. This independence supplies the variance factorization used below.

For either group $g \in \{T, H\}$, let $\overline{\widetilde{x}}_g$ and $\bar{u}_g$ denote the group sample means. The OLS slope with an intercept satisfies $\widehat{\beta}_g - \beta_g = \sum_{i\in g}(\widetilde{x}_i - \overline{\widetilde{x}}_g)u_i / \sum_{i\in g}(\widetilde{x}_i - \overline{\widetilde{x}}_g)^2$. The law of large numbers gives $n_g^{-1}\sum_{i\in g}(\widetilde{x}_i - \overline{\widetilde{x}}_g)^2 \to V$ in probability.

For the numerator, $n_g^{-1/2}\sum_{i\in g}(\widetilde{x}_i - \overline{\widetilde{x}}_g)u_i = n_g^{-1/2}\sum_{i\in g}\widetilde{x}_i u_i - \sqrt{n_g}\,\overline{\widetilde{x}}_g \bar{u}_g$. Finite second moments give $\overline{\widetilde{x}}_g = O_p(n_g^{-1/2})$ and $\bar{u}_g = O_p(n_g^{-1/2})$, so the second term is $O_p(n_g^{-1/2}) = o_p(1)$. The central limit theorem applied to $\widetilde{x}_i u_i$ gives $n_g^{-1/2}\sum_{i\in g}\widetilde{x}_i u_i \Rightarrow N(0, V\sigma_u^2)$. Dividing by the denominator limit yields $\sqrt{n_g}(\widehat{\beta}_g - \beta_g) \Rightarrow N(0, \sigma_u^2/V)$.

Applying this result to the two groups gives $\sqrt{n_T}(\widehat{\beta}_T - \beta_T) \Rightarrow N(0, \sigma_u^2/V)$ and $\sqrt{n_H}(\widehat{\beta}_H - \beta_H) \Rightarrow N(0, \sigma_u^2/V)$. Multiplying the first expression by $\sqrt{n/n_T}$ and the second by $\sqrt{n/n_H}$, Slutsky's theorem gives $\sqrt{n}(\widehat{\beta}_T - \beta_T) \Rightarrow N(0, \sigma_u^2/[V(1-h)])$ and $\sqrt{n}(\widehat{\beta}_H - \beta_H) \Rightarrow N(0, \sigma_u^2/[Vh])$.

Conditional on the assignment, the two regressions use disjoint sets of independent units. Their leading influence sums are therefore independent. Since the sample proportions converge to constants, the joint limiting distribution is Gaussian with zero covariance between the two slope errors. The asymptotic variance of their difference is consequently the sum of the two variances.

Finally, substituting $\beta_T - \beta_H = k\lambda$ gives $\sqrt{n}[(\widehat{\beta}_T - \widehat{\beta}_H) - k\lambda] \Rightarrow N(0, (\sigma_u^2/V)[1/(1-h) + 1/h])$. This is equation (71). □

Proof of Proposition 9. Index the first cross-section by $s = 1$, and let λ_1 be the initial deployed coefficient, independent of all subsequent cross-sections. Write $x_{is}^c = \widetilde{x}_{i,s} - \bar{x}_s$ and $u_{is}^c = u_{i,s} - \bar{u}_s$. The centered outcome equation is $y_{is} - \bar{y}_s = (\rho + k\lambda_s)x_{is}^c + u_{is}^c$. Define $S_s = \sum_i (x_{is}^c)^2$ and $\xi_s = \sum_i x_{is}^c u_{is}^c$. The contribution of cross-section s to the pooled numerator is therefore $\sum_i x_{is}^c(y_{is} - \bar{y}_s) = (\rho + k\lambda_s)S_s + \xi_s$.

Let $A_t = \sum_{s=1}^t S_s$, $\mu_t = A_t^{-1}\sum_{s=1}^t S_s\lambda_s$, and $\nu_t = A_t^{-1}\sum_{s=1}^t \xi_s$. Summing the numerator contributions and dividing by A_t gives $\lambda_{t+1} = \rho + k\mu_t + \nu_t$. These quantities are well defined because $S_s > 0$ almost surely. In particular, $\mu_1 = \lambda_1$ and $\nu_1 = \xi_1/S_1$.

The Gaussian sampling assumptions imply that S_s/V has a chi-square distribution with $n-1$ degrees of freedom. Conditional on the design sums, the variables ξ_s are independent Gaussian variables with mean zero and variance $\sigma_u^2 S_s$. Thus $\mathbb{E}\xi_s = 0$ and $\mathbb{E}\xi_s^2 = \sigma_u^2\mathbb{E}S_s = \sigma_u^2 V(n-1) < \infty$. Fresh cross-sections make the pairs (S_s, ξ_s) i.i.d. Consequently, the strong law gives $A_t/t \to \mathbb{E}S_1 = V(n-1) > 0$ and $t^{-1}\sum_{s=1}^t \xi_s \to 0$ almost surely. Dividing these limits yields $\nu_t \to 0$ almost surely.

For $s \geq 2$, define $\gamma_s = S_s/A_s \in (0,1)$. Since $A_{t+1} = A_t + S_{t+1}$, the weighted mean satisfies $\mu_{t+1} = (1-\gamma_{t+1})\mu_t + \gamma_{t+1}\lambda_{t+1}$. Substitution of $\lambda_{t+1} = \rho + k\mu_t + \nu_t$ gives $\mu_{t+1} = \mu_t + \gamma_{t+1}[\rho + (k-1)\mu_t + \nu_t]$ for every $t \geq 1$.

We first establish the gain properties used below. Since S_s has a finite second moment, $\Pr(S_s > \epsilon s) \le \mathbb{E}[S_1^2]/(\epsilon^2 s^2)$ for every $\epsilon > 0$. Borel–Cantelli therefore gives $S_s/s \to 0$ almost surely. Together with $A_s/s \to \mathbb{E}S_1 > 0$, this implies $\gamma_s \to 0$ almost surely.

Also, $\mathbb{E}[\sum_{s=2}^{\infty} S_s^2/s^2] = \mathbb{E}[S_1^2] \sum_{s=2}^{\infty} s^{-2} < \infty$, so $\sum_{s=2}^{\infty} S_s^2/s^2 < \infty$ almost surely. Eventually $A_s \ge cs$ for a fixed $c > 0$, which implies $\sum_{s=2}^{\infty} \gamma_s^2 < \infty$ almost surely.

For $s \ge 2$, the identity $1 - \gamma_s = A_{s-1}/A_s$ gives $-\log(1-\gamma_s) = \log A_s - \log A_{s-1}$. Because $\gamma_s \to 0$ and its squares are summable, the nonnegative series $\sum_{s=2}^{\infty}[-\log(1-\gamma_s) - \gamma_s]$ converges. Telescoping therefore gives $\sum_{s=2}^{t} \gamma_s = \log A_t - \log A_1 + O(1) = \log t + O(1)$ almost surely. In particular, $\sum_{s=2}^{\infty} \gamma_s = \infty$.

Part (i). Suppose $k < 1$, define $\lambda^* = \rho/(1-k)$, and set $e_t = \mu_t - \lambda^*$. Since $\rho + (k-1)\lambda^* = 0$, the weighted-mean recursion becomes $e_{t+1} = [1 - (1-k)\gamma_{t+1}]e_t + \gamma_{t+1}\nu_t$.

Fix $\epsilon > 0$. Almost surely, there is a finite t_0 such that $|\nu_t| \le (1-k)\epsilon$ and $0 < (1-k)\gamma_{t+1} < 1$ for every $t \ge t_0$. Hence $|e_{t+1}| \le [1 - (1-k)\gamma_{t+1}]|e_t| + (1-k)\gamma_{t+1}\epsilon$. Define $z_t = \max\{|e_t| - \epsilon, 0\}$. Subtracting ϵ from the preceding inequality and taking positive parts gives $z_{t+1} \le [1 - (1-k)\gamma_{t+1}]z_t$.

Iteration yields $z_t \le z_{t_0} \prod_{s=t_0+1}^{t}[1-(1-k)\gamma_s] \le z_{t_0} \exp[-(1-k)\sum_{s=t_0+1}^{t} \gamma_s]$. The sum in the exponent diverges, so $z_t \to 0$. Thus $\limsup_t |e_t| \le \epsilon$. Letting ϵ decrease to zero proves $e_t \to 0$ almost surely. Therefore $\mu_t \to \lambda^*$, and $\lambda_{t+1} = \rho + k\mu_t + \nu_t \to \rho + k\lambda^* = \lambda^*$ almost surely. This argument applies to every $k < 1$, including $k \le -1$.

Part (ii). Suppose $k = 1$. Iterating the weighted-mean recursion from date one gives $\mu_t = \mu_1 + \rho\sum_{s=2}^{t} \gamma_s + \sum_{s=2}^{t} \gamma_s \nu_{s-1}$. We show that the final series converges absolutely almost surely.

Let $\mathcal{D}_{s-1} = \sigma(S_1, \ldots, S_{s-1})$. Conditional on $\mathcal{D}_{s-1}$, the numerator of ν_{s-1} is Gaussian with variance $\sigma_u^2 A_{s-1}$. Therefore $\mathbb{E}[|\nu_{s-1}| \mid \mathcal{D}_{s-1}] = \sqrt{2/\pi}\,\sigma_u A_{s-1}^{-1/2}$. Since $\gamma_s \le S_s/A_{s-1}$ and S_s is independent of all preceding cross-sections, it follows that $\mathbb{E}[\gamma_s|\nu_{s-1}|] \le \sqrt{2/\pi}\,\sigma_u \mathbb{E}[S_1]\mathbb{E}[A_{s-1}^{-3/2}]$ whenever the inverse moment is finite.

The variable A_{s-1}/V is chi-square with $(n-1)(s-1)$ degrees of freedom. Its inverse moment of order $3/2$ is finite for all sufficiently large s and is $O(s^{-3/2})$. Hence, for some finite integer $s_0 \ge 2$, $\sum_{s=s_0}^{\infty} \mathbb{E}[\gamma_s|\nu_{s-1}|] < \infty$. Tonelli's theorem implies that the corresponding random series is finite almost surely. Each of the finitely many terms with $2 \le s < s_0$ is also finite almost surely, even if its expectation is infinite. Thus $\sum_{s=2}^{\infty} \gamma_s|\nu_{s-1}| < \infty$ almost surely.

Combining this convergence with $\sum_{s=2}^{t} \gamma_s = \log t + O(1)$ gives $\mu_t = \rho \log t + O(1)$ almost surely. Since $\lambda_{t+1} = \rho + \mu_t + \nu_t$ and $\nu_t \to 0$, we obtain $\lambda_t/\log t \to \rho$, proving (73). If $\rho > 0$, the coefficient diverges to $+\infty$. If $\rho = 0$, the gain-sum term disappears, so μ_t converges to a finite random limit and $\lambda_{t+1} = \mu_t + \nu_t$ converges to the same limit.

Part (iii). Suppose $k > 1$, let $\widetilde{\lambda} = \rho/(1-k)$, and set $e_t = \mu_t - \widetilde{\lambda}$. Then $e_{t+1} = [1 + (k-1)\gamma_{t+1}]e_t + \gamma_{t+1}\nu_t$. Define $\Pi_1 = 1$ and $\Pi_t = \prod_{s=2}^{t}[1 + (k-1)\gamma_s]$ for $t \ge 2$.

Since $\gamma_s \to 0$ and $\sum_s \gamma_s^2 < \infty$, expansion of the logarithm gives $\log \Pi_t = (k-1)\sum_{s=2}^{t} \gamma_s + O(1) = (k-1)\log t + O(1)$ almost surely. Exponentiating shows that $\Pi_t \asymp t^{k-1}$ almost surely. Also $\Pi_t \ge 1$.

Dividing the error recursion by Π_{t+1} gives $e_{t+1}/\Pi_{t+1} = e_t/\Pi_t + \gamma_{t+1}\nu_t/\Pi_{t+1}$. Iterating from date one yields $e_t/\Pi_t = e_1 + L_t$, where $L_t = \sum_{s=2}^{t} \gamma_s \nu_{s-1}/\Pi_s$. The absolute summability established in part (ii) and $\Pi_s \ge 1$ imply that L_t converges absolutely almost surely to a finite random variable L.

It remains to show that $e_1 + L$ is nonzero almost surely. We isolate the contribution of the first regression disturbance. Since $\nu_{s-1} = A_{s-1}^{-1}\sum_{r=1}^{s-1} \xi_r$, write $L_t = w_{1,t}\xi_1 + R_t$, where $w_{1,t} = \sum_{s=2}^{t} \gamma_s/(\Pi_s A_{s-1})$ and R_t depends only on the design sums and on $\xi_2, \ldots, \xi_{t-1}$.

The coefficients $w_{1,t}$ are increasing and strictly positive for $t \ge 2$. Moreover, $\gamma_s/A_{s-1} = S_s/(A_s A_{s-1}) = A_{s-1}^{-1} - A_s^{-1}$. Since $\Pi_s \ge 1$, telescoping gives $w_{1,t} \le A_1^{-1} - A_t^{-1} \le A_1^{-1}$. Thus $w_{1,t} \to w_1$ for a finite random variable satisfying $0 < w_1 \le A_1^{-1}$. Because $L_t \to L$ and ξ_1 is finite almost surely, $R_t = L_t - w_{1,t}\xi_1$ also converges almost surely to a finite limit R. This limit is measurable with respect to the design sums and the disturbances $\{\xi_s\}_{s \ge 2}$.

Conditional on the entire design sequence, the initial coefficient, and $\{\xi_s\}_{s\geq 2}$, the variable ξ_1 is Gaussian with mean zero and variance $\sigma_u^2 S_1 > 0$. The variables $e_1 = \lambda_1 - \widetilde{\lambda}$, w_1, and R are fixed under this conditioning. Therefore, $e_1 + L = e_1 + R + w_1\xi_1$ has a conditionally nondegenerate Gaussian distribution. Its conditional probability of equaling zero is zero, so $\Pr(e_1 + L = 0) = 0$.

Consequently, $e_t/\Pi_t = e_1 + L_t \to e_1 + L \neq 0$ almost surely. Together with $\Pi_t \asymp t^{k-1}$, this gives $|e_t| \asymp t^{k-1}$ almost surely. Finally, $\lambda_{t+1} = \rho + k(\widetilde{\lambda} + e_t) + \nu_t = \widetilde{\lambda} + ke_t + \nu_t$. Since $\nu_t \to 0$ and $|e_t| \to \infty$, we have $|\lambda_{t+1}|/|e_t| \to k > 0$. Hence $|\lambda_t| \asymp t^{k-1}$ and $|\lambda_t| \to \infty$ almost surely, including when $\lambda_1 = \widetilde{\lambda}$. □

Proof of Corollary 10. The population update is the affine scalar map $F(\lambda) = \rho + \kappa\lambda$. Its derivative is the constant $F'(\lambda) = \kappa$. Since the maintained primitives imply $\kappa \geq 0$, the fixed point is attracting if and only if $|\kappa| < 1$, which is equivalent to $\kappa < 1$. Because $\kappa = k\psi$ and $\psi > 0$, this condition is $k < 1/\psi$. Substituting equation (78) gives $k_c = 1/\psi = (1 - \delta\phi)/[1 - \delta\phi(1 - \rho)]$, proving equation (80).

Let $N = 1 - \delta\phi$ and $D = 1 - \delta\phi(1 - \rho)$. Then $k_c = N/D$. Since $D > 0$, the quotient rule is valid. Differentiate with respect to δ. We have $N_\delta = -\phi$ and $D_\delta = -\phi(1 - \rho)$. Hence

$$\frac{\partial k_c}{\partial \delta} = \frac{(-\phi)D - N[-\phi(1-\rho)]}{D^2} = \frac{\phi[-D + N(1-\rho)]}{D^2}.$$

Now $-D + N(1 - \rho) = -[1 - \delta\phi(1 - \rho)] + (1 - \delta\phi)(1 - \rho) = -1 + \delta\phi(1 - \rho) + 1 - \rho - \delta\phi(1 - \rho) = -\rho$. Therefore $\partial k_c/\partial\delta = -\rho\phi/D^2$.

Similarly, $N_\phi = -\delta$ and $D_\phi = -\delta(1 - \rho)$, so the same algebra gives $\partial k_c/\partial\phi = -\rho\delta/D^2$.

For ρ, $N_\rho = 0$ and $D_\rho = \delta\phi$. Hence $\partial k_c/\partial\rho = -(N\delta\phi)/D^2 = -\delta\phi(1 - \delta\phi)/D^2$.

All denominators are positive. The three derivatives are therefore weakly negative. The first is strictly negative exactly when $\rho\phi > 0$, the second exactly when $\rho\delta > 0$, and the third exactly when $\delta\phi > 0$ because $1 - \delta\phi > 0$.

If $\delta = 0$ or $\phi = 0$, then $\delta\phi = 0$, so $\psi = 1$ and $\kappa = k$. If $\rho = 0$, then $\psi = [1 - \delta\phi]/[1 - \delta\phi] = 1$, so again $\kappa = k$. In all three cases $k_c = 1$. Finally, for fixed $\rho > 0$, as $\delta\phi \uparrow 1$, the numerator $1 - \delta\phi \to 0$ while the denominator $1 - \delta\phi(1 - \rho) \to \rho > 0$. Hence $k_c \to 0$. □

Proof of Proposition 10. We first derive the probability limit of the coefficient matrix. Equation (89) gives $A_t = \begin{pmatrix} k & \beta\delta\phi/s_t^{xx} \\ \gamma s_t^{\theta x} & \delta\phi \end{pmatrix}$. For each fixed date, the panel observations are i.i.d. across units. By the cross-sectional law of large numbers, $s_t^{xx} = n^{-1}\sum_i (\widetilde{x}_{i,t}^c)^2 \to_p \mathrm{Var}(\widetilde{x}_{i,t}) = V$ and $s_t^{\theta x} = n^{-1}\sum_i \widetilde{\theta}_{i,t}^c \widetilde{x}_{i,t}^c \to_p \mathrm{Cov}(\widetilde{\theta}_{i,t}, \widetilde{x}_{i,t}) = \sigma_\theta^2$. Since $V > 0$, continuous mapping gives $1/s_t^{xx} \to_p 1/V$. Therefore, each entry of A_t converges in probability to the corresponding entry of $\overline{A}$ in equation (90).

Write $d = \delta\phi$. The characteristic polynomial is

$$\det(\mu I - \overline{A}) = \det \begin{pmatrix} \mu - k & -\beta d/V \\ -\gamma\sigma_\theta^2 & \mu - d \end{pmatrix} = (\mu - k)(\mu - d) - \frac{\beta\gamma d\sigma_\theta^2}{V}.$$

Since $\beta\gamma = k$ and $\sigma_\theta^2/V = \rho$, the last term is $kd\rho$. Expanding the first product gives $(\mu - k)(\mu - d) = \mu^2 - (k + d)\mu + kd$. Therefore the polynomial equals $\mu^2 - (k + d)\mu + kd - kd\rho = \mu^2 - (k + d)\mu + kd(1 - \rho)$, proving equation (91).

The discriminant is

$$\Delta = (k + d)^2 - 4kd(1 - \rho) = k^2 + 2kd + d^2 - 4kd + 4kd\rho = (k - d)^2 + 4kd\rho \geq 0.$$

Hence both eigenvalues are real. The quadratic formula gives equation (92). Their sum is $k + d \geq 0$, and their product is $kd(1 - \rho) \geq 0$. A pair of real roots with nonnegative product has the same sign or contains

zero. Their nonnegative sum rules out two negative roots. Thus both roots are nonnegative. The smaller root is strictly positive exactly when the product is strictly positive, which is exactly $kd(1-\rho) > 0$.

Assume now that $kd\rho > 0$. Let $P(\mu) = \mu^2 - (k+d)\mu + kd(1-\rho)$. Evaluating at k gives $P(k) = k^2 - (k+d)k + kd(1-\rho) = k^2 - k^2 - kd + kd - kd\rho = -kd\rho < 0$. Likewise $P(d) = d^2 - (k+d)d + kd(1-\rho) = -kd\rho < 0$. Since P is an upward-opening parabola and is negative strictly between its two real roots, both k and d lie strictly between μ_2 and μ_1. Therefore $\mu_1 > \max\{k, d\}$ and $\mu_2 < \min\{k, d\}$.

We next prove the stability equivalence. Evaluate the characteristic polynomial at one:

$$P(1) = 1 - k - d + kd(1-\rho).$$

From equation (78), $\psi = [1 - d(1-\rho)]/(1-d)$. Hence $(1-d)(1-\kappa) = (1-d) - k[1-d(1-\rho)] = 1 - d - k + kd(1-\rho) = P(1)$. Thus equation (94) holds.

If $\kappa > 1$, then $P(1) < 0$. Since P is negative between the two roots, 1 lies strictly between μ_2 and μ_1, so $\mu_1 > 1$. Hence the spectral radius exceeds one. If $\kappa = 1$, then $P(1) = 0$, so one root equals one, and the spectral radius is one.

Suppose $\kappa < 1$. Then $P(1) > 0$. Since $\psi \geq 1$, $k \leq \kappa < 1$. Also $d = \delta\phi < 1$. Hence the sum of the roots is $k + d < 2$. If both nonnegative roots were at least one, their sum would be at least two, a contradiction. Since $P(1) > 0$, the point one is outside the interval between the roots. It cannot lie below both roots because that would make both roots exceed one and their sum exceed two. Therefore, one lies above both roots, so $\mu_1 < 1$. Thus, the spectral radius is below one. This proves that the spectral radius is below one if and only if $\kappa < 1$.

Finally, solve for the stationary state of the deterministic recursion $z_{t+1} = \overline{A} z_t + (\rho, 0)'$. Write $z^* = (\bar{\lambda}^*, \bar{p}^*)'$. The second row of $z^* = \overline{A} z^* + (\rho, 0)'$ is $\bar{p}^* = \gamma\sigma_\theta^2 \bar{\lambda}^* + d\bar{p}^*$. Rearranging gives $(1-d)\bar{p}^* = \gamma\sigma_\theta^2 \bar{\lambda}^*$, so $\bar{p}^* = \gamma\sigma_\theta^2 \bar{\lambda}^*/(1-d)$.

The first row is $\bar{\lambda}^* = k\bar{\lambda}^* + (\beta d/V)\bar{p}^* + \rho$. Substitute the expression for $\bar{p}^*$ to obtain

$$\bar{\lambda}^* = k\bar{\lambda}^* + \frac{\beta d}{V}\frac{\gamma\sigma_\theta^2}{1-d}\bar{\lambda}^* + \rho = k\bar{\lambda}^* + \frac{kd\rho}{1-d}\bar{\lambda}^* + \rho.$$

Move the slope terms to the left:

$$\left[1 - k - \frac{kd\rho}{1-d}\right]\bar{\lambda}^* = \rho.$$

The bracket equals $1 - k[1 + d\rho/(1-d)] = 1 - k\psi = 1 - \kappa$. Hence $\bar{\lambda}^* = \rho/(1-\kappa)$. Substituting this into the second-row solution gives equation (93). □

Proof of Lemma 2. Let $x_t^c = (\widetilde{x}_{1,t}^c, \ldots, \widetilde{x}_{n,t}^c)'$ and similarly for the other centered vectors. Cross-sectional OLS with an intercept gives

$$\lambda_{t+1} = \frac{(x_t^c)' y_t^c}{S_t}, \qquad S_t = (x_t^c)' x_t^c.$$

The centered outcome equation is $y_t^c = \widetilde{\theta}_t^c + \beta K_t^c + \varepsilon_t^c$. The centered capital equation is $K_t^c = \delta K_{t-1}^c + \gamma\lambda_t x_t^c$. Substitute the capital equation into the outcome equation and then into OLS:

$$\lambda_{t+1} = \frac{(x_t^c)'(\widetilde{\theta}_t^c + \varepsilon_t^c)}{S_t} + \frac{\beta(x_t^c)'[\delta K_{t-1}^c + \gamma\lambda_t x_t^c]}{S_t}.$$

Expand the second term:

$$\lambda_{t+1} = g_t + \frac{\beta\delta}{S_t}(x_t^c)' K_{t-1}^c + \beta\gamma\lambda_t \frac{(x_t^c)' x_t^c}{S_t}.$$

The final ratio equals one and $\beta\gamma = k$, so

$$\lambda_{t+1} = k\lambda_t + \frac{\beta\delta}{S_t}(x_t^c)'K_{t-1}^c + g_t.$$

The capital vector satisfies $K_t^c = \gamma x_t^c \lambda_t + \delta K_{t-1}^c$. Thus the full state $W_t = (\lambda_t, (K_{t-1}^c)')'$ obeys an affine linear recurrence.

Because all capital vectors are cross-sectionally centered, $K_{t-1}^c \in \mathbb{1}^\perp$. Let Q have orthonormal columns spanning $\mathbb{1}^\perp$. Then QQ' is the orthogonal projection onto $\mathbb{1}^\perp$, so $K_{t-1}^c = QQ'K_{t-1}^c$. Define $q_{t-1} = Q'K_{t-1}^c$. Then $(x_t^c)'K_{t-1}^c = (x_t^c)'Qq_{t-1}$. Multiplying the capital equation by Q' gives $q_t = \gamma Q'x_t^c\lambda_t + \delta q_{t-1}$. Combining this equation with the OLS equation yields

$$\begin{pmatrix}\lambda_{t+1}\\ q_t\end{pmatrix} = \begin{pmatrix}k & (\beta\delta/S_t)(x_t^c)'Q\\ \gamma Q'x_t^c & \delta I_{n-1}\end{pmatrix}\begin{pmatrix}\lambda_t\\ q_{t-1}\end{pmatrix} + \begin{pmatrix}g_t\\ 0\end{pmatrix},$$

which is exactly equations (96)–(98).

We now derive the marginal distribution of g_t. At a fixed date, the pair $(\widetilde{\theta}_{i,t}, \widetilde{x}_{i,t})$ is jointly Gaussian with $\mathrm{Cov}(\widetilde{\theta}, \widetilde{x}) = \sigma_\theta^2$ and $\mathrm{Var}(\widetilde{x}) = V$. Hence the linear projection is $\widetilde{\theta}_{i,t} = \rho\widetilde{x}_{i,t} + v_{i,t}$, where $v_{i,t}$ is Gaussian, independent of $\widetilde{x}_{i,t}$, with variance $\sigma_\theta^2 - \rho^2 V = \sigma_\theta^2 - \sigma_\theta^4/V = \sigma_\theta^2(1-\rho)$. Adding the independent outcome shock gives $r_{i,t} = v_{i,t} + \varepsilon_{i,t}$, which is independent of the feature and has variance $\sigma_r^2 = \sigma_\theta^2(1-\rho) + \sigma_\varepsilon^2$.

After demeaning, $\widetilde{\theta}_t^c + \varepsilon_t^c = \rho x_t^c + r_t^c$. Therefore

$$g_t = \frac{(x_t^c)'(\rho x_t^c + r_t^c)}{S_t} = \rho + \frac{(x_t^c)'r_t^c}{S_t}.$$

Since $(x_t^c)'\mathbb{1} = 0$, $(x_t^c)'r_t^c = (x_t^c)'r_t$. Conditional on x_t^c, the numerator $(x_t^c)'r_t$ is Gaussian with mean zero and variance $\sigma_r^2 S_t$. Thus $g_t - \rho = \sigma_r Z_t/\sqrt{S_t}$ for a standard normal Z_t independent of S_t. The Gaussian feature design gives $S_t/V \sim \chi_{n-1}^2$. Hence

$$g_t - \rho = \frac{\sigma_r}{\sqrt{V(n-1)}}\frac{Z_t}{\sqrt{(S_t/V)/(n-1)}},$$

which is equation (99).

For $n = 2$, the Student variable has one degree of freedom and is Cauchy, so its mean is undefined. For $n = 3$, it has two degrees of freedom, so the mean exists and equals zero but the variance is infinite. For $n \geq 4$, its variance is $(n-1)/(n-3)$. Multiplying by the squared scale gives $\mathrm{Var}(g_t) = \sigma_r^2/[V(n-3)]$, which is equation (100).

We next verify the logarithmic and dependence claims. The lower-left block of $\widetilde{B}_t$ has norm $|\gamma|\|Q'x_t^c\| \leq |\gamma|\|x_t^c\| = |\gamma|\sqrt{S_t}$. The upper-right block has norm $|\beta|\delta\|(x_t^c)'Q\|/S_t \leq |\beta|\delta\|x_t^c\|/S_t = |\beta|\delta/\sqrt{S_t}$. Therefore, for a finite constant C, $\|\widetilde{B}_t\| \leq C[1 + \sqrt{S_t} + S_t^{-1/2}]$. Taking positive logarithms gives $\log^+\|\widetilde{B}_t\| \leq C' + \frac{1}{2}|\log S_t|$. Since $S_t/V \sim \chi_{n-1}^2$, $S_t > 0$ almost surely and $\mathbb{E}|\log S_t| < \infty$ for every $n \geq 2$. Thus $\mathbb{E}\log^+\|\widetilde{B}_t\| < \infty$. Likewise, g_t is a shifted scaled Student variable, and Student variables have finite logarithmic moments, so $\mathbb{E}\log^+\|\widetilde{E}_t\| < \infty$.

Finally, the vector of panel fundamentals is an n-dimensional stable Gaussian AR(1) with coefficient $\phi \in [0, 1)$ and is geometrically β-mixing. Adding independent contemporaneous measurement and outcome shocks preserves stationarity and geometric mixing. The pair $(\widetilde{B}_t, \widetilde{E}_t)$ is a measurable function of the period-t primitive panel state. Measurable transformations cannot increase the β-mixing coefficients, so the coefficient-forcing sequence is stationary and geometrically β-mixing. The formulas contain no functional dependence on the deployed score or inherited capital, although serial dependence remains through the persistent primitive fundamentals. □

Proof of Theorem 8. Lemma 2 establishes that $\{(\widetilde{B}_t, \widetilde{E}_t)\}$ is stationary and ergodic and that the logarithmic moments in (102) are finite. The theorem also assumes that the top Lyapunov exponent $\gamma_{L,n}$ is negative. We first establish convergence of the causal series, allowing $\gamma_{L,n} = -\infty$, and then verify the remaining claims.

Fix a date t. Choose a finite number b satisfying $\gamma_{L,n} < b < 0$, and choose $\epsilon > 0$ such that $b + \epsilon < 0$. These choices are possible whether $\gamma_{L,n}$ is finite or equals $-\infty$. The almost-sure backward Lyapunov limit implies that there is a finite random integer J_1 such that $\|\widetilde{B}_{t-1} \cdots \widetilde{B}_{t-j}\| \le e^{bj}$ for every $j \ge J_1$.

The logarithmic moment condition on the forcing sequence controls its growth backward in time. Define $Y_j = \log^+ \|\widetilde{E}_{t-j-1}\|$ for $j \ge 0$. Stationarity implies that every Y_j has the same distribution as Y_0, and (102) gives $\mathbb{E}Y_0 < \infty$. The tail-sum bound therefore yields $\sum_{j=1}^{\infty} \Pr(Y_j > \epsilon j) = \sum_{j=1}^{\infty} \Pr(Y_0 > \epsilon j) < \infty$. By the first Borel–Cantelli lemma, there is a finite random integer J_2 such that $\|\widetilde{E}_{t-j-1}\| \le e^{\epsilon j}$ for every $j \ge J_2$, almost surely. This argument does not require independence across dates.

For every $j \ge \max\{J_1, J_2\}$, submultiplicativity gives $\|\widetilde{B}_{t-1} \cdots \widetilde{B}_{t-j}\widetilde{E}_{t-j-1}\| \le e^{bj}e^{\epsilon j} = e^{(b+\epsilon)j}$. Since $b + \epsilon < 0$, the geometric series $\sum_{j=0}^{\infty} e^{(b+\epsilon)j}$ converges. The finitely many preceding terms are finite almost surely. Thus the vector series in (103) converges absolutely almost surely. Taking a countable intersection over integer dates gives simultaneous convergence for every $t \in \mathbb{Z}$.

Define the sum as $\widetilde{W}_t^*$. Absolute convergence permits multiplication by the finite matrix $\widetilde{B}_{t-1}$ and reindexing of the series. Multiplying the series for $\widetilde{W}_{t-1}^*$ by $\widetilde{B}_{t-1}$ reproduces every term in the series for $\widetilde{W}_t^*$ except its first term, $\widetilde{E}_{t-1}$. Consequently, $\widetilde{W}_t^* = \widetilde{B}_{t-1}\widetilde{W}_{t-1}^* + \widetilde{E}_{t-1}$, which is the time-shifted form of (96).

The solution is causal because $\widetilde{W}_t^*$ is measurable with respect to the coefficient-forcing pairs dated at or before $t-1$. Its defining function commutes with time shifts, so stationarity of the input sequence implies stationarity of $\{\widetilde{W}_t^*\}$. Ergodicity follows because a measurable function of a stationary ergodic sequence that commutes with time shifts is itself stationary and ergodic.

To establish contraction of initial conditions, let $\widetilde{W}_t$ and $\widetilde{W}_t'$ be two solutions driven by the same primitive sequence, with finite initial states. Their difference $D_t = \widetilde{W}_t - \widetilde{W}_t'$ satisfies $D_{t+1} = \widetilde{B}_t D_t$, and iteration gives $D_t = \widetilde{B}_{t-1} \cdots \widetilde{B}_0 D_0$. The forward Lyapunov limit and the choice $\gamma_{L,n} < b < 0$ imply that $\|\widetilde{B}_{t-1} \cdots \widetilde{B}_0\| \le e^{bt}$ for all sufficiently large t, almost surely. Hence $\|D_t\| \le e^{bt}\|D_0\|$ eventually. Since $\|D_0\| < \infty$ almost surely, the difference converges to zero exponentially almost surely. This argument also covers $D_0 = 0$ and products that become zero.

The stationary random-affine-recurrence results of Brandt (1986) and Bougerol and Picard (1992) give uniqueness of the causal stationary solution under stationarity, ergodicity, the stated logarithmic moment conditions, and a negative top Lyapunov exponent. All these conditions hold here. Finally, the first coordinate of a stationary ergodic vector process is stationary and ergodic, so the vintage-slope process has the stated property. □

Proof of Theorem 9. We first treat $\delta > 0$. The boundary case $\delta = 0$ is handled at the end. The proof has three steps.

Step 1. Balance the finite-panel matrices and take a fixed-block large-panel limit. For each panel size n, define $D_n = \operatorname{diag}(1, n^{-1/2}I_{n-1})$ and $\widehat{B}_{n,t} = D_n\widetilde{B}_{n,t}D_n^{-1}$. For a product $P_{n,t} = \widetilde{B}_{n,t} \cdots \widetilde{B}_{n,1}$, the corresponding balanced product is $\widehat{P}_{n,t} = D_n P_{n,t} D_n^{-1}$. Submultiplicativity gives $\|\widehat{P}_{n,t}\| \le \|D_n\|\,\|P_{n,t}\|\,\|D_n^{-1}\|$. Conversely, $P_{n,t} = D_n^{-1}\widehat{P}_{n,t}D_n$, so $\|P_{n,t}\| \le \|D_n^{-1}\|\,\|\widehat{P}_{n,t}\|\,\|D_n\|$. Taking logs of both inequalities yields

$$|\log\|\widehat{P}_{n,t}\| - \log\|P_{n,t}\|| \le \log\|D_n\| + \log\|D_n^{-1}\|.$$

Divide by t. For fixed n, the right-hand side is a constant divided by t, so it tends to zero. Thus the similarity transformation leaves the Lyapunov exponent unchanged.

Define $v_{n,t} = n^{-1/2}Q'\widetilde{x}_t^c$ and $s_{n,t} = S_t/n$. Since $\|Q'\widetilde{x}_t^c\|^2 = (\widetilde{x}_t^c)'QQ'\widetilde{x}_t^c = S_t$, we have $s_{n,t} = \|v_{n,t}\|^2$. We now compute the balanced matrix explicitly. The upper-right block of $\widetilde{B}_{n,t}$ is $(\beta\delta/S_t)(\widetilde{x}_t^c)'Q$. Right

multiplication by D_n^{-1} multiplies this block by $\sqrt{n}$, so it becomes $(\beta\delta\sqrt{n}/S_t)(\widetilde{x}_t^c)'Q$. Since $(\widetilde{x}_t^c)'Q = \sqrt{n}\, v_{n,t}'$ and $S_t = n s_{n,t}$, this equals $(\beta\delta/s_{n,t})v_{n,t}'$. Left multiplication by D_n multiplies the lower-left block $\gamma Q'\widetilde{x}_t^c$ by $n^{-1/2}$, giving $\gamma v_{n,t}$. The lower-right block remains δI_{n-1}. Hence

$$\widehat{B}_{n,t} = \begin{pmatrix} k & (\beta\delta/s_{n,t})v_{n,t}' \\ \gamma v_{n,t} & \delta I_{n-1} \end{pmatrix}. \tag{A.6}$$

Fix a positive integer block length m. Let $V_{n,m}$ be the $(n-1)\times m$ matrix with columns $v_{n,1},\dots,v_{n,m}$, and let $G_{n,m} = V_{n,m}'V_{n,m}$. Its (s,t) entry is

$$(G_{n,m})_{st} = v_{n,s}'v_{n,t} = \frac{1}{n}(\widetilde{x}_s^c)'QQ'\widetilde{x}_t^c = \frac{1}{n}(\widetilde{x}_s^c)'\widetilde{x}_t^c,$$

because every centered feature vector belongs to $\mathbb{1}^\perp$. For fixed s and t, the cross-sectional law of large numbers gives this sample covariance converging in probability to the population covariance of the stationary feature process. At equal dates, that covariance is V. At distinct dates it is $\mathrm{Cov}(\widetilde{x}_{i,s},\widetilde{x}_{i,t}) = \sigma_\theta^2\phi^{|s-t|} = \rho V\phi^{|s-t|}$. Since m is fixed, all m^2 entries converge jointly. Thus $G_{n,m}\to_p G_m$, where

$$(G_m)_{st} = \begin{cases} V, & s = t, \\ \rho V\phi^{|s-t|}, & s \neq t. \end{cases} \tag{A.7}$$

The matrix G_m is positive definite because it is the covariance matrix of a nondegenerate Gaussian vector consisting of m consecutive feature observations. For fixed m and $n-1\geq m$, the Gaussian vectors $v_{n,1},\dots,v_{n,m}$ are in general position and are linearly independent almost surely, so $G_{n,m}$ is positive definite almost surely. Define $J_{n,m} = V_{n,m}G_{n,m}^{-1/2}$. Then $J_{n,m}'J_{n,m} = G_{n,m}^{-1/2}V_{n,m}'V_{n,m}G_{n,m}^{-1/2} = I_m$, so the columns of $J_{n,m}$ form an orthonormal basis of $\mathcal{H}_{n,m} = \mathrm{span}\{v_{n,1},\dots,v_{n,m}\}$. Also $V_{n,m} = J_{n,m}G_{n,m}^{1/2}$, so $v_{n,t} = J_{n,m}G_{n,m}^{1/2}e_t$.

Consider the restriction of $\widehat{B}_{n,t}$ to the score coordinate plus $\mathcal{H}_{n,m}$. If the capital coordinate is $z = J_{n,m}a$, then $v_{n,t}'z = e_t'G_{n,m}^{1/2}a$, while the coordinates of $v_{n,t}$ in the $J_{n,m}$ basis are $G_{n,m}^{1/2}e_t$. Since $s_{n,t} = v_{n,t}'v_{n,t} = (G_{n,m})_{tt}$, the restricted matrix in the basis consisting of the score coordinate and the columns of $J_{n,m}$ is exactly

$$C_t(G_{n,m}) = \begin{pmatrix} k & \dfrac{\beta\delta}{(G_{n,m})_{tt}}e_t'G_{n,m}^{1/2} \\ \gamma G_{n,m}^{1/2}e_t & \delta I_m \end{pmatrix}. \tag{A.8}$$

Every vector orthogonal to $\mathcal{H}_{n,m}$ is orthogonal to every $v_{n,t}$ in the block. Hence each $\widehat{B}_{n,t}$ multiplies such a capital vector by δ and does not couple it to the score coordinate. The product over the block therefore acts as $\delta^m I$ on the common orthogonal complement. When $n\geq m+2$, this complement has positive dimension. The state space is an orthogonal direct sum of the restricted score-plus-span subspace and that complement. Therefore $\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\| = \max\{\|C_m(G_{n,m})\cdots C_1(G_{n,m})\|,\delta^m\}$. The map from a positive definite Gram matrix G to $G^{1/2}$ is continuous, and matrix multiplication and operator norm are continuous. Since $G_{n,m}\to_p G_m$, the continuous mapping theorem gives

$$\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\| \xrightarrow{p} L_m, \qquad L_m = \max\{\|C_m(G_m)\cdots C_1(G_m)\|,\delta^m\}. \tag{A.9}$$

We next upgrade convergence in probability of the logarithm to convergence in expectation. First, bound the norm of one balanced matrix. The norm of a block matrix is at most the sum of the norms of its

four blocks, so

$$\|\widehat{B}_{n,t}\| \le |k| + \delta + \frac{|\beta|\delta}{s_{n,t}}\|v_{n,t}\| + |\gamma|\|v_{n,t}\| = |k| + \delta + \frac{|\beta|\delta}{\sqrt{s_{n,t}}} + |\gamma|\sqrt{s_{n,t}}. \tag{A.10}$$

Since $s_{n,t} = V\chi^2_{n-1}/n$, every logarithmic moment of $s_{n,t}$ is finite. For fixed m, we only need $n \ge m+2$, and the family $\{|\log s_{n,t}|^p : n \ge m+2\}$ has uniformly bounded expectation for every fixed $p > 1$. This follows from the gamma representation of the chi-square distribution and the standard logarithmic moments of gamma variables, together with convergence of $s_{n,t}$ to V.

From the one-period bound there is a constant C such that $\log^+ \|\widehat{B}_{n,t}\| \le C + C|\log s_{n,t}|$. For the fixed block, submultiplicativity gives $\log^+ \|\widehat{B}_{n,m} \cdots \widehat{B}_{n,1}\| \le \sum_{t=1}^m \log^+ \|\widehat{B}_{n,t}\|$. Hence, the positive parts of the fixed-block logarithms are uniformly bounded in L^p for some $p > 1$.

For the negative parts, the invariant orthogonal capital directions give $\|\widehat{B}_{n,m} \cdots \widehat{B}_{n,1}\| \ge \delta^m$. Since $\delta > 0$, $\log \|\widehat{B}_{n,m} \cdots \widehat{B}_{n,1}\| \ge m \log \delta$. Therefore, the negative part is bounded above by $m|\log \delta|$. Positive-part L^p boundedness and the deterministic negative-part bound imply uniform integrability. Convergence in probability from equation (A.9) therefore implies

$$\mathbb{E} \log \|\widehat{B}_{n,m} \cdots \widehat{B}_{n,1}\| \longrightarrow \log L_m. \tag{A.11}$$

Step 2. Compute the exponential growth rate of the deterministic fixed-block limit. Because every G_m is positive definite and consistent across m, we may realize vectors $X_1, X_2, \ldots$ in a Hilbert space $\mathcal{H}$ with inner products $\langle X_s, X_t\rangle = (G_m)_{st}$ whenever $s, t \le m$. Define deterministic operators on $\mathbb{R} \oplus \mathcal{H}$ by

$$\mathcal{C}_t \begin{pmatrix} \lambda \\ z \end{pmatrix} = \begin{pmatrix} k\lambda + \dfrac{\beta\delta}{V}\langle X_t, z\rangle \\ \gamma X_t \lambda + \delta z \end{pmatrix}. \tag{A.12}$$

Their restriction to the span of $X_1, \ldots, X_m$ has exactly the matrices $C_t(G_m)$ above, while every direction orthogonal to that span is multiplied by δ. Therefore $L_m = \|\mathcal{C}_m \cdots \mathcal{C}_1\|$.

We first establish the synthesis bound. For arbitrary scalars $a_1, \ldots, a_m$,

$$\left\|\sum_{t=1}^m a_t X_t\right\|^2 = \sum_{s=1}^m \sum_{t=1}^m a_s a_t \langle X_s, X_t\rangle.$$

Using the Toeplitz covariance structure and taking absolute values gives

$$\left\|\sum a_t X_t\right\|^2 \le V \sum_t a_t^2 + 2\rho V \sum_{h=1}^{m-1} \phi^h \sum_{t=1}^{m-h} |a_t a_{t+h}|.$$

For each h, the inequality $2|ab| \le a^2 + b^2$ gives $\sum_{t=1}^{m-h} |a_t a_{t+h}| \le \frac{1}{2}\sum_{t=1}^{m-h}(a_t^2 + a_{t+h}^2) \le \sum_{t=1}^m a_t^2$. Therefore

$$\left\|\sum_{t=1}^m a_t X_t\right\|^2 \le \Gamma \sum_{t=1}^m a_t^2, \qquad \Gamma = V + 2\rho V \frac{\phi}{1-\phi}. \tag{A.13}$$

The analysis operator is the adjoint of the synthesis operator and has the same norm. Hence for every $z \in \mathcal{H}$,

$$\sum_{t=1}^m |\langle X_t, z\rangle|^2 \le \Gamma \|z\|^2. \tag{A.14}$$

Now trace the deterministic recursion. If the state before the first operator is (λ_1, z_1), the capital

coordinate obeys $z_{t+1} = \gamma X_t \lambda_t + \delta z_t$. Iterating gives

$$z_t = \delta^{t-1} z_1 + \gamma \sum_{j=1}^{t-1} \delta^{t-1-j} X_j \lambda_j. \tag{A.15}$$

The score coordinate obeys $\lambda_{t+1} = k\lambda_t + (\beta\delta/V)\langle X_t, z_t\rangle$. Substitute the expansion for z_t:

$$\lambda_{t+1} = k\lambda_t + \frac{\beta}{V}\delta^t \langle X_t, z_1\rangle + \frac{\beta\gamma\delta}{V} \sum_{j=1}^{t-1} \delta^{t-1-j} \langle X_t, X_j\rangle \lambda_j.$$

For $j < t$, $\langle X_t, X_j\rangle = \rho V \phi^{t-j}$. Since $\beta\gamma = k$, the coefficient in the sum is

$$\frac{\beta\gamma\delta}{V}\delta^{t-1-j}\rho V \phi^{t-j} = k\rho\delta^{t-j}\phi^{t-j} = k\rho d^{t-j},$$

where $d = \delta\phi$. Define $q_{t-1} = \sum_{j=1}^{t-1} d^{t-1-j}\lambda_j$ and $b_t = (\beta/V)\delta^t\langle X_t, z_1\rangle$. Then $\sum_{j=1}^{t-1} d^{t-j}\lambda_j = dq_{t-1}$, so $\lambda_{t+1} = k\lambda_t + k\rho d q_{t-1} + b_t$. Also

$$q_t = \sum_{j=1}^{t} d^{t-j}\lambda_j = \lambda_t + d\sum_{j=1}^{t-1} d^{t-1-j}\lambda_j = \lambda_t + dq_{t-1}.$$

Combining the two equations gives

$$\begin{pmatrix}\lambda_{t+1}\\ q_t\end{pmatrix} = A\begin{pmatrix}\lambda_t\\ q_{t-1}\end{pmatrix} + \begin{pmatrix}b_t\\ 0\end{pmatrix}, \qquad A = \begin{pmatrix}k & k\rho d\\ 1 & d\end{pmatrix}. \tag{A.16}$$

The characteristic polynomial of A is $\mu^2 - (k+d)\mu + kd(1-\rho)$, so its dominant eigenvalue is the same μ_1 as in Proposition 10. Let $r = \max\{\mu_1, \delta\}$. Since this part of the proof treats $\delta > 0$, we have $r > 0$. The two-dimensional Jordan form implies that there is a finite constant C_A, independent of j, such that $\|A^j\| \le C_A(j+1)r^j$ for every integer $j \ge 0$. The factor $j+1$ accommodates a repeated eigenvalue. If $\mu_1 = 0$, both eigenvalues vanish and the Cayley–Hamilton theorem gives $A^2 = 0$. In that case, the same bound holds by choosing C_A large enough to cover $j = 0$ and $j = 1$.

Let $Y_t = (\lambda_t, q_{t-1})'$. Iterating the forced recursion gives

$$Y_t = A^{t-1}Y_1 + \sum_{s=1}^{t-1} A^{t-1-s}(b_s, 0)'.$$

Since $q_0 = 0$, $\|Y_1\| = |\lambda_1|$. Let $a_s = |\langle X_s, z_1\rangle|$. Then $|b_s| \le (|\beta|/V)\delta^s a_s$. Applying the power bound gives

$$\|Y_t\| \le C_A t r^{t-1}|\lambda_1| + C\sum_{s=1}^{t-1}(t-s)r^{t-1-s}\delta^s a_s.$$

Because $\delta \le r$, $r^{t-1-s}\delta^s \le r^{t-1}$. Also $t - s \le t$. Therefore the sum is at most $Ctr^{t-1}\sum_{s=1}^{t-1} a_s$. Cauchy–Schwarz and equation (A.14) give $\sum_{s=1}^{t-1} a_s \le \sqrt{t}\,[\sum_{s=1}^{t-1} a_s^2]^{1/2} \le \sqrt{t\Gamma}\|z_1\|$. Hence

$$|\lambda_t| \le \|Y_t\| \le C_1(t+1)^{3/2}r^{t-1}(|\lambda_1| + \|z_1\|). \tag{A.17}$$

Return to the capital expansion at date $m+1$: $z_{m+1} = \delta^m z_1 + \gamma\sum_{j=1}^{m}\delta^{m-j}X_j\lambda_j$. The synthesis bound

gives

$$\left\|\sum_{j=1}^{m}\delta^{m-j}X_j\lambda_j\right\| \le \sqrt{\Gamma}\left[\sum_{j=1}^{m}\delta^{2(m-j)}\lambda_j^2\right]^{1/2}.$$

Using the preceding slope bound, there is a constant C such that $\lambda_j^2 \le C(j+1)^3 r^{2j-2}R_0^2$, where $R_0 = |\lambda_1| + \|z_1\|$. Since $\delta \le r$, $\delta^{2(m-j)}r^{2j-2} \le r^{2m-2}$. Therefore

$$\sum_{j=1}^{m}\delta^{2(m-j)}\lambda_j^2 \le Cr^{2m-2}R_0^2\sum_{j=1}^{m}(j+1)^3 \le C'm^4r^{2m-2}R_0^2.$$

Taking square roots gives a bound of order $m^2r^{m-1}R_0$. Since $\delta^m \le r^m$, the initial-capital term is bounded by the same exponential rate. Absorbing the fixed factor $1/r$ into the constant gives

$$\|z_{m+1}\| \le C_2(m+1)^2r^m(|\lambda_1| + \|z_1\|). \tag{A.18}$$

Together with the score bound, this yields $L_m \le C_3(m+1)^2r^m$.

We now obtain a matching lower exponential rate. For every m, choose a unit capital vector z_1 orthogonal to $\operatorname{span}\{X_1,\dots,X_m\}$ and set $\lambda_1 = 0$. Such a vector exists because the positive definiteness of every G_m implies that $\mathcal{H}$ is infinite dimensional. Along this trajectory, the score coordinate remains zero, and the capital coordinate is multiplied by δ in every period. Therefore, $\|(\lambda_{m+1}, z_{m+1})\| = \delta^m$, which gives $L_m \ge \delta^m$.

Suppose next that $\mu_1 > \delta$. Choose the unit initial state $\lambda_1 = 1$ and $z_1 = 0$. Every forcing term b_t then vanishes, so $Y_t = A^{t-1}e_1$. We establish a lower bound for the original score coordinate λ_{m+1}.

Let μ_2 denote the smaller eigenvalue of A. The eigenvalues are $\mu_{1,2} = [k+d \pm \sqrt{(k-d)^2 + 4k\rho d}]/2$. Since $k, d \ge 0$ and $0 \le \rho \le 1$, they satisfy $0 \le \mu_2 \le \min\{k,d\}$ and $\mu_1 \ge \max\{k,d\}$. Moreover, $d = \delta\phi < \delta < \mu_1$, so $\mu_2 < \mu_1$. Using $\mu_1 + \mu_2 = k + d$ gives $k - \mu_2 = \mu_1 - d > 0$.

The Cayley–Hamilton identity is $A^2 - (k+d)A + kd(1-\rho)I = 0$. Applying it to Y_t and taking the first coordinate yields $\lambda_{t+2} = (k+d)\lambda_{t+1} - kd(1-\rho)\lambda_t$. The initial conditions are $\lambda_1 = 1$ and $\lambda_2 = k$. Since the characteristic roots μ_1 and μ_2 are distinct, the solution is $\lambda_{m+1} = c_1\mu_1^m + c_2\mu_2^m$, where $c_1 + c_2 = 1$ and $c_1\mu_1 + c_2\mu_2 = k$. Solving these equations gives $c_1 = (k-\mu_2)/(\mu_1-\mu_2) > 0$ and $c_2 = (\mu_1 - k)/(\mu_1 - \mu_2) \ge 0$. Consequently, $\lambda_{m+1} \ge c_1\mu_1^m$ for every $m \ge 1$.

The initial state $(1,0)$ has unit norm in $\mathbb{R} \oplus \mathcal{H}$. By the definition of the operator norm, $L_m \ge \|(\lambda_{m+1}, z_{m+1})\| \ge |\lambda_{m+1}| \ge c_1\mu_1^m$. This lower bound concerns the original state and does not require a comparison with the auxiliary coordinate q_m.

If $r = \delta \ge \mu_1$, the bound $L_m \ge \delta^m = r^m$ applies. If $r = \mu_1 > \delta$, the slope bound gives $L_m \ge c_1 r^m$. Thus, in either case, there is a constant $c > 0$, independent of m, such that $L_m \ge cr^m$. Combining this inequality with the upper bound gives $\log r + m^{-1}\log c \le m^{-1}\log L_m \le \log r + m^{-1}[\log C + 2\log(m+1)]$. Both correction terms tend to zero, so

$$\lim_{m\to\infty}\frac{1}{m}\log L_m = \log r = \log\max\{\mu_1, \delta\}. \tag{A.19}$$

Step 3. Use subadditivity to bound the finite-panel Lyapunov exponents. For each fixed n, Kingman's theorem gives both the almost-sure Lyapunov limit and the expectation formula

$$\gamma_{L,n} = \inf_{m\ge 1}\frac{1}{m}\mathbb{E}\log\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\|.$$

Therefore, for every fixed m, $\gamma_{L,n} \le m^{-1}\mathbb{E}\log\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\|$. Taking $\limsup_{n\to\infty}$ on both sides and using

the fixed-block expectation limit gives

$$\limsup_{n\to\infty} \gamma_{L,n} \le \frac{1}{m} \log L_m.$$

This inequality holds for every fixed m. Let $m \to \infty$. Equation (A.19) gives $m^{-1} \log L_m \to \log r = \log \max\{\mu_1, \delta\}$. Hence

$$\limsup_{n\to\infty} \gamma_{L,n} \le \log \max\{\mu_1, \delta\},$$

which is equation (104).

Now suppose $\kappa < 1$. Proposition 10 gives $\mu_1 < 1$, and the primitive assumption gives $\delta < 1$. Therefore $r = \max\{\mu_1, \delta\} < 1$ and $\log r < 0$. Let $\epsilon = -\frac{1}{2} \log r > 0$. By the definition of the limit superior, there exists N such that for every $n \ge N$, $\gamma_{L,n} < \log r + \epsilon = \frac{1}{2} \log r < 0$. Thus all sufficiently large panels have negative Lyapunov exponent, and Theorem 8 gives the unique causal stationary and ergodic solution.

Finally, consider $\delta = 0$. There is no inherited capital in the outcome. The score recursion reduces to the static slope channel with population multiplier k, and $\kappa = k$. The large-panel matrix has eigenvalues k and zero, so $\mu_1 = k$. The reduced homogeneous product has an exponential rate at most $\log k$ when $k > 0$, with the degenerate $k = 0$ case having exponent $-\infty$. Thus equation (104) holds with $\log \max\{\mu_1, 0\} = \log k$. If $\kappa = k < 1$, the exponent is negative, completing the boundary case. □

Proof of Proposition 11. First consider the boundary case $\delta = 0$. Then $\kappa = k$, so $\kappa > 1$ implies $k > 1$. Equation (97) reduces to

$$\widetilde{B}_t = \begin{pmatrix} k & 0 \\ \gamma Q' \widetilde{x}_t^c & 0 \end{pmatrix}.$$

A direct multiplication gives

$$\widetilde{B}_m \cdots \widetilde{B}_1 = \begin{pmatrix} k^m & 0 \\ \gamma k^{m-1} Q' \widetilde{x}_m^c & 0 \end{pmatrix}.$$

Applying this product to the score basis vector e_1 shows that its operator norm in any norm that leaves the first coordinate unscaled is at least k^m. In particular, $\|\widetilde{B}_m \cdots \widetilde{B}_1\|_n \ge k^m$. Choose any ϵ with $0 < \epsilon < \log k$. Then $m^{-1} \mathbb{E} \log \|\widetilde{B}_m \cdots \widetilde{B}_1\|_n \ge \log k > \epsilon$ for every m and n, and the probability in equation (106) equals one. Thus the proposition holds when $\delta = 0$.

Now suppose $\delta > 0$. Use the balanced matrices introduced in the proof of Theorem 9: $\widehat{B}_{n,t} = D_n \widetilde{B}_{n,t} D_n^{-1}$ with $D_n = \mathrm{diag}(1, n^{-1/2} I_{n-1})$. By definition of the norm in the proposition,

$$\left\| \widetilde{B}_m \cdots \widetilde{B}_1 \right\|_n = \left\| D_n \widetilde{B}_m \cdots \widetilde{B}_1 D_n^{-1} \right\|_2 = \left\| \widehat{B}_{n,m} \cdots \widehat{B}_{n,1} \right\|_2.$$

Thus the fixed-block convergence established in equations (A.9) and (A.11) applies directly to the norm used here.

Suppose $\kappa > 1$. Proposition 10 implies that the dominant population eigenvalue satisfies $\mu_1(\kappa) > 1$. Since $\delta < 1$, we have $r = \max\{\mu_1, \delta\} = \mu_1 > 1$. Equation (A.19) gives

$$\lim_{m\to\infty} \frac{1}{m} \log L_m = \log \mu_1 > 0.$$

Choose $\epsilon > 0$ such that $4\epsilon < \log \mu_1$. By convergence of $m^{-1} \log L_m$ to $\log \mu_1$, there exists a finite block length m such that

$$\frac{1}{m} \log L_m > 3\epsilon.$$

Equivalently, $L_m > e^{3m\epsilon}$.

For this fixed m, equation (A.11) gives

$$\mathbb{E}\log\left\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\right\| \longrightarrow \log L_m.$$

Since $\log L_m > 3m\epsilon$, there exists N such that for every $n \geq N$,

$$\mathbb{E}\log\left\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\right\| > 2m\epsilon.$$

Divide by m and use the equality between the balanced Euclidean norm and $\|\cdot\|_n$ to obtain

$$\frac{1}{m}\mathbb{E}\log\left\|\widetilde{B}_m\cdots\widetilde{B}_1\right\|_n > 2\epsilon > \epsilon,$$

which proves equation (105).

For the probability statement, equation (A.9) gives

$$\left\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\right\| \xrightarrow{p} L_m.$$

Because $L_m > e^{3m\epsilon} > e^{m\epsilon}$, define the positive gap $c = L_m - e^{m\epsilon} > 0$. Then

$$\Pr\left(\left\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\right\| \leq e^{m\epsilon}\right) \leq \Pr\left(\left|\left\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\right\| - L_m\right| \geq c\right).$$

The right-hand side tends to zero by convergence in probability. Therefore

$$\Pr\left(\left\|\widehat{B}_{n,m}\cdots\widehat{B}_{n,1}\right\| > e^{m\epsilon}\right) \to 1.$$

Replacing the balanced Euclidean norm by the equivalent notation $\|\cdot\|_n$ gives equation (106).

The proof stops at a fixed block because Kingman's formula is subadditive:

$$\gamma_{L,n} = \inf_{\ell\geq 1}\frac{1}{\ell}\mathbb{E}\log\left\|\widetilde{B}_\ell\cdots\widetilde{B}_1\right\|_n.$$

A positive value at one block length gives no lower bound on this infimum. Completing the positive-side bridge therefore requires an additional argument that controls loss of alignment across successive expanding blocks. □

Appendix B Numerical Details

Calibration and initialization. The dynamic simulations set $\sigma_\theta^2 = \sigma_\eta^2 = \sigma_\varepsilon^2 = 1$, so $V = 2$ and $\rho = 0.5$. We normalize $\beta = 1$ and set $\gamma = k$, keeping the contemporaneous gain $k = \beta\gamma$. The baseline has $\delta = 0.8$ and $\phi = 0.875$, which imply $\delta\phi = 0.7$, $\psi = 13/6$, and $\kappa = (13/6)k$. The population frontier is therefore $k_c = 6/13 \approx 0.46154$.

Each fundamental starts independently in its stationary law, $\widetilde{\theta}_{i,0} \sim N(0,1)$. Subsequent innovations satisfy $w_{i,t} \sim N(0, 1-\phi^2)$, $\eta_{i,t} \sim N(0,1)$, and $\varepsilon_{i,t} \sim N(0,1)$, independently across units and dates and across shock types. Economic paths start with zero capital. The long-run comparisons start from $\lambda_0 = 0.5$ and discard 10,000 updates. The frontier and threshold-crossing experiments have no economic burn-in and use the initial slopes stated below.

The code implements the exact centered-capital recurrence. Write x_t^c for the demeaned feature vector and $S_t = (x_t^c)'x_t^c$. The coefficient update is $\lambda_{t+1} = k\lambda_t + \beta\delta(x_t^c)'K_{t-1}^c/S_t + g_t$, and capital updates as $K_t^c = \delta K_{t-1}^c + (k/\beta)\lambda_t x_t^c$. The forcing term is generated as $g_t = (x_t^c)'\widetilde{\theta}_t/S_t + \sigma_\varepsilon Z_t/\sqrt{S_t}$, where Z_t is an independent standard normal. This last step integrates out the outcome-noise vector and has exactly the conditional distribution of its OLS projection. It retains the dependence between the fundamental component of g_t, current features, and past capital.

Simulation lengths and uncertainty. For each dynamic calibration and panel size, four independent runs retain 200,000 updates each. Lyapunov calculations set the additive forcing to zero and iterate a normalized vector in the admissible centered-capital space. Normalization uses $\|(\lambda, K^c)\|_n = [\lambda^2 + \|K^c\|^2/n]^{1/2}$. For fixed n, this norm has the same asymptotic exponent as the Euclidean norm. Each run discards 5,000 direction updates before accumulating log gains. Figure B.2 depicts the excluded common-capital direction.

Reported Monte Carlo standard errors use 40 nonoverlapping batches of length 5,000 per run, giving 160 batches across the four runs. The standard error is the sample standard deviation of the batch averages divided by $\sqrt{160}$. Approximate 95% Monte Carlo intervals use the corresponding Student critical value. Doubling the batch length changes the coupled-mode Lyapunov standard errors by at most 13% and the slope-mean standard errors by at most 9% over the reported panel sizes. These calculations measure simulation variability. They do not bound burn-in bias, establish the existence of stationary moments, or resolve the conjectured large-panel limit.

The static check uses $n = 12$, $k = 0.6$, $\rho = 0.5$, $V = 1$, and $\sigma_\varepsilon^2 = 0.25$, so $\sigma_u^2/V = 0.5$. It starts at $\lambda_0 = 1.25$, discards 2,000 updates, and retains 200,000. Its standard errors use 40 batches of length 5,000. The fitted dynamic ACF decay rates are averaged across the four runs, with standard errors computed from those four estimates.

Population benchmarks and simulated averages. Table B.1 separates exact static moment formulas from dynamic population benchmarks. At $n = 200$ and $k = 0.30$, the simulated time-average slope is 1.4325 with Monte Carlo standard error 0.00061, compared with the population fixed point 1.4286. At $k = 0.44$, it is 11.1014 with standard error 0.04160, compared with 10.7143. These differences exceed the simulation errors. The probability limit of the coefficient matrix need not equal its expectation, and its coefficients can be correlated with the inherited state. The population fixed point therefore need not equal a finite-panel stationary mean.

Across $n \in \{25, 50, 100, 200, 400, 800, 1000\}$, the time-average slopes move toward their population benchmarks. At $k = 0.30$, the gap falls from 0.0405 at $n = 25$ to 0.0005 at $n = 1000$. At $k = 0.44$, it falls from 4.9419 to 0.0608. These results support a finite-panel interpretation of the discrepancy. They do not establish a $1/n$ expansion or its coefficient.

At $k = 0.30$, the fitted ACF decay over lags 10 through 24 is 0.8878, compared with the dominant

Table B.1: Simulated quantities and population benchmarks. The static benchmarks are exact moment formulas. Dynamic time averages are compared with population fixed points, fitted ACF rates with μ_1, and Lyapunov estimates with the conjectured limit $\log \max\{\mu_1, \delta\}$. The dynamic benchmarks are not finite-panel moment formulas. Monte Carlo standard errors are computed as described in the text.

Quantity	Estimate	MC s.e.	Benchmark
Static economy, $k = 0.60$, $n = 12$, $\rho = 0.5$, $\sigma_u^2/V = 0.5$			
$\mathbb{E}[U_t]$	0.00018	0.000545	0.00000
$\mathrm{Var}(U_t)$	0.05539	0.000213	0.05556
$\mathbb{E}[\lambda_\infty]$	1.25044	0.001359	1.25000
$\mathrm{Var}(\lambda_\infty)$	0.08616	0.000467	0.08681
$\mathrm{Corr}(\lambda_{t+1}, \lambda_t)$	0.59767	0.001602	0.60000
Excess kurtosis of U_t	0.89030	0.044359	0.85714
Dynamic panel, $n = 200$, $\delta = 0.8$, $\phi = 0.875$			
Time-average slope, $k = 0.30$	1.43254	0.000607	1.42857
Time-average slope, $k = 0.44$	11.10143	0.041596	10.71429
Fitted ACF decay, $k = 0.30$	0.88780	0.001807	0.88079
Fitted ACF decay, $k = 0.44$	0.98377	0.000489	0.98340
Lyapunov exponent, $k = 0.05$	-0.22295	0.000001	-0.22314
Lyapunov exponent, $k = 0.30$	-0.12849	0.000073	-0.12694
Lyapunov exponent, $k = 0.44$	-0.01784	0.000065	-0.01674
Lyapunov exponent, $k = 0.50$	0.02865	0.000063	0.02967

Table B.2: Estimated finite-panel zero crossings. The final gain grid has step 0.0005. Intervals and standard deviations come from 999 joint resamples of the log-gain batches. The population frontier is $k_c \approx 0.46154$.

Panel size n	Crossing	95% MC interval	Bootstrap s.d.
25	0.47628	[0.47580, 0.47682]	0.00026
100	0.46444	[0.46422, 0.46467]	0.00012
200	0.46292	[0.46276, 0.46307]	0.00008
1000	0.46179	[0.46171, 0.46187]	0.00004

population eigenvalue $\mu_1 \approx 0.8808$. At $k = 0.44$, the fit over lags 20 through 39 is 0.9838, compared with $\mu_1 \approx 0.9834$. These are descriptive comparisons. In particular, primitive persistence $\phi = 0.875$ is close to μ_1 in the first calibration, so its fitted ACF does not identify which source of persistence governs the long-lag decay.

Lyapunov estimates and zero crossings. At $n = 200$, the weak-loop calibration $k = 0.05$ gives -0.22295, compared with $\log \delta \approx -0.22314$. The coupled-mode estimates at $k = 0.30$, 0.44, and 0.50 are -0.12849, -0.01784, and 0.02865. Their respective population benchmarks are -0.12694, -0.01674, and 0.02967. The estimated gaps exceed their Monte Carlo standard errors and shrink with panel size in Figure B.1. At $n = 1000$, the four estimates are -0.22310, -0.12722, -0.01694, 0.02948. Each lies within 3×10^{-4} of its benchmark.

For the zero-crossing calculation, a pilot grid with step 0.005 locates a neighborhood of the crossing. The final grid has step 0.0005 and uses the four longer runs at every grid point. Within a run, all gains use the same primitive feature path. We linearly interpolate between the adjacent estimates that bracket zero. Approximate 95% Monte Carlo intervals come from 999 resamples of the batch vectors, retaining the dependence across gains. Table B.2 reports the results. Using every second point to give a 0.001 grid changes each estimated crossing by less than 6×10^{-7}. This is a grid-refinement check, not a bound on all simulation bias. The estimated crossings approach $k_c = 6/13$ from above over these panel sizes.

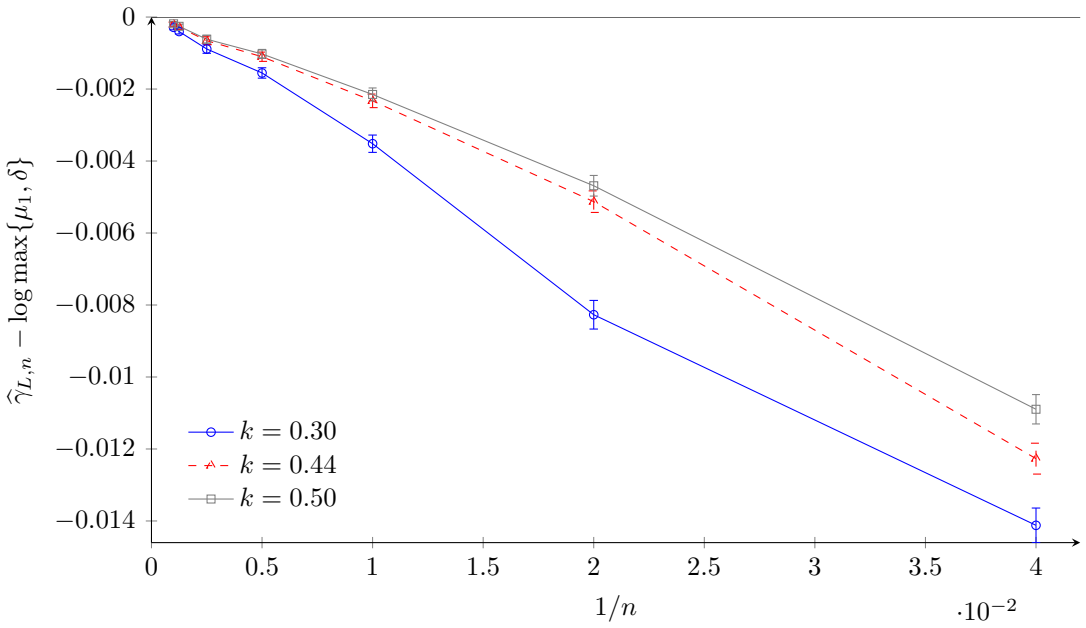


Figure B.1: Finite-panel Lyapunov estimates relative to the conjectured large-panel limit. The baseline calibration has $\rho = 0.5$, $\delta = 0.8$, and $\phi = 0.875$. Each point averages four runs of 200,000 retained log gains after direction burn-in. Error bars show approximate 95% batch-based Monte Carlo intervals. The gaps decrease as n rises from 25 to 1000. The figure supplies numerical evidence for convergence, without establishing its rate or a matching analytical lower bound.

Table B.3: Durability variations with $\delta\phi = 0.7$ and $n = 200$. The benchmark is $\log\max\{\mu_1, \delta\}$. Each estimate uses four runs of 200,000 retained log gains.

δ	ϕ	k	Estimate	MC s.e.	Benchmark
0.72	0.97222	0.05	-0.32118	0.000063	-0.32035
0.72	0.97222	0.30	-0.12725	0.000067	-0.12694
0.80	0.87500	0.05	-0.22295	0.000001	-0.22314
0.80	0.87500	0.30	-0.12849	0.000073	-0.12694
0.90	0.77778	0.05	-0.10523	0.000001	-0.10536
0.90	0.77778	0.30	-0.09969	0.000037	-0.10536

Durability and the capital mode. To vary durability while holding $d = \delta\phi = 0.7$ fixed, we set $\phi = 0.7/\delta$. The population eigenvalue μ_1 then remains unchanged at a given k, while the capital decay factor δ varies. Table B.3 reports the resulting estimates at $n = 200$. At $k = 0.05$ and $\delta = 0.72$, the coupled mode governs the benchmark because $\mu_1 \approx 0.7259 > \delta$. At $\delta = 0.8$ and 0.9, the capital mode governs. At $k = 0.30$ and $\delta = 0.9$, the finite-panel estimate lies visibly above $\log\delta$. This discrepancy shows why a finite-panel estimate near a change in the dominant mode should not be treated as the large-panel limit.

The population frontier. At $k = k_c$ and $n = 200$, the average running Lyapunov estimates after 5,000, 20,000, 80,000, and 200,000 retained updates are -0.00146, -0.00134, -0.00099, -0.00107. Their standard errors across the four independent runs are 0.00022, 0.00024, 0.00015, 0.00006. The slightly negative estimates agree with the estimated finite-panel zero crossing lying above k_c. They do not establish the behavior of the exponent at the population frontier as $n \to \infty$.

For economic paths at k_c, Table B.4 reports the median and upper quantiles of $|\lambda_t|$ from 500 independent paths at each panel size. These paths start from $\lambda_0 = 0$ and zero capital, with stationary fundamentals. The median intervals use 999 independent-path bootstrap resamples. The replication files also report intervals

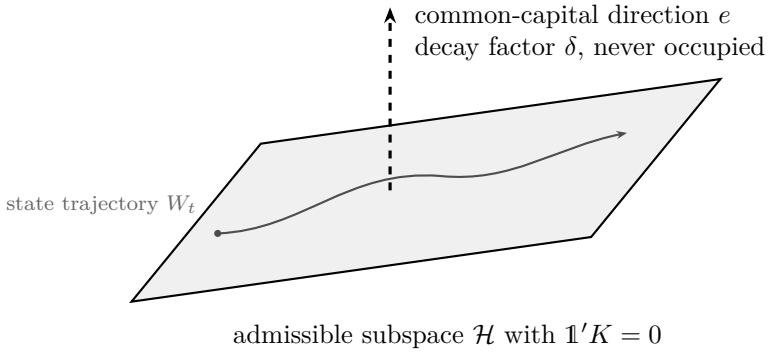


Figure B.2: The admissible state space of the dynamic recurrence. Demeaned capital lies in the hyperplane $\mathcal{H}$, which the coefficient matrices preserve. The common-capital direction has eigenvalue δ but is never occupied by the economic state or the forcing term. The simulations keep capital centered and therefore exclude this direction from the Lyapunov calculation.

Table B.4: Slope magnitudes at the population frontier $k = 6/13$. Each panel size uses 500 independent paths without economic burn-in. Quantiles refer to $|\lambda_t|$, and the median interval is a 95% independent-path bootstrap interval. Dates count completed coefficient updates.

n	Updates	Median	Median interval	90th pct.	95th pct.
50	1,000	66.1	[56.7, 81.0]	848.4	2,301.5
50	5,000	75.1	[65.0, 82.2]	1,491.0	5,405.2
50	20,000	65.9	[57.2, 77.1]	1,273.6	3,491.6
200	1,000	120.5	[108.4, 130.2]	525.4	905.4
200	5,000	275.2	[243.0, 317.5]	2,769.9	5,659.0
200	20,000	317.3	[264.7, 381.5]	4,255.8	9,898.4

for the upper quantiles and results at intermediate dates.

The median at $n = 50$ is about 66 after 1,000 updates and remains near that level after 20,000. At $n = 200$, it rises from about 120 to 317 over the same dates. Upper quantiles are much larger and remain imprecisely estimated, especially late in the sample. These results document substantial finite-horizon dispersion. They do not support a common power-law growth rate for the slope distribution. Cross-path means and variances are retained in the data files for comparison, but a few large paths make them unstable summaries of this experiment.

Threshold-crossing diagnostics. Table B.5 reports first-passage results for 500 paths of length 5,000 at $n = 200$. At $k = 0.50$, all paths cross all three thresholds from either initial slope. The second initialization sets the slope to the repelling population value -6 but still starts capital at zero. It is therefore not an initialization of the entire state at its population fixed point. The spacing of the median passage times across thresholds is broadly consistent with geometric growth. The baseline exponent estimate implies about 80.4 updates per tenfold increase once transient effects have dissipated.

At $k = 0.44$, no path crosses 10^3 within the simulated horizon. A missing passage time is recorded as right-censored, rather than as evidence that the threshold can never be crossed. Likewise, universal crossing in the $k = 0.50$ sample is numerical evidence of expansion, not a proof of almost-sure divergence for the dynamic panel.

Tail diagnostics. The Hill calculation is retained as an exploratory diagnostic using one separate 200,000-update path after 10,000 burn-in updates. At $n = 200$, estimates based on the largest 0.5%, 1%, and

Table B.5: Threshold crossings within 5,000 updates at $n = 200$. Counts give the number of paths with $|\lambda_t|$ exceeding each threshold. Median times use paths that cross. All paths start with zero capital, and a dash denotes no observed crossing.

		Paths crossing			Median passage time		
k	λ_0	10^3	10^4	10^5	10^3	10^4	10^5
0.50	0	500/500	500/500	500/500	174.0	255.5	338.0
0.50	-6	500/500	500/500	500/500	190.0	270.0	352.0
0.44	0	0/500	0/500	0/500	–	–	–

2% of mean-centered absolute slopes are 8.72, 7.91, 7.17 at $k = 0.30$, and 5.25, 4.54, 3.76 at $k = 0.44$. The dependence on the chosen tail fraction is substantial. The code reports dispersion across 399 circular block-bootstrap resamples using blocks of 2,000 observations. These dispersions are not presented as validated confidence intervals for a dynamic tail index. Regular variation and the conditions needed for Hill inference have not been established for this panel recurrence.

Internet Appendix

Self-Fulfilling Accuracy

This Internet Appendix develops function-space archive identification, Gaussian and elliptical linear specializations, additional stateful results, and archive diagnostics. It also derives the stochastic-approximation representation of pooled retraining. The final section collects the proofs, grouped by result type and ordered numerically within each group.

IA.1 Function-space archive identification

The main text uses the exact scalar transition function as the operational nonlinear archive result. This section studies the corresponding local restriction for the entire prediction function. Let P_X denote the distribution of the feature and let $\mathcal{H} = L^2(P_X)$ be the space of square-integrable prediction functions, with functions equal P_X-almost surely treated as the same element. The population operator is $(Tm)(x) = u(x) + f(m(x))$. Identification of local feedback can be exact even when its estimation is sensitive to small covariance errors.

Theorem IA.1 (Function-space archive linearization)**.** *Suppose $u \in \mathcal{H}$ and f is differentiable. Assume $|f'(s)| \le \kappa_{\text{abs}} < 1$ for every $s \in \mathbb{R}$. Let $\{\xi_t\}_{t\in\mathbb{Z}}$ be i.i.d. $\mathcal{H}$-valued random elements with $\mathbb{E}[\xi_t] = 0$ and $\mathbb{E}\|\xi_t\|_{\mathcal{H}}^2 < \infty$. For each $\tau > 0$, consider*

$$M_{t+1}^{\tau} = TM_t^{\tau} + \tau\xi_{t+1}. \tag{IA.1}$$

Define the bounded multiplication operator $A\colon \mathcal{H} \to \mathcal{H}$ by $(Aq)(x) = a(x)q(x)$, where $a(x) = f'(m^*(x))$. Then (IA.1) has a unique causal stationary solution in $L^2(\Omega;\mathcal{H})$. Construct these stationary solutions using the same innovation sequence. With $Z_t^{\tau} = (M_t^{\tau} - m^*)/\tau$, we have $Z_t^{\tau} \to Z_t$ in $L^2(\Omega;\mathcal{H})$ as $\tau \downarrow 0$ for every fixed t, where

$$Z_{t+1} = AZ_t + \xi_{t+1}, \qquad Z_t = \sum_{j=0}^{\infty} A^j \xi_{t-j}. \tag{IA.2}$$

Define $C_0 h = \mathbb{E}[\langle Z_t, h\rangle_{\mathcal{H}} Z_t]$ and $C_1 h = \mathbb{E}[\langle Z_t, h\rangle_{\mathcal{H}} Z_{t+1}]$. Then

$$C_1 = AC_0. \tag{IA.3}$$

The lag-zero and lag-one covariance operators of Z_t^{τ} converge in operator norm to C_0 and C_1. The pair (C_0, C_1) determines the restriction of A to $\overline{\text{Ran}(C_0)}$. In particular, it identifies A on $\mathcal{H}$ if $\overline{\text{Ran}(C_0)} = \mathcal{H}$.

Equation (IA.3) is a population identification statement rather than a well-conditioned infinite-dimensional estimator. Under the maintained second-moment condition, the covariance operator is compact. If it is injective on an infinite-dimensional space, its inverse on its range is unbounded. The next proposition makes the operational content explicit on finite-dimensional archive directions.

Proposition IA.1 (Finite-dimensional archive identification and sensitivity)**.** *Let C_0 have orthonormal eigenfunctions $e_1, e_2, \ldots$ with eigenvalues $\nu_1 \ge \nu_2 \ge \cdots > 0$ on the directions considered. Fix $J < \infty$, let $\mathcal{H}_J = \text{span}\{e_1, \ldots, e_J\}$, and let P_J be the orthogonal projection onto $\mathcal{H}_J$. Regard $C_{0,J} = P_J C_0|_{\mathcal{H}_J}$ and $C_{1,J} = P_J C_1|_{\mathcal{H}_J}$ as operators from $\mathcal{H}_J$ to itself, and define $A_J = P_J A|_{\mathcal{H}_J}$. Then*

$$A_J = C_{1,J} C_{0,J}^{-1}. \tag{IA.4}$$

Suppose estimated operators satisfy $\|\widehat{C}_{0,J} - C_{0,J}\| \le \epsilon_0 < \nu_J$ and $\|\widehat{C}_{1,J} - C_{1,J}\| \le \epsilon_1$, and define $\widehat{A}_J = \widehat{C}_{1,J}\widehat{C}_{0,J}^{-1}$. Then

$$\|\widehat{A}_J - A_J\| \le \frac{\epsilon_1}{\nu_J - \epsilon_0} + \frac{\|C_{1,J}\|\epsilon_0}{\nu_J(\nu_J - \epsilon_0)}. \tag{IA.5}$$

The finite-dimensional result separates identification from statistical conditioning. Small covariance eigenvalues make recovery sensitive to estimation error even though the corresponding projected feedback operator is point-identified. Rich archive variation is valuable twice. It expands the subspace on which feedback can be recovered and it keeps the relevant covariance eigenvalues away from zero, reducing sensitivity to sampling error.

IA.2 Linear and elliptical specializations

The main text uses a scalar linear-Gaussian specialization only to obtain exact archive dynamics. This section records a multivariate restriction that can be useful when institutions retain coefficient vectors. Let $\widetilde{x} \in \mathbb{R}^d$ be centered with nonsingular covariance matrix Σ, let θ and ε be square integrable, and define $q \coloneqq \mathrm{Cov}(\widetilde{x}, \theta) \neq 0$ and $v \coloneqq \Sigma^{-1} q$. The modeler deploys the linear score $s = \lambda' \widetilde{x}$ and the outcome satisfies

$$y = \theta + f(\lambda' \widetilde{x}) + \varepsilon, \qquad \mathbb{E}[\varepsilon \widetilde{x}] = 0. \tag{IA.6}$$

Population least squares maps the deployed coefficient into

$$T\lambda = \Sigma^{-1} \mathrm{Cov}(\widetilde{x}, y^{(\lambda)}). \tag{IA.7}$$

A stable coefficient satisfies $\lambda^* = T\lambda^*$.

Under Gaussian features, scalar feedback has no direction in feature space other than the deployed index itself. This yields a no-rotation restriction on the stable population coefficient. For $\lambda \neq 0$, write $\overline{k}(\lambda) = \mathbb{E}[f'(\lambda' \widetilde{x})]$ for the average derivative along the deployed index.

Theorem IA.2 (No rotation under Gaussian features)**.** *Suppose $\widetilde{x}$ is centered Gaussian with nonsingular covariance matrix Σ, $q \neq 0$, $\mathbb{E}[\varepsilon \widetilde{x}] = 0$, and f is Lipschitz with $f'(s) \leq \kappa^+ < 1$ for Lebesgue-almost every s. Then there is a unique stable coefficient. It is nonzero and satisfies*

$$\lambda^* = \frac{v}{1 - \overline{k}(\lambda^*)}, \qquad \overline{k}(\lambda^*) \leq \kappa^+ < 1. \tag{IA.8}$$

If f is affine with slope k, then

$$\lambda^* = \frac{v}{1 - k}. \tag{IA.9}$$

For any nonzero deployed coefficient λ, Stein's identity gives

$$\mathrm{Cov}\big(\widetilde{x}, f(\lambda' \widetilde{x})\big) = \Sigma \lambda \, \mathbb{E}\big[f'(\lambda' \widetilde{x})\big] \,. \tag{IA.10}$$

At $\lambda = 0$, the feedback contribution is zero and $T0 = v \neq 0$, so zero cannot be stable. For $\lambda \neq 0$, multiplication by Σ^{-1} shows that the feedback contribution to the coefficient update is collinear with the deployed coefficient. At a fixed point, population least squares combines the no-feedback coefficient v with a scalar multiple of λ^*, forcing the stable coefficient to lie on the ray spanned by v.

The same geometry extends beyond Gaussian features when one-dimensional projections have linear conditional expectations.

Recall that $v = \Sigma^{-1} q$, where $q = \mathrm{Cov}(\widetilde{x}, \theta)$, is the population coefficient without feedback. For a nonzero deployed coefficient λ, define the effective feedback gain by

$$\overline{k}(\lambda) \coloneqq \frac{\mathrm{Cov}(\lambda' \widetilde{x}, f(\lambda' \widetilde{x}))}{\mathrm{Var}(\lambda' \widetilde{x})}. \tag{IA.11}$$

Proposition IA.2 (No rotation under elliptical features)**.** *Suppose $\widetilde{x}$ is elliptically distributed with mean zero, nonsingular covariance matrix Σ, and finite second moments. Suppose also that $q \neq 0$, $\mathbb{E}[\varepsilon \widetilde{x}] = 0$,*

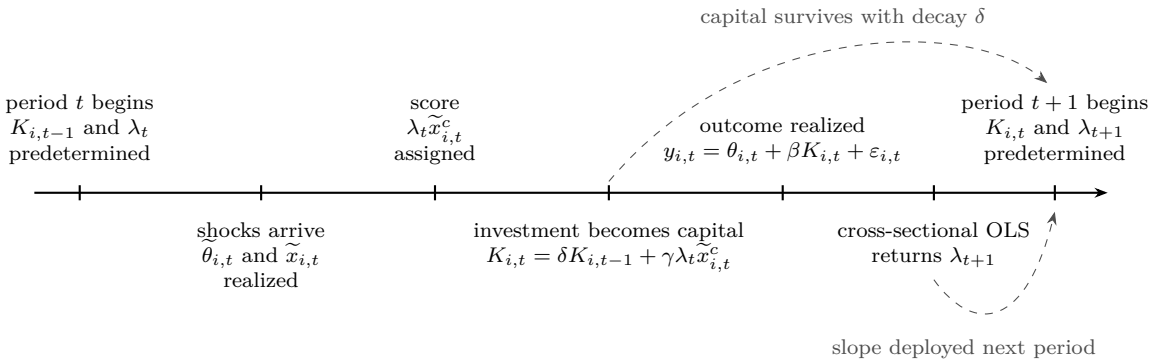


Figure IA.1: The within-period timing of dynamic retraining. Inherited capital and the previously estimated slope are predetermined when the period-t shocks arrive, the assigned score converts current features into new capital, and the realized outcome feeds the cross-sectional regression that produces the next slope. The dashed arrows mark the two channels through which the current deployment enters the next period: durable capital carried forward with decay δ and the reestimated coefficient carried forward by deployment.

and f is Lipschitz with $f'(t) \le \kappa^+ < 1$ almost everywhere. Then there is a unique stable coefficient, and it satisfies

$$\lambda^* = \frac{v}{1 - \overline{k}(\lambda^*)}, \qquad \overline{k}(\lambda^*) \le \kappa^+ < 1. \tag{IA.12}$$

Under Gaussian features, Stein's identity gives $\overline{k}(\lambda) = \mathbb{E}[f'(\lambda' \widetilde{x})]$, recovering the gain formula in Theorem IA.2. Under either distributional specification, the stable coefficient is a positive scalar multiple of the coefficient without feedback. Scalar feedback therefore changes the magnitude of the stable linear predictor while preserving its direction and the ranking generated by $v'\widetilde{x}$.

This restriction applies at the population fixed point. During retraining, the coefficient vector can change direction even when feedback is scalar. For example, under affine feedback, the update is $T\lambda = v + k\lambda$. If the deployed coefficient is not collinear with v, this update can rotate it toward or away from the no-feedback direction. Rotation across model vintages alone therefore does not reject the benchmark.

The empirical implication concerns a stable population coefficient evaluated against a no-feedback benchmark under the same fundamental and feature environment. If that coefficient is not a positive scalar multiple of v, the joint restrictions of the scalar-feedback model and the Gaussian or elliptical feature distribution are rejected. Such a violation may reflect feature-specific outcome responses or a failure of the distributional restriction. Changes in the information environment and sampling variation must also be accounted for when applying this restriction to estimated coefficients.

IA.3 Additional stateful results

The main text keeps only the stateful objects needed for the stability argument. This section records the within-period timing, one-period eigenstructure, the implication of a positive Lyapunov exponent, and the block-alignment condition that isolates the remaining positive-side difficulty. Throughout, retain the stationary Gaussian fundamental process and the nonanticipative initialization of the stateful model. Current primitive innovations are independent of the history that determines the initial slope and inherited capital. The reduced coefficient sequence is stationary and geometrically β-mixing, and the logarithmic moment conditions in (102) hold.

Lemma IA.1 (Spectral structure of the coefficient matrix)**.** *For every realization of the feature vector with*

$S_t > 0$, the matrix $\widetilde{B}_t$ has characteristic polynomial

$$\det(zI - \widetilde{B}_t) = z\,(z - k - \delta)\,(z - \delta)^{n-2}. \tag{IA.13}$$

Its eigenvalues, counted with algebraic multiplicity, are 0, $k + \delta$, and $n - 2$ copies of δ. These values may coincide.

The spectrum separates a feature-aligned score-capital mode from $n - 2$ capital directions that decay at the survival rate δ. Because the feature-aligned subspace rotates over time, one-period eigenvalues do not determine long-run stability. The state-space discussion below explains why the Lyapunov calculation uses the centered-capital subspace.

Proposition IA.3 (Expansion under a positive Lyapunov exponent)**.** *Suppose $\gamma_{L,n} > 0$. Then the reduced coefficient products satisfy*

$$\lim_{t\to\infty} \frac{1}{t} \log \left\| \widetilde{B}_t \cdots \widetilde{B}_1 \right\| = \gamma_{L,n} > 0 \quad \textit{almost surely.} \tag{IA.14}$$

The proposition establishes an exponentially expanding propagation mechanism rather than divergence of the state itself. Divergence of the affine state additionally requires that the forcing term repeatedly excites the expanding directions, and a primitive nondegeneracy condition guaranteeing this excitation remains open. The simulations in Appendix Appendix B show rapid state expansion above the population frontier over the reported horizons. They do not establish the required excitation condition or almost-sure divergence. Together, Theorem 8 and Proposition IA.3 replace the scalar contraction condition of the static economy with the sign of the growth rate of random matrix products. They fall short of a complete finite-panel phase transition, because the positive side concerns coefficient products rather than the state and because the critical case is untreated.

Lemma IA.2 (Block-alignment sufficient condition)**.** *Fix a panel size n and a block length m. Let $P_j = \widetilde{B}_{jm} \cdots \widetilde{B}_{(j-1)m+1}$ denote consecutive length-m coefficient products. Suppose the block sequence is stationary and ergodic and $\mathbb{E}|\log \|P_1\|_n| < \infty$. Assume there is a constant $c \in (0, 1]$ such that, almost surely, for every $r \geq 2$,*

$$\|P_r P_{r-1} \cdots P_1\|_n \geq c^{r-1} \prod_{j=1}^{r} \|P_j\|_n. \tag{IA.15}$$

Then

$$\gamma_{L,n} \geq \frac{1}{m} \mathbb{E} \log \|P_1\|_n + \frac{1}{m} \log c. \tag{IA.16}$$

In particular, if $m^{-1}\mathbb{E} \log \|P_1\|_n > \epsilon > 0$ and $c > e^{-m\epsilon}$, then $\gamma_{L,n} > 0$.

Geometric mixing of the Gaussian panel implies ergodicity of its nonoverlapping blocks. Proposition 11 supplies the positive expected block growth used in the final implication for sufficiently large panels above the population frontier. The additional alignment condition says that concatenating expanding blocks cannot repeatedly rotate the maximizing direction into a nearly orthogonal contracting direction. Lemma IA.2 is therefore a diagnostic of the missing proof step rather than an additional primitive assumption of the model. The paper does not claim that equation (IA.15) follows from the Gaussian panel primitives.

The main text uses the probability limit $\overline{A}$ as the deterministic large-panel benchmark. Two qualifications complete that connection.

Remark IA.1 (The large-panel matrix is not a mean)**.** *The matrix $\overline{A}$ is the probability limit of A_t. When $\mathbb{E}[A_t]$ exists, it need not equal $\overline{A}$, because taking expectations does not remove the finite-sample effects of demeaning and division by s_t^{xx}. For small panels, the required inverse design moment may not exist. Even when all relevant expectations are finite, correlation between A_t and the inherited state prevents replacing*

$\mathbb{E}[A_t Z_t]$ *by* $\mathbb{E}[A_t]\mathbb{E}[Z_t]$*. The finite-panel stationary mean therefore need not equal the fixed point of the deterministic recursion. The simulated time-average slopes reported in Appendix Appendix B exceed the population fixed points, and the gaps decrease across the reported panel sizes. These calculations do not establish a* $1/n$ *expansion, its sign, or its leading coefficient. Interpreting a time average as a stationary mean also requires the relevant stationary first moment to exist.*

Proposition IA.4 (Finite-horizon large-panel convergence)**.** *Fix* $T < \infty$*. Suppose* $Z_0 = (\lambda_0, p_{-1})'$ *converges in probability to a deterministic limit* z_0 *as* $n \to \infty$*, and suppose the initial capital cross-section has uniformly bounded fourth moments. Then*

$$\max_{0 \le t \le T} \|Z_t - z_t\| \xrightarrow{p} 0, \qquad z_{t+1} = \overline{A} z_t + (\rho, 0)'. \tag{IA.17}$$

Proposition IA.4 justifies reading $\overline{A}$ as the deterministic large-panel dynamics over every fixed horizon. It does not interchange the limits in time and panel size. Unbounded-horizon behavior is governed by the Lyapunov results in Section 4.2.

Remark IA.2 (Full-state and slope-specific persistence)**.** *The Lyapunov exponent governs the full reduced state, while an auditor typically observes the deployed slope. These persistence rates need not coincide when the dominant Lyapunov direction lies mainly in capital coordinates. Equality would require an irreducibility or alignment condition that rules out asymptotically hidden capital directions. Capital directions disconnected from the score decay at rate* $\delta < 1$*. A positive full-state exponent therefore requires interaction with the score, but this observation alone does not establish the growth rate of a particular slope path or its stationary autocorrelation.*

Remark IA.3 (Moments and the critical case)**.** *A negative Lyapunov exponent guarantees a stationary solution but does not by itself guarantee finite first or second moments. When the Gaussian forcing has positive Student scale, it has a finite absolute first moment exactly for* $n \ge 3$ *and a finite second moment exactly for* $n \ge 4$*. These forcing moments do not by themselves establish the corresponding moments of the stationary state, because random coefficient products can amplify rare disturbances. At* $\gamma_{L,n} = 0$*, the dynamic theory does not characterize recurrence or fluctuation rates. These qualifications are distinct from the exact scalar frontier in Theorem 7.*

The admissible capital state lies in the centered hyperplane $\mathbb{1}^{\perp}$. The common-capital direction has eigenvalue δ in the unreduced system but is absent from both the centered economic state and the forcing term. Figure B.2 depicts this geometry. The one-period spectral decomposition in Lemma IA.1 explains the coexistence of the coupled score-capital mode and capital directions that decay at rate δ, but rotating feature-aligned subspaces prevent those one-period eigenvalues from determining the long-run Lyapunov exponent.

IA.4 Robustness and archive diagnostics

The exact Gaussian benchmark delivers a closed-form innovation distribution. This section gives its stationary tail implications, then studies which archive results survive weaker sampling assumptions. For the Gaussian diagnostics, retain the fresh-sample environment of Theorem 6, with $n \ge 2$, $V > 0$, and positive innovation scale $\sigma_u > 0$.

Corollary IA.1 (Stationary tail index under Gaussian sampling)**.** *Suppose* $|k| < 1$*. The stationary coefficient* λ_∞ *has regularly varying tails with index* $n - 1$*. Consequently, for every* $p > 0$*,*

$$\mathbb{E}[|\lambda_\infty|^p] < \infty \quad \Longleftrightarrow \quad p < n - 1. \tag{IA.18}$$

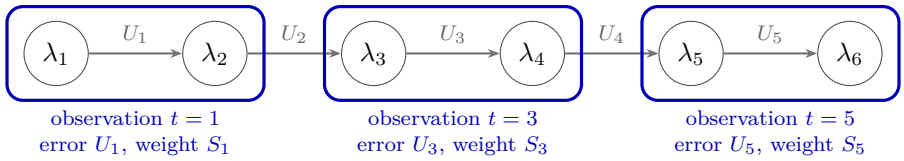


Figure IA.2: The exact archive test. Vintage slopes form a chain in which each innovation U_t carries slope λ_t into slope λ_{t+1}, so on the full archive every slope serves as regressand of one observation and regressor of the next. Under the null $k = 0$, using only the odd observations, boxed, keeps the error set $\{U_1, U_3, U_5, \dots\}$ disjoint from the innovations embedded in the regressors, and the weighted regression with the period-t design sum S_t as weight is then a classical normal linear model with an exactly Student-t slope statistic.

The tail index determines which population moments are available for archive diagnostics. In particular, a stationary coefficient can exist even when its variance or mean does not. These thresholds depend on the exact Student innovation law and need not survive a change in the sampling distribution.

For the non-Gaussian comparison, retain the affine fresh-sample economy and write $\widetilde{x}_{i,t} = \widetilde{\theta}_{i,t} + \eta_{i,t}$. The population linear projection coefficient is $\rho = \sigma_\theta^2/V$, and its outcome error is $u_{i,t} = (1-\rho)\widetilde{\theta}_{i,t} - \rho\eta_{i,t} + \varepsilon_{i,t}$. These quantities describe a linear projection even when the conditional fundamental mean is nonlinear. Define $S_t = \sum_i (\widetilde{x}_{i,t} - \overline{x}_t)^2$, $N_t = \sum_i (\widetilde{x}_{i,t} - \overline{x}_t)u_{i,t}$, and $U_t = N_t/S_t$. When its expectation exists, write $b_n = \mathbb{E}[U_t]$ and $\widetilde{U}_t = U_t - b_n$. Let $\widehat{k}_T$ and $\widehat{a}_T$ be the slope and intercept from OLS of λ_{t+1} on an intercept and λ_t over T archive observations.

Proposition IA.5 (Robustness beyond Gaussianity)**.** *Suppose the centered primitives $\widetilde{\theta}_{i,t}$, $\eta_{i,t}$, and $\varepsilon_{i,t}$ are mutually independent within units, have mean zero and finite fourth moments, and are identically distributed and independent across units and periods. Assume $V > 0$, $\mathbb{E}[S_t^{-2}] < \infty$, and a finite nonanticipative initial coefficient. Then $\mathbb{E}[U_t^2] < \infty$, and retraining satisfies*

$$\lambda_{t+1} = (\rho + b_n) + k\lambda_t + \widetilde{U}_t. \tag{IA.19}$$

The innovations $\widetilde{U}_t$ are i.i.d., centered, square integrable, and independent of the previously deployed coefficient. The following statements hold.

(i) If $|k| < 1$, the causal recursion has a unique stationary distribution, with $\mathbb{E}[\lambda_\infty] = (\rho + b_n)/(1-k)$ and $\mathrm{Var}(\lambda_\infty) = \mathrm{Var}(\widetilde{U}_t)/(1-k^2)$.

(ii) If $|k| < 1$, $\mathrm{Var}(\widetilde{U}_t) > 0$, and the process is stationary, then $\widehat{k}_T \to k$ and $\widehat{a}_T \to \rho + b_n$ almost surely. Moreover, $\sqrt{T}(\widehat{k}_T - k) \Rightarrow N(0, 1-k^2)$.

(iii) If $|k| > 1$ and the innovation distribution has an atomless component, then $|\lambda_t| \to \infty$ almost surely at geometric rate $|k|^t$. More precisely, for $\widetilde{\lambda} = (\rho + b_n)/(1-k)$, the ratio $(\lambda_t - \widetilde{\lambda})/k^t$ converges almost surely to a finite nonzero random variable. If $k < -1$, the sign of $\lambda_t - \widetilde{\lambda}$ eventually alternates each period.

Proposition IA.5 separates the implications of linear feedback from those of Gaussian sampling. Linearity and fresh cross-sections preserve the autoregressive law and the exogeneity of vintage innovations. Causal stationarity holds for $|k| < 1$. Under the stated atomless-component condition, divergence holds for $|k| > 1$, including alternating divergence under strong negative feedback. Gaussianity delivers the stronger finite-sample statements. It eliminates the intercept bias, gives symmetric Student innovations with an exact variance, and ties the stationary tail index and moment thresholds to the cross-sectional sample size. The weaker sampling assumptions do not preserve these distributional conclusions. At $k = 1$, the average drift becomes $\rho + b_n$ and need not equal the positive Gaussian drift. At $k = -1$, multiplying $\lambda_t - (\rho + b_n)/2$ by $(-1)^t$ still produces independent centered increments, but they need not be identically distributed without innovation symmetry. Figure IA.2 instead illustrates the alternate-vintage exact test in Proposition 8(ii).

This distinction determines which diagnostics are robust. Persistence in the coefficient archive continues to identify the loop gain under the conditions of Proposition IA.5. Diagnostics based on symmetry, exact variances, or tail indices require the Gaussian specification. Slope-based auditing therefore rests on a broader set of assumptions than tail-based auditing.

Corollary IA.2 (Noise amplification near the stability frontier)**.** *In the Gaussian fresh-sample environment, suppose $n \geq 4$, $\sigma_u > 0$, and $k \in [0, 1)$. The stationary standard deviation of the deployed slope is*

$$\mathrm{sd}(\lambda_\infty) = \frac{\sigma_u}{\sqrt{V(n-3)}} \frac{1}{\sqrt{1-k^2}}. \tag{IA.20}$$

It is strictly increasing in k and satisfies $\mathrm{sd}(\lambda_\infty) \longrightarrow \infty$ as $k \uparrow 1$.

Holding sample size, feature variance, reliability, and outcome-noise variance fixed, higher stationary dispersion of coefficient levels corresponds to a higher nonnegative feedback gain. Level dispersion differs from the volatility of adjacent coefficient revisions. Under stationarity,

$$\mathrm{Var}\left(\lambda_{t+1} - \lambda_t\right) = \frac{2\sigma_u^2}{V(n-3)} \frac{1}{1+k}. \tag{IA.21}$$

This variance falls as the loop gain rises. Near the frontier, the coefficient changes little from one vintage to the next because it is highly persistent, even though its long-run level varies widely across the archive. A sequence of models can therefore appear locally stable while drifting over a much wider range at longer horizons.

An auditor who monitors only adjacent coefficient revisions may interpret this local smoothness as evidence of stability. Within a family of stationary regimes approaching the frontier, the model predicts rising persistence and level dispersion together with falling revision variance. A rolling implementation requires the underlying gain and sampling environment to remain approximately stable over the estimation window. Changes in sample size, feature variance, feature reliability, or residual outcome uncertainty must also be controlled before a dispersion trend is attributed to feedback.

The same archive provides identifying variation that is absent from a single deterministic deployment.

Remark IA.4 (An overidentifying restriction)**.** *In the stationary Gaussian archive with $n \geq 4$, the innovation variance satisfies $\mathrm{Var}(U_t) = \sigma_u^2/[V(n-3)]$, while the structural residual variance satisfies $\sigma_u^2 = \rho(1-\rho)V + \sigma_\varepsilon^2$. Because $\sigma_\varepsilon^2 \geq 0$, the model imposes the population restriction $\sigma_u^2 \geq \rho(1-\rho)V$, together with the restriction $0 \leq \rho \leq 1$ on feature reliability.*

Define the archive-implied residual variance by $\widehat{\sigma}_u^2 \coloneqq (n-3)\widehat{V}\,\widehat{\mathrm{Var}}(U_t)$, where $\widehat{V}$ estimates the cross-sectional feature variance and $\widehat{\mathrm{Var}}(U_t)$ estimates the variance of the archive regression residuals. A specification test assesses the joint restrictions on feature reliability and residual variance using $\widehat{\rho}$ and $\widehat{\sigma}_u^2 - \widehat{\rho}(1-\widehat{\rho})\widehat{V}$, allowing for their sampling uncertainty. The estimates need not satisfy the population restrictions exactly in a finite sample. Assessing the restrictions jointly matters because the variance inequality can become mechanically uninformative when $\widehat{\rho}$ lies outside $[0, 1]$.

This test requires no external estimate of outcome noise. Its implementation must account for the innovation tails. A finite innovation variance alone does not justify Gaussian standard errors for an estimated second moment. Under the Student law, the fourth innovation moment exists only when $n \geq 6$. If an independent estimate $\widehat{\sigma}_\varepsilon^2$ is available, the population equality $\sigma_u^2 = \rho(1-\rho)V + \sigma_\varepsilon^2$ supplies a further restriction, which can be assessed using the discrepancy $\widehat{\sigma}_u^2 - \widehat{\rho}(1-\widehat{\rho})\widehat{V} - \widehat{\sigma}_\varepsilon^2$ and its sampling uncertainty.

The restriction in Remark IA.4 tests more than the autoregressive fit of the coefficient archive. A time series may satisfy the reduced-form law in equation (61) while failing the variance decomposition implied by the latent-fundamental model. A violation rejects the joint structural restrictions. Satisfaction is compatible with the feedback mechanism but does not by itself establish that mechanism as the source of persistence.

IA.5 Pooled retraining as stochastic approximation

The pooled estimator admits an exact stochastic-approximation representation. Pool the within-period centered cross-sections, allowing a separate intercept for each period. Let $G_t = \sum_{s=1}^{t} S_s$ and $\alpha_t = S_t/G_t$. For $t \geq 2$, the previous pooled numerator is $G_{t-1}\lambda_t$, while the new cross-section contributes $S_t(\rho + k\lambda_t + U_t)$. Dividing their sum by G_t gives

$$\lambda_{t+1} = \lambda_t + \alpha_t \left[\rho + (k-1)\lambda_t + U_t\right]. \tag{IA.22}$$

The same identity holds at $t = 1$ because $\alpha_1 = 1$. Under the Gaussian design, $G_t/t \to V(n-1) > 0$ almost surely, so $\alpha_t = O_p(t^{-1})$. More precisely, $\alpha_t \to 0$ almost surely, $\sum_{s=2}^{t} \alpha_s = \log t + O(1)$ almost surely, and $\sum_{t=2}^{\infty} \alpha_t^2 < \infty$ almost surely. To obtain the first and third claims, use $G_t \asymp t$ and the finite second moment of S_t, which imply $S_t/t \to 0$ and $\sum_t S_t^2/t^2 < \infty$ almost surely. For the cumulative gain, sum $-\log(1-\alpha_t) = \log G_t - \log G_{t-1}$ from $t = 2$ and use the summable quadratic remainder. The gain remains random, and $t\alpha_t$ need not converge to a constant. Sampling error receives progressively less weight, while the cumulative adjustment remains unbounded.

For $k \neq 1$, the associated mean differential equation $d\lambda/d\tau = \rho + (k-1)\lambda$ has the unique rest point $\lambda^* = \rho/(1-k)$. This rest point is globally attracting exactly when $k < 1$. At $k = 1$, the equation becomes $d\lambda/d\tau = \rho$ and has no rest point under the maintained assumption $\rho > 0$. The decreasing-gain stability region therefore includes negative-feedback cases in which repeated application of the population retraining map is unstable.

Part (i) of Proposition 9 gives a global version of the E-stability principle developed in Marcet and Sargent (1989) and Evans and Honkapohja (2001). The general theory of self-referential learning often relies on projection devices or boundedness assumptions to keep beliefs within a compact region. The affine retraining map and regime-invariant disturbance make those auxiliary restrictions unnecessary here. Below the frontier, the decreasing gain eventually suppresses sampling variation, and the coefficient converges almost surely to the performatively stable value.

A credit-scoring system illustrates the mechanism. Early cross-sections may produce unusually high or low estimates of the sensitivity of repayment to borrower characteristics. As the lending archive expands, each new estimation error receives less weight, while the accumulated data continue to pull the coefficient toward the stable slope. The model eventually settles at a fixed coefficient even though every historical outcome was affected by the lending policy in place when that outcome was produced.

Part (iii) of Proposition 9 strengthens the usual nonconvergence result for an unstable stochastic-approximation equilibrium (Pemantle, 1990). Above the frontier, the deterministic force in equation (IA.22) pushes the coefficient away from the rest point. Conditional Gaussianity and linearity rule out permanent cancellation by the sampling disturbances, so the coefficient diverges almost surely at the exact polynomial rate in equation (74).

Fresh-sample and pooled retraining describe two limiting approaches to model updating. Under fresh-sample retraining, the learning gain remains constant, and the coefficient fluctuates indefinitely around the stable point only when $|k| < 1$. Under pooled retraining, the gain vanishes, and the coefficient converges to the fixed point for every $k < 1$, including the strongly negative region in which fresh-sample retraining oscillates or explodes. Rolling windows and exponential downweighting lie between these cases. They retain a positive effective gain because older observations eventually leave the estimation sample or receive negligible weight, so persistent coefficient fluctuations remain relevant for such systems.

Near the frontier, a positive effective gain interacts with the heavy-tailed estimation disturbances characterized in Theorem 6. This interaction may produce occasional departures from the neighborhood of the stable coefficient of the kind studied in the escape dynamics literature (Sargent, 1999; Cho et al., 2002). A full characterization of these excursions under rolling or exponentially weighted retraining remains open.

IA.6 Proofs

IA.6.1 Theorems

Proof of Theorem IA.1. Write $\kappa = \kappa_{\text{abs}} < 1$ and $\|Y\|_2 = \|Y\|_{L^2(\Omega;\mathcal{H})}$. The mean value theorem gives $|f(a)-f(b)| \le \kappa|a-b|$. Since $|f(a)| \le |f(0)|+\kappa|a|$ and $u \in \mathcal{H}$, the operator T maps $\mathcal{H}$ into itself. Moreover, $\|Tm_1 - Tm_2\|_{\mathcal{H}} \le \kappa\|m_1 - m_2\|_{\mathcal{H}}$. The stable score $m^* = \Phi(u)$ belongs to $\mathcal{H}$ because Φ is Lipschitz and $u \in \mathcal{H}$.

For $\tau > 0$, set $Z_t^\tau = (M_t^\tau - m^*)/\tau$. Subtracting $Tm^* = m^*$ from (IA.1) gives $Z_{t+1}^\tau = G_\tau(Z_t^\tau) + \xi_{t+1}$, where $(G_\tau z)(x) = \frac{f(m^*(x)+\tau z(x))-f(m^*(x))}{\tau}$. The Lipschitz bound on f implies $G_\tau 0 = 0$ and $\|G_\tau z - G_\tau z'\|_{\mathcal{H}} \le \kappa\|z - z'\|_{\mathcal{H}}$.

Part (i). Start this recursion from zero at date $t-J$ and call its value at date t $Z_{t,J}^\tau$. Repeated application of the contraction bound gives $\|Z_{t,J}^\tau\|_{\mathcal{H}} \le \sum_{j=0}^{J-1}\kappa^j\|\xi_{t-j}\|_{\mathcal{H}}$, $\|Z_{t,J}^\tau\|_2 \le \frac{\|\xi_0\|_2}{1-\kappa}$. The backward constructions of lengths $J+1$ and J differ initially by ξ_{t-J}. Their subsequent J updates use the same innovations, so $\|Z_{t,J+1}^\tau - Z_{t,J}^\tau\|_2 \le \kappa^J\|\xi_0\|_2$. Telescoping over construction lengths shows that, for $K > J$, $\|Z_{t,K}^\tau - Z_{t,J}^\tau\|_2 \le \sum_{\ell=J}^{K-1}\kappa^\ell\|\xi_0\|_2 \le \frac{\kappa^J}{1-\kappa}\|\xi_0\|_2$. Thus $Z_{t,J}^\tau$ converges in $L^2(\Omega;\mathcal{H})$ as $J \to \infty$. Its limit Z_t^τ is causal and stationary, and continuity of G_τ lets us pass to the limit in the recursion. Consequently $M_t^\tau = m^* + \tau Z_t^\tau$ is a causal stationary solution of (IA.1).

If Y_t and $\widetilde{Y}_t$ are two L^2 stationary solutions driven by the same innovations, contraction over J periods gives $\|Y_t - \widetilde{Y}_t\|_2 \le \kappa^J(\|Y_{t-J}\|_2 + \|\widetilde{Y}_{t-J}\|_2)$. Stationarity keeps the two norms on the right bounded as J grows. Letting $J \to \infty$ proves uniqueness.

Part (ii). The multiplication operator A is bounded because $|f'(m^*(x))| \le \kappa$ almost everywhere, so $\|A\|_{\text{op}} \le \kappa$. For fixed $z \in \mathcal{H}$, differentiability of f gives $G_\tau z(x) \to Az(x)$ almost everywhere. The bound $|G_\tau z(x) - Az(x)| \le 2\kappa|z(x)|$ and dominated convergence imply $\|G_\tau z - Az\|_{\mathcal{H}} \to 0$. If $Y \in L^2(\Omega;\mathcal{H})$, this deterministic convergence applies to $Y(\omega)$ for almost every ω. Since $\|G_\tau Y - AY\|_{\mathcal{H}}^2 \le 4\kappa^2\|Y\|_{\mathcal{H}}^2$, a second application of dominated convergence gives $\|G_\tau Y - AY\|_2 \to 0$.

Let $Z_{t,J}$ be the length-J backward construction for $Z_{t+1} = AZ_t + \xi_{t+1}$. For $J = 1$, both constructions equal ξ_t. Suppose $Z_{t-1,J-1}^\tau \to Z_{t-1,J-1}$ in $L^2(\Omega;\mathcal{H})$. Adding and subtracting $G_\tau(Z_{t-1,J-1})$ gives $\|Z_{t,J}^\tau - Z_{t,J}\|_2 \le \kappa\|Z_{t-1,J-1}^\tau - Z_{t-1,J-1}\|_2 + \|G_\tau(Z_{t-1,J-1}) - AZ_{t-1,J-1}\|_2$. Both terms tend to zero. Induction therefore proves $Z_{t,J}^\tau \to Z_{t,J}$ for every fixed J.

The nonlinear backward construction has truncation error at most $\kappa^J\|\xi_0\|_2/(1-\kappa)$. The same bound holds for the linear construction because $\|A\|_{\text{op}} \le \kappa$. Hence $\|Z_t^\tau - Z_t\|_2 \le \frac{2\kappa^J}{1-\kappa}\|\xi_0\|_2 + \|Z_{t,J}^\tau - Z_{t,J}\|_2$. First choose J large and then let $\tau \downarrow 0$. This proves $Z_t^\tau \to Z_t$ in $L^2(\Omega;\mathcal{H})$. Since $Z_{t,J} = \sum_{j=0}^{J-1}A^j\xi_{t-j}$ and $\|A^j\|_{\text{op}} \le \kappa^j$, passing to the limit in J gives the series in (IA.2). The series and $\mathbb{E}[\xi_t] = 0$ also give $\mathbb{E}[Z_t] = 0$.

Part (iii). For $h \in \mathcal{H}$, substitute $Z_{t+1} = AZ_t + \xi_{t+1}$ into the definition of C_1. The variable Z_t depends only on innovations through date t, so it is independent of ξ_{t+1}. Thus $C_1 h = \mathbb{E}[\langle Z_t, h\rangle_{\mathcal{H}} AZ_t] + \mathbb{E}[\langle Z_t, h\rangle_{\mathcal{H}}\xi_{t+1}] = A\mathbb{E}[\langle Z_t, h\rangle_{\mathcal{H}} Z_t] = AC_0 h$. This proves (IA.3).

To address the covariance operators at positive noise scales, let $\mu_\tau = \mathbb{E}[Z_t^\tau]$ and $\overline{Z}_t^\tau = Z_t^\tau - \mu_\tau$. Stationarity makes μ_τ independent of t. Since $Z_t^\tau \to Z_t$ in L^2 and $\mathbb{E}[Z_t] = 0$, we have $\mu_\tau \to 0$ and $\overline{Z}_t^\tau \to Z_t$ in L^2. Define $C_{0,\tau}h = \mathbb{E}[\langle \overline{Z}_t^\tau, h\rangle_{\mathcal{H}}\overline{Z}_t^\tau]$, $C_{1,\tau}h = \mathbb{E}[\langle \overline{Z}_t^\tau, h\rangle_{\mathcal{H}}\overline{Z}_{t+1}^\tau]$. Adding and subtracting the corresponding mixed product, then applying Cauchy–Schwarz, gives $\|C_{0,\tau} - C_0\|_{\text{op}} \le \|\overline{Z}_t^\tau - Z_t\|_2(\|\overline{Z}_t^\tau\|_2 + \|Z_t\|_2) \longrightarrow 0$. The same argument applied to the pair $(\overline{Z}_t^\tau, \overline{Z}_{t+1}^\tau)$ gives $\|C_{1,\tau} - C_1\|_{\text{op}} \to 0$.

Finally, let $\mathcal{K} = \overline{\text{Ran}(C_0)}$. If two bounded operators A and $\widetilde{A}$ satisfy $AC_0 = \widetilde{A}C_0 = C_1$, they agree on $\text{Ran}(C_0)$. Boundedness extends their agreement to $\mathcal{K}$. Thus (C_0, C_1) determines A on $\mathcal{K}$. If $\mathcal{K} = \mathcal{H}$, it determines A everywhere. In that case it also determines the multiplication function almost everywhere, since the constant function $\mathbb{1}$ belongs to $\mathcal{H}$ and $a = A\mathbb{1}$. When $\mathcal{K}$ is smaller, this argument still determines A on $\mathcal{K}$, but makes no further nonidentification claim within the class of multiplication operators. □

Proof of Theorem IA.2. For a deployed coefficient λ, equation (IA.6) gives $\mathrm{Cov}(\widetilde{x}, y^{(\lambda)}) = \mathrm{Cov}(\widetilde{x}, \theta) + \mathrm{Cov}(\widetilde{x}, f(\lambda'\widetilde{x})) + \mathrm{Cov}(\widetilde{x}, \varepsilon) = q + \mathrm{Cov}(\widetilde{x}, f(\lambda'\widetilde{x}))$, because $\mathbb{E}[\varepsilon\widetilde{x}] = 0$ and $\widetilde{x}$ is centered.

If $\lambda = 0$, the feedback is constant and $T0 = v \neq 0$, so zero is not stable. Fix $\lambda \neq 0$ and let $\varphi(z) = f(\lambda' z)$. Since f is Lipschitz and differentiable almost everywhere, φ is weakly differentiable and $\nabla\varphi(z) = f'(\lambda' z)\lambda$ almost everywhere. Multivariate Stein's identity for $\widetilde{x} \sim N(0, \Sigma)$ gives $\mathbb{E}[\widetilde{x}\varphi(\widetilde{x})] = \Sigma\mathbb{E}[\nabla\varphi(\widetilde{x})]$. Since $\mathbb{E}[\widetilde{x}] = 0$, the left-hand side equals the covariance $\mathrm{Cov}(\widetilde{x}, f(\lambda'\widetilde{x}))$. The right-hand side is $\Sigma\mathbb{E}[f'(\lambda'\widetilde{x})\lambda] = \Sigma\lambda\mathbb{E}[f'(\lambda'\widetilde{x})] = \bar{k}(\lambda)\Sigma\lambda$. Hence equation (IA.10) holds.

The population OLS update is $T\lambda = \Sigma^{-1}[q + \bar{k}(\lambda)\Sigma\lambda] = \Sigma^{-1}q + \bar{k}(\lambda)\lambda = v + \bar{k}(\lambda)\lambda$. At a stable coefficient λ^*, $\lambda^* = v + \bar{k}(\lambda^*)\lambda^*$. Rearranging gives $[1 - \bar{k}(\lambda^*)]\lambda^* = v$. The derivative bound implies $\bar{k}(\lambda) = \mathbb{E}[f'(\lambda'\widetilde{x})] \leq \kappa^+ < 1$, so $1 - \bar{k}(\lambda^*) > 0$. Dividing gives equation (IA.8). In particular, λ^* is a positive scalar multiple of v, so it is collinear with the no-feedback coefficient.

It remains to establish existence. Since $q \neq 0$ and Σ is nonsingular, $v = \Sigma^{-1}q \neq 0$. Search for a fixed point on the ray $\lambda = cv$ with $c \geq 0$. Define $K(c) = \bar{k}(cv) = \mathbb{E}[f'(cv'\widetilde{x})]$ for $c > 0$, set $\xi(c) = c[1 - K(c)]$ for $c > 0$, and define $\xi(0) = 0$. If $\xi(c) = 1$, then $[1 - K(c)]cv = v$, which is exactly the fixed-point equation.

We verify that such a c exists. At $c = 0$, $\xi(0) = 0$. Also $K(c) \leq \kappa^+$ for every $c > 0$, so $1 - K(c) \geq 1 - \kappa^+ > 0$ and $\xi(c) \geq c(1 - \kappa^+) \to \infty$ as $c \to \infty$. It remains to verify continuity of ξ. For $c > 0$, the scalar $cv'\widetilde{x}$ is Gaussian with variance $c^2\sigma_v^2$, where $\sigma_v^2 = v'\Sigma v > 0$. Hence $K(c) = \int f'(w)\varphi_{c\sigma_v}(w)\,dw$, where φ_s is the centered Gaussian density with standard deviation s. Since f' is bounded by the Lipschitz constant of f and the Gaussian densities are continuous in L^1 as c varies away from zero, $K(c)$ is continuous on $(0, \infty)$. At zero, we do not need continuity of K itself because $|\xi(c)| = |c[1-K(c)]| \leq c[1+|K(c)|] \leq c(1+\mathrm{Lip}(f)) \to 0 = \xi(0)$. Thus ξ is continuous on $[0, \infty)$. By the intermediate value theorem, there exists $c^* > 0$ with $\xi(c^*) = 1$. Then $\lambda^* = c^*v$ is a stable coefficient.

Uniqueness follows from the same one-sided slope bound. Lipschitz continuity makes f absolutely continuous on bounded intervals, so $f(b) - f(a) \leq \kappa^+(b-a)$ whenever $b > a$. Therefore $(b-a)[f(b)-f(a)] \leq \kappa^+(b-a)^2$ for all a, b. Suppose λ_1 and λ_2 are stable and write $\Delta\lambda = \lambda_1 - \lambda_2$. Their normal equations give $\Sigma\Delta\lambda = \mathbb{E}[\widetilde{x}\{f(\lambda_1'\widetilde{x}) - f(\lambda_2'\widetilde{x})\}]$. Premultiplying by $(\Delta\lambda)'$ and applying the scalar inequality yields $(\Delta\lambda)'\Sigma\Delta\lambda \leq \kappa^+\mathbb{E}[(\Delta\lambda'\widetilde{x})^2] = \kappa^+(\Delta\lambda)'\Sigma\Delta\lambda$. Hence $(1 - \kappa^+)(\Delta\lambda)'\Sigma\Delta\lambda \leq 0$. Positive definiteness of Σ and $\kappa^+ < 1$ force $\Delta\lambda = 0$.

If $f(t) = a + kt$, then $f'(t) = k$ everywhere, so $\bar{k}(\lambda) = k$ for every λ. The fixed-point equation becomes $(1 - k)\lambda^* = v$. Since $k \leq \kappa^+ < 1$, the denominator is nonzero and the unique solution is $\lambda^* = v/(1 - k)$, proving equation (IA.9). □

IA.6.2 Propositions

Proof of Proposition IA.1. Because $e_1, \ldots, e_J$ are eigenfunctions of C_0, the subspace $\mathcal{H}_J$ is invariant under C_0. For every $h \in \mathcal{H}_J$, equation (IA.3) gives $C_{1,J}h = P_J C_1 h = P_J A C_0 h$. Since $C_0 h \in \mathcal{H}_J$, inserting P_J before $C_0 h$ does not change the vector. By definition of $A_J = P_J A|_{\mathcal{H}_J}$, $C_{1,J}h = A_J C_{0,J}h$. Thus $C_{1,J} = A_J C_{0,J}$ on $\mathcal{H}_J$. The smallest eigenvalue of $C_{0,J}$ is $\nu_J > 0$, so $C_{0,J}$ is invertible and $\|C_{0,J}^{-1}\| = 1/\nu_J$. Right multiplication by the inverse gives equation (IA.4).

For the sensitivity bound, write $\widehat{C}_0 = \widehat{C}_{0,J}$, $\widehat{C}_1 = \widehat{C}_{1,J}$, $C_0 = C_{0,J}$, and $C_1 = C_{1,J}$. For every $h \in \mathcal{H}_J$, the triangle inequality gives $\|\widehat{C}_0 h\| \geq \|C_0 h\| - \|(\widehat{C}_0 - C_0)h\| \geq (\nu_J - \epsilon_0)\|h\|$. Thus $\widehat{C}_0$ is injective on the finite-dimensional space $\mathcal{H}_J$, hence invertible, and $\|\widehat{C}_0^{-1}\| \leq \frac{1}{\nu_J - \epsilon_0}$. Use the decomposition $\widehat{A}_J - A_J = (\widehat{C}_1 - C_1)\widehat{C}_0^{-1} + C_1(\widehat{C}_0^{-1} - C_0^{-1})$. The inverse identity $\widehat{C}_0^{-1} - C_0^{-1} = \widehat{C}_0^{-1}(C_0 - \widehat{C}_0)C_0^{-1}$ implies $\|\widehat{C}_0^{-1} - C_0^{-1}\| \leq \frac{\epsilon_0}{(\nu_J - \epsilon_0)\nu_J}$. Taking norms in the decomposition and using $\|\widehat{C}_1 - C_1\| \leq \epsilon_1$ therefore gives $\|\widehat{A}_J - A_J\| \leq \frac{\epsilon_1}{\nu_J - \epsilon_0} + \frac{\|C_1\|\epsilon_0}{\nu_J(\nu_J - \epsilon_0)}$, which is equation (IA.5). The factors involving $1/\nu_J$ show directly why directions with little archive variance are poorly conditioned. □

Proof of Proposition IA.2. Since $T0 = v \neq 0$, zero cannot be a stable coefficient. Fix a nonzero deployed

coefficient λ and let $S = \lambda'\widetilde{x}$. Under ellipticity with finite second moments, the conditional expectation of the feature vector given the scalar projection is linear: $\mathbb{E}[\widetilde{x} \mid S] = \mathrm{Cov}(\widetilde{x}, S)\mathrm{Var}(S)^{-1}S$. Since $\mathrm{Cov}(\widetilde{x}, S) = \mathrm{Cov}(\widetilde{x}, \lambda'\widetilde{x}) = \Sigma\lambda$ and $\mathrm{Var}(S) = \lambda'\Sigma\lambda$, this becomes $\mathbb{E}[\widetilde{x} \mid S] = (\Sigma\lambda/(\lambda'\Sigma\lambda))S$.

Because $\widetilde{x}$ is centered, $\mathrm{Cov}(\widetilde{x}, f(S)) = \mathbb{E}[\widetilde{x}f(S)] - \mathbb{E}[\widetilde{x}]\mathbb{E}[f(S)] = \mathbb{E}[\widetilde{x}f(S)]$. Apply iterated expectations: $\mathbb{E}[\widetilde{x}f(S)] = \mathbb{E}[\mathbb{E}[\widetilde{x} \mid S]f(S)]$. Substituting the linear conditional expectation gives $\mathrm{Cov}(\widetilde{x}, f(S)) = (\Sigma\lambda/(\lambda'\Sigma\lambda))\mathbb{E}[Sf(S)]$. Since $\mathbb{E}[S] = 0$, $\mathbb{E}[Sf(S)] = \mathrm{Cov}(S, f(S))$. The scalar multiplier is therefore $\mathrm{Cov}(S, f(S))/\mathrm{Var}(S) = \bar{k}(\lambda)$, giving $\mathrm{Cov}(\widetilde{x}, f(S)) = \bar{k}(\lambda)\Sigma\lambda$.

The population OLS update is therefore $T\lambda = \Sigma^{-1}[q + \bar{k}(\lambda)\Sigma\lambda] = v + \bar{k}(\lambda)\lambda$. At a fixed point, $(1 - \bar{k}(\lambda^*))\lambda^* = v$, so any stable coefficient is collinear with v provided $1 - \bar{k}(\lambda^*) \neq 0$.

We now establish the bound $\bar{k}(\lambda) \leq \kappa^+$. Let S' be an independent copy of S. The covariance identity $2\mathrm{Cov}(S, f(S)) = \mathbb{E}[(S - S')(f(S) - f(S'))]$ follows by expanding the right-hand side. Since f is Lipschitz and has derivative at most κ^+ almost everywhere, for $s_2 > s_1$ the fundamental theorem for absolutely continuous Lipschitz functions gives $f(s_2) - f(s_1) = \int_{s_1}^{s_2} f'(v)\,dv \leq \kappa^+(s_2 - s_1)$. Therefore $(S - S')(f(S) - f(S')) \leq \kappa^+(S - S')^2$ pointwise. Taking expectations yields $2\mathrm{Cov}(S, f(S)) \leq \kappa^+\mathbb{E}[(S - S')^2] = 2\kappa^+\mathrm{Var}(S)$. Dividing by $2\mathrm{Var}(S) > 0$ gives $\bar{k}(\lambda) \leq \kappa^+ < 1$. Thus the denominator in equation (IA.12) is strictly positive.

For existence, again restrict to the ray $\lambda = cv$, $c \geq 0$. Define $K(c) = \bar{k}(cv)$ and $\xi(c) = c[1 - K(c)]$ for $c > 0$, and set $\xi(0) = 0$. The preceding bound implies $\xi(c) \geq c(1 - \kappa^+) \to \infty$. Also $\xi(0) = 0$. It remains to establish continuity for $c > 0$. Let $S_0 = v'\widetilde{x}$. Then for $c > 0$, $\bar{k}(cv) = \mathrm{Cov}(cS_0, f(cS_0))/\mathrm{Var}(cS_0) = \mathrm{Cov}(S_0, f(cS_0))/(c\mathrm{Var}(S_0))$. If $c_n \to c > 0$, Lipschitz continuity gives $|f(c_nS_0) - f(cS_0)| \leq L|c_n - c||S_0|$. Since S_0 has finite second moment, $|S_0[f(c_nS_0) - f(cS_0)]| \leq L|c_n - c|S_0^2$, whose expectation tends to zero. Thus the numerator covariance is continuous in c, and the denominator $c\mathrm{Var}(S_0)$ is continuous and nonzero. Therefore $K(c)$ and $\xi(c)$ are continuous on $(0, \infty)$. The independent-copy covariance identity and $|f(b) - f(a)| \leq L|b - a|$ also give $|K(c)| \leq L$. Thus $|\xi(c)| \leq c(1 + L) \to 0$ as $c \downarrow 0$, establishing continuity at zero. The intermediate value theorem gives a $c^* > 0$ with $\xi(c^*) = 1$. Then $(1 - K(c^*))c^*v = v$, so $\lambda^* = c^*v$ is stable.

It remains to establish uniqueness. Since f is Lipschitz, it is absolutely continuous on every bounded interval. The derivative bound therefore implies $f(b) - f(a) \leq \kappa^+(b - a)$ whenever $b > a$. Reversing the arguments when necessary gives $(b - a)[f(b) - f(a)] \leq \kappa^+(b - a)^2$ for all $a, b \in \mathbb{R}$.

Suppose λ_1 and λ_2 are stable coefficients and write $\Delta\lambda = \lambda_1 - \lambda_2$. Their population normal equations are $\Sigma\lambda_j = q + \mathbb{E}[\widetilde{x}f(\lambda_j'\widetilde{x})]$ for $j = 1, 2$. Subtracting gives $\Sigma\Delta\lambda = \mathbb{E}[\widetilde{x}\{f(\lambda_1'\widetilde{x}) - f(\lambda_2'\widetilde{x})\}]$. Premultiplying by $(\Delta\lambda)'$ yields $(\Delta\lambda)'\Sigma\Delta\lambda = \mathbb{E}[(\Delta\lambda'\widetilde{x})\{f(\lambda_1'\widetilde{x}) - f(\lambda_2'\widetilde{x})\}]$.

Apply the preceding scalar inequality with $b = \lambda_1'\widetilde{x}$ and $a = \lambda_2'\widetilde{x}$. The right-hand side is at most $\kappa^+\mathbb{E}[(\Delta\lambda'\widetilde{x})^2] = \kappa^+(\Delta\lambda)'\Sigma\Delta\lambda$, where the equality uses $\mathbb{E}[\widetilde{x}] = 0$. Rearranging gives $(1 - \kappa^+)(\Delta\lambda)'\Sigma\Delta\lambda \leq 0$. Since $1 - \kappa^+ > 0$ and Σ is positive definite, this inequality requires $\Delta\lambda = 0$. Thus $\lambda_1 = \lambda_2$, proving uniqueness.

Under Gaussian features, Stein's identity from Theorem IA.2 gives $\mathrm{Cov}(S, f(S)) = \mathrm{Var}(S)\mathbb{E}[f'(S)]$. Dividing by $\mathrm{Var}(S)$ recovers $\bar{k}(\lambda) = \mathbb{E}[f'(\lambda'\widetilde{x})]$. □

Proof of Proposition IA.3. For integers $0 \leq r < s$, define $X_{r,s} = \log\|\widetilde{B}_s\widetilde{B}_{s-1}\cdots\widetilde{B}_{r+1}\|$. The operator norm is submultiplicative, so for $r < s < t$, $\|\widetilde{B}_t\cdots\widetilde{B}_{r+1}\| \leq \|\widetilde{B}_t\cdots\widetilde{B}_{s+1}\|\,\|\widetilde{B}_s\cdots\widetilde{B}_{r+1}\|$. Taking logarithms gives the subadditivity relation $X_{r,t} \leq X_{s,t} + X_{r,s}$. Lemma 2 gives stationarity and ergodicity of the coefficient sequence, and equation (102) gives the integrable positive part required by Kingman's subadditive ergodic theorem. Therefore $\lim_{t\to\infty}\frac{1}{t}\log\|\widetilde{B}_t\cdots\widetilde{B}_1\| = \gamma_{L,n}$ almost surely, where ergodicity makes the limit constant rather than merely invariant. Under the hypothesis $\gamma_{L,n} > 0$, the limit is strictly positive. This is exactly equation (IA.14). □

Proof of Proposition IA.4. We prove the result by induction on t and simultaneously show that the empirical capital second moment $M_{K,t} = n^{-1}\sum_i (K_{i,t}^c)^2$ remains bounded in probability over the fixed horizon.

At $t = 0$, $Z_0 \to_p z_0$ by assumption. The initial capital fourth-moment bound implies in particular a uniform second-moment bound. Hence $\mathbb{E}[M_{K,-1}]$ is uniformly bounded, so Markov's inequality gives $M_{K,-1} = O_p(1)$. This establishes the induction base.

Suppose $Z_t \to_p z_t$ and $M_{K,t-1} = O_p(1)$. The contemporaneous panel variables are i.i.d. across units. Therefore the law of large numbers gives $s_t^{xx} \to_p V$, $s_t^{\theta x} \to_p \sigma_\theta^2$, and $s_t^{x\varepsilon} \to_p 0$.

We next control $\zeta_t = n^{-1} \sum_i w_{i,t}^c K_{i,t-1}^c$. Conditional on $\mathcal{F}_{t-1}$, the capital vector is fixed and the innovations $w_{i,t}$ have conditional mean zero. Demeaning does not change the weighted sum because $\sum_i K_{i,t-1}^c = 0$, so $\sum_i w_{i,t}^c K_{i,t-1}^c = \sum_i w_{i,t} K_{i,t-1}^c$. Hence $\mathbb{E}[\zeta_t \mid \mathcal{F}_{t-1}] = 0$. Conditional independence across units gives $\mathrm{Var}(\zeta_t \mid \mathcal{F}_{t-1}) = \frac{\sigma_w^2}{n^2} \sum_i (K_{i,t-1}^c)^2 = \frac{\sigma_w^2}{n} M_{K,t-1}$. Since $M_{K,t-1} = O_p(1)$, this conditional variance converges to zero in probability. For any $\epsilon > 0$ and any $C > 0$, conditional Chebyshev gives $\Pr(|\zeta_t| > \epsilon, M_{K,t-1} \le C) \le \sigma_w^2 C/(n\epsilon^2)$. Choose C so that $\Pr(M_{K,t-1} > C)$ is arbitrarily small. Then let $n \to \infty$. This proves $\zeta_t \to_p 0$.

The same argument applies to $\iota_t = n^{-1} \sum_i \eta_{i,t}^c K_{i,t-1}^c$. Conditional on the past, the measurement errors are independent across units with mean zero and variance σ_η^2, so $\mathrm{Var}(\iota_t \mid \mathcal{F}_{t-1}) = (\sigma_\eta^2/n) M_{K,t-1}$ and therefore $\iota_t \to_p 0$.

Equation (87) is $\nu_t = [s_t^{x\theta} + s_t^{x\varepsilon} + \beta\delta(\zeta_t + \iota_t)]/s_t^{xx}$. The numerator converges in probability to σ_θ^2 and the denominator converges to $V > 0$. Therefore $\nu_t \to_p \sigma_\theta^2/V = \rho$.

It follows entry by entry from equation (89) that $A_t \to_p \overline{A}$, and the forcing vector satisfies $\xi_t = (\nu_t, \delta\zeta_t)' \to_p (\rho, 0)'$. The exact recursion is $Z_{t+1} = A_t Z_t + \xi_t$. Since $A_t \to_p \overline{A}$ and $Z_t \to_p z_t$, matrix multiplication is continuous and $A_t Z_t \to_p \overline{A} z_t$. Adding the forcing limit gives $Z_{t+1} \to_p \overline{A} z_t + (\rho, 0)' = z_{t+1}$.

It remains to propagate the capital-moment bound. The centered capital law is $K_{i,t}^c = \delta K_{i,t-1}^c + \gamma\lambda_t \widetilde{x}_{i,t}^c$. Using $(a+b)^2 \le 2a^2 + 2b^2$ gives $(K_{i,t}^c)^2 \le 2\delta^2 (K_{i,t-1}^c)^2 + 2\gamma^2\lambda_t^2 (\widetilde{x}_{i,t}^c)^2$. Averaging across units yields $M_{K,t} \le 2\delta^2 M_{K,t-1} + 2\gamma^2\lambda_t^2 s_t^{xx}$. By the induction hypothesis $M_{K,t-1} = O_p(1)$. Since $Z_t \to_p z_t$, $\lambda_t = O_p(1)$. Also $s_t^{xx} \to_p V$, so $s_t^{xx} = O_p(1)$. Therefore the right-hand side is $O_p(1)$, and $M_{K,t} = O_p(1)$.

This completes the induction for every $t \le T$. For each fixed t, $\|Z_t - z_t\| \to_p 0$. Since the horizon contains only finitely many dates, for any $\epsilon > 0$, $\Pr(\max_{0 \le t \le T} \|Z_t - z_t\| > \epsilon) \le \sum_{t=0}^T \Pr(\|Z_t - z_t\| > \epsilon) \to 0$. This proves equation (IA.17). □

Proof of Proposition IA.5. The OLS decomposition is algebraic and does not require Gaussianity. The projection error is $u_{i,t} = (1-\rho)\widetilde{\theta}_{i,t} - \rho\eta_{i,t} + \varepsilon_{i,t}$. Independence and centering of the primitives give $\mathbb{E}[u_{i,t}] = 0$ and $\mathrm{Cov}(\widetilde{x}_{i,t}, u_{i,t}) = (1-\rho)\sigma_\theta^2 - \rho\sigma_\eta^2 = 0$. This is a linear projection identity and does not require $\mathbb{E}[u_{i,t} \mid \widetilde{x}_{i,t}] = 0$.

Within period t, centering the outcome gives $y_{i,t} - \overline{y}_t = (\rho + k\lambda_t)(\widetilde{x}_{i,t} - \overline{x}_t) + (u_{i,t} - \overline{u}_t)$. The centered features sum to zero, so the OLS numerator is $(\rho + k\lambda_t) S_t + N_t$. Dividing by $S_t > 0$ gives $\lambda_{t+1} = \rho + k\lambda_t + U_t$, with $U_t = N_t/S_t$.

Each U_t is a measurable function only of the period-t primitive draws. Fresh i.i.d. cross-sections therefore make the innovations i.i.d. and independent of the past. Their mean need not be zero because the projection error need not have conditional mean zero given the design. In particular, zero unconditional covariance between the feature and projection error does not imply that the ratio N_t/S_t is centered.

The assumptions imply a finite second moment, not just a finite mean. Cauchy–Schwarz for the finite cross-sectional sum gives $N_t^2 \le S_t \sum_i u_{i,t}^2$. Hence $U_t^2 \le S_t^{-1} \sum_i u_{i,t}^2$. Applying Cauchy–Schwarz to the expectation yields $\mathbb{E}[U_t^2] \le \{\mathbb{E}[(\sum_i u_{i,t}^2)^2]\mathbb{E}[S_t^{-2}]\}^{1/2} < \infty$. The first factor is finite because each $u_{i,t}$ is a linear combination of primitives with finite fourth moments, and the second is finite by assumption. Consequently $b_n = \mathbb{E}[U_t]$ exists. Setting $\widetilde{U}_t = U_t - b_n$ gives centered square-integrable innovations and proves (IA.19).

Part (i). Suppose $|k| < 1$. Write $a = \rho + b_n$ and $\lambda^\dagger = a/(1-k)$. Since $a + k\lambda^\dagger = \lambda^\dagger$, the centered recursion is $\lambda_{t+1} - \lambda^\dagger = k(\lambda_t - \lambda^\dagger) + \widetilde{U}_t$. On a two-sided innovation sequence, define $\lambda_t^* = \lambda^\dagger + \sum_{j=0}^\infty k^j \widetilde{U}_{t-1-j}$. The

bound $\mathbb{E}[\sum_{j=0}^{\infty}|k|^j|\widetilde{U}_{t-1-j}|] = \mathbb{E}|\widetilde{U}_0|/(1-|k|) < \infty$ shows that the series converges absolutely almost surely. It also converges in L^2, since the sum of the L^2 norms is at most $\|\widetilde{U}_0\|_2/(1-|k|)$. Shifting the series verifies the recursion. Its dependence only on past innovations makes it causal, stationary, and ergodic.

To establish uniqueness without imposing moments on a candidate stationary solution, iterate any stationary recursion backward by J dates. This gives $\lambda_t - \lambda^\dagger = k^J(\lambda_{t-J} - \lambda^\dagger) + \sum_{j=0}^{J-1} k^j \widetilde{U}_{t-1-j}$. Stationarity makes $\lambda_{t-J} - \lambda^\dagger$ a tight family, so the first term tends to zero in probability. The second term converges to the series just constructed. Thus every stationary solution has the same representation. Equivalently, any invariant law for the causal transition kernel must be this law. Coupling a process from any finite nonanticipative initial coefficient with the stationary solution and using the same future innovations gives the difference $k^t(\lambda_0 - \lambda_0^*)$, which tends to zero almost surely.

Absolute integrability allows expectations to pass through the stationary series. Since $\mathbb{E}[\widetilde{U}_t] = 0$, its mean is $\lambda^\dagger = (\rho + b_n)/(1-k)$. Independence and L^2 convergence give $\mathrm{Var}(\lambda_\infty) = \sum_{j=0}^{\infty} k^{2j}\mathrm{Var}(\widetilde{U}_t) = \mathrm{Var}(\widetilde{U}_t)/(1-k^2)$.

Part (ii). Suppose also that $v_U = \mathrm{Var}(\widetilde{U}_t) > 0$, and let $\mu_\lambda = \mathbb{E}[\lambda_t]$ and $v_\lambda = \mathrm{Var}(\lambda_t) = v_U/(1-k^2) > 0$. Write $\overline{\lambda}_T = T^{-1}\sum_{t=1}^{T}\lambda_t$ and $\overline{\lambda}_{+,T} = T^{-1}\sum_{t=1}^{T}\lambda_{t+1}$. Substitution of the archive recursion into OLS gives $\widehat{k}_T - k = \sum_{t=1}^{T}(\lambda_t - \overline{\lambda}_T)\widetilde{U}_t / \sum_{t=1}^{T}(\lambda_t - \overline{\lambda}_T)^2$. The stationary archive and innovation pairs are ergodic functions of the i.i.d. innovation sequence. The ergodic theorem therefore gives $T^{-1}\sum_{t=1}^{T}(\lambda_t - \overline{\lambda}_T)^2 \to v_\lambda$ almost surely. Applying it to $\lambda_t\widetilde{U}_t$ and $\widetilde{U}_t$, whose required first moments are finite, makes the numerator divided by T converge almost surely to zero. Hence $\widehat{k}_T \to k$. The intercept estimator is $\widehat{a}_T = \overline{\lambda}_{+,T} - \widehat{k}_T\overline{\lambda}_T$. Both sample means converge to μ_λ, so $\widehat{a}_T \to (1-k)\mu_\lambda = a$ almost surely.

For asymptotic normality, decompose the slope numerator as $\sum_{t=1}^{T}(\lambda_t - \overline{\lambda}_T)\widetilde{U}_t = \sum_{t=1}^{T}(\lambda_t - \mu_\lambda)\widetilde{U}_t - (\overline{\lambda}_T - \mu_\lambda)\sum_{t=1}^{T}\widetilde{U}_t$. Let $\mathcal{F}_t = \sigma(\lambda_1, \widetilde{U}_1, \ldots, \widetilde{U}_{t-1})$. Then $M_t = (\lambda_t - \mu_\lambda)\widetilde{U}_t$ is $\mathcal{F}_{t+1}$-measurable and satisfies $\mathbb{E}[M_t \mid \mathcal{F}_t] = 0$. Its conditional variance is $v_U(\lambda_t - \mu_\lambda)^2$. Ergodicity gives $T^{-1}\sum_{t=1}^{T}\mathbb{E}[M_t^2 \mid \mathcal{F}_t] \to v_U v_\lambda$ almost surely.

The conditional Lindeberg condition follows from the second moments already established. Independence of λ_t and $\widetilde{U}_t$ gives $\mathbb{E}[M_t^2] = v_U v_\lambda < \infty$. For any $\epsilon > 0$, consider the conditional sum $T^{-1}\sum_{t=1}^{T}\mathbb{E}[M_t^2\mathbb{1}\{|M_t| > \epsilon\sqrt{T}\} \mid \mathcal{F}_t]$. Its expectation equals $\mathbb{E}[M_1^2\mathbb{1}\{|M_1| > \epsilon\sqrt{T}\}]$ by stationarity and tends to zero by dominated convergence. Markov's inequality then makes the conditional sum converge to zero in probability. The martingale central limit theorem yields $T^{-1/2}\sum_{t=1}^{T}M_t \Rightarrow N(0, v_U v_\lambda)$.

The correction from demeaning is negligible after division by $\sqrt{T}$. Ergodicity gives $\overline{\lambda}_T - \mu_\lambda = o_p(1)$, while the i.i.d. central limit theorem gives $T^{-1/2}\sum_{t=1}^{T}\widetilde{U}_t = O_p(1)$. Their product is $o_p(1)$. Dividing by the OLS denominator and applying Slutsky's theorem therefore gives $\sqrt{T}(\widehat{k}_T - k) \Rightarrow N(0, v_U/v_\lambda) = N(0, 1-k^2)$. No moment above order two is required for this argument.

Part (iii). Suppose $|k| > 1$ and let $\widetilde{\lambda} = (\rho + b_n)/(1-k)$. Subtracting this fixed point and iterating gives $\lambda_t - \widetilde{\lambda} = k^t[\lambda_0 - \widetilde{\lambda} + \sum_{j=0}^{t-1}k^{-j-1}\widetilde{U}_j]$. Because $\mathbb{E}|\widetilde{U}_0| < \infty$ and $\sum_{j=0}^{\infty}|k|^{-j-1} < \infty$, the weighted series converges absolutely almost surely to $M_\infty = \sum_{j=0}^{\infty}k^{-j-1}\widetilde{U}_j$.

It remains to rule out exact cancellation. Write the innovation law as $pQ + (1-p)R$, where $p > 0$ and Q is atomless. Realize each innovation by an independent mixture label followed by a draw from the selected component. Almost surely at least one innovation is drawn from Q, because the probability of no such draw in the first J periods is $(1-p)^J \to 0$. Let J be the first index with a Q draw. Conditional on the mixture labels, the initial coefficient, and every innovation except $\widetilde{U}_J$, the limit $\lambda_0 - \widetilde{\lambda} + M_\infty$ is a fixed finite quantity plus $k^{-J-1}\widetilde{U}_J$. The coefficient is nonzero, and the remaining draw has the atomless law Q. Nonanticipativity ensures that conditioning on the initial coefficient does not change this draw's law. Thus the conditional probability of exact cancellation is zero, and integration gives $\Pr(\lambda_0 - \widetilde{\lambda} + M_\infty = 0) = 0$.

Consequently $(\lambda_t - \widetilde{\lambda})/k^t \to C = \lambda_0 - \widetilde{\lambda} + M_\infty$ almost surely, with $0 < |C| < \infty$. It follows that $|\lambda_t - \widetilde{\lambda}|/|k|^t \to |C|$ and $|\lambda_t| \to \infty$. If $k < -1$, the sign of the bracket is eventually the sign of C, while the sign of k^t alternates. This proves the alternating geometric divergence. □

IA.6.3 Lemmas

Proof of Lemma IA.1. Fix a realization with $S_t > 0$ and write $q = Q'x_t^c \in \mathbb{R}^{n-1}$. Since $x_t^c \in \mathbb{1}^\perp$ and QQ' is the orthogonal projection onto $\mathbb{1}^\perp$, $q'q = (x_t^c)'QQ'x_t^c = (x_t^c)'x_t^c = S_t$.

Consider first any $v \in \mathbb{R}^{n-1}$ satisfying $q'v = 0$. Apply $\widetilde{B}_t$ to the vector $(0, v')'$. The first coordinate is $(\beta\delta/S_t)(x_t^c)'Qv = (\beta\delta/S_t)q'v = 0$. The capital coordinate is δv. Hence $\widetilde{B}_t(0, v')' = \delta(0, v')'$. The subspace orthogonal to q has dimension $n-2$, so δ is an eigenvalue with multiplicity at least $n-2$.

The orthogonal complement of those eigenvectors is the two-dimensional subspace spanned by $e_1 = (1, 0, \ldots, 0)'$ and $u = (0, q'/\sqrt{S_t})'$. We compute the matrix representation on this basis. First,

$$\widetilde{B}_t e_1 = \begin{pmatrix} k \\ \gamma q \end{pmatrix} = k e_1 + \gamma\sqrt{S_t}\, u.$$

Second,

$$\widetilde{B}_t u = \begin{pmatrix} (\beta\delta/S_t)(x_t^c)'Q(q/\sqrt{S_t}) \\ \delta q/\sqrt{S_t} \end{pmatrix}.$$

Since $(x_t^c)'Qq = q'q = S_t$, the first coordinate is $\beta\delta/\sqrt{S_t}$, and the second block is exactly δ times the capital component of u. Therefore $\widetilde{B}_t u = (\beta\delta/\sqrt{S_t})e_1 + \delta u$. The two-dimensional block is

$$C_t = \begin{pmatrix} k & \beta\delta/\sqrt{S_t} \\ \gamma\sqrt{S_t} & \delta \end{pmatrix}.$$

Its characteristic polynomial is $(z-k)(z-\delta) - \beta\gamma\delta = z^2 - (k+\delta)z + k\delta - k\delta = z[z - (k+\delta)]$, since $\beta\gamma = k$. Hence its eigenvalues are 0 and $k+\delta$.

Together with the $n-2$ eigenvalues equal to δ, the full characteristic polynomial is $z(z-k-\delta)(z-\delta)^{n-2}$, proving equation (IA.13) and the stated spectrum. □

Proof of Lemma IA.2. Fix n and m, and let $P_j = \widetilde{B}_{jm} \cdots \widetilde{B}_{(j-1)m+1}$. The assumed integrability of $\log \|P_1\|_n$ and stationarity imply that every block norm is strictly positive almost surely. Condition (IA.15) therefore permits taking logarithms and gives $\log \|P_r \cdots P_1\|_n \geq \sum_{j=1}^r \log \|P_j\|_n + (r-1)\log c$ for every integer $r \geq 2$. Dividing by rm yields $(rm)^{-1} \log \|P_r \cdots P_1\|_n \geq m^{-1}[r^{-1}\sum_{j=1}^r \log \|P_j\|_n] + (r-1)(rm)^{-1}\log c$.

Under the maintained Gaussian panel assumptions, the coefficient sequence is stationary and geometrically β-mixing. Forming nonoverlapping blocks of fixed length preserves stationarity and geometric mixing, and taking the matrix product within each block is a measurable transformation. Thus $\{P_j\}$ is stationary and ergodic. The assumption $\mathbb{E}|\log \|P_1\|_n| < \infty$ allows the ergodic theorem to give $r^{-1}\sum_{j=1}^r \log \|P_j\|_n \to \mathbb{E}\log \|P_1\|_n$ almost surely.

The product $P_r \cdots P_1$ equals $\widetilde{B}_{rm} \cdots \widetilde{B}_1$. Its normalized log norm therefore converges almost surely to $\gamma_{L,n}$ along the subsequence of dates rm. The norm $\|\cdot\|_n$ gives the same exponent because n is fixed and equivalent norms have the same asymptotic growth rate. Since $(r-1)/(rm) \to 1/m$, taking limits in the preceding inequality gives $\gamma_{L,n} \geq m^{-1}\mathbb{E}\log \|P_1\|_n + m^{-1}\log c$, proving (IA.16).

Finally, suppose $m^{-1}\mathbb{E}\log \|P_1\|_n > \epsilon > 0$. If $c > e^{-m\epsilon}$, we also have $m^{-1}\log c > -\epsilon$. The lower bound in (IA.16) is consequently strictly positive, so $\gamma_{L,n} > 0$. □

IA.6.4 Corollaries

Proof of Corollary IA.1. Let $\alpha = n-1$. Theorem 6 gives $U_t = c_n T_{n-1}$, where $c_n = \sigma_u/\sqrt{V(n-1)} > 0$. The Student distribution has regularly varying tails with index α, and $\mathbb{E}|U_t|^p < \infty$ holds exactly when $0 < p < \alpha$.

By Theorem 7(i), the stationary coefficient satisfies $\lambda_\infty = \lambda^* + \sum_{j=0}^{\infty} k^j U_{-j}$. If $k = 0$, its tail index is that of U_0. Suppose $0 < |k| < 1$. For every $\epsilon \in (0, \alpha)$,

$$\sum_{j=0}^{\infty} |k^j|^{\alpha-\epsilon} = \sum_{j=0}^{\infty} |k|^{j(\alpha-\epsilon)} = \frac{1}{1-|k|^{\alpha-\epsilon}} < \infty. \tag{IA.23}$$

The coefficients are also absolutely summable. The weighted-sum theorem for independent regularly varying random variables therefore applies to the symmetric Student innovations and implies that $\sum_{j=0}^{\infty} k^j U_{-j}$ has regularly varying tails with index α (Embrechts et al., 1997, Appendix A3.3). Adding the constant λ^* does not change this index. Thus λ_∞ has tail index $n-1$.

We establish the moment statement separately, including the boundary $p = \alpha$. Suppose first that $0 < p < \alpha$. If $p \le 1$, subadditivity and Fatou's lemma give $\mathbb{E}|\lambda_\infty|^p \le |\lambda^*|^p + \mathbb{E}|U_0|^p \sum_{j=0}^{\infty} |k|^{jp} < \infty$. If $p > 1$, Minkowski's inequality bounds the L^p norm of every finite partial sum by $|\lambda^*| + \|U_0\|_p \sum_{j=0}^{\infty} |k|^j < \infty$. Fatou's lemma then gives $\mathbb{E}|\lambda_\infty|^p < \infty$.

Now suppose $p \ge \alpha$. Write $\lambda_\infty = U_0 + R$, where $R = \lambda^* + \sum_{j=1}^{\infty} k^j U_{-j}$. The stationary-series construction makes R finite almost surely and independent of U_0. For any finite r, the inequality $|U_0 + r| \ge |U_0|/2$ holds whenever $|U_0| > 2|r|$. Therefore $\mathbb{E}|U_0 + r|^p \ge 2^{-p}\mathbb{E}[|U_0|^p \mathbf{1}_{\{|U_0|>2|r|\}}] = \infty$, because the Student innovation has an infinite pth absolute moment. Conditioning on R and applying Tonelli's theorem yields $\mathbb{E}|\lambda_\infty|^p = \infty$. Consequently, for every $p > 0$, the stationary coefficient has a finite pth absolute moment if and only if $p < n - 1$, proving (IA.18). □

Proof of Corollary IA.2. Let $v_U = \mathrm{Var}(U_t) = \sigma_u^2/[V(n-3)] > 0$. For $0 \le k < 1$, the stationary representation gives $\mathrm{Var}(\lambda_\infty) = v_U \sum_{j=0}^{\infty} k^{2j} = v_U/(1-k^2)$. Taking the positive square root yields (IA.20). If $0 \le k_1 < k_2 < 1$, then $1 - k_2^2 < 1 - k_1^2$, so the standard deviation is strictly larger at k_2. It diverges as $k \uparrow 1$ because v_U is held fixed and positive.

For completeness, independence of U_t and λ_t gives $\mathrm{Cov}(\lambda_{t+1}, \lambda_t) = k\mathrm{Var}(\lambda_\infty)$. Stationarity therefore implies $\mathrm{Var}(\lambda_{t+1} - \lambda_t) = 2(1-k)\mathrm{Var}(\lambda_\infty) = 2v_U/(1+k)$, which proves (IA.21). Its derivative with respect to k is $-2v_U/(1+k)^2 < 0$. □