"The 11-month detection peak could systematically reflect electoral timing rather than learning dynamics, as federal enforcement often intensifies before elections as administrations demonstrate anti-corruption credentials, then declines post-election as priorities shift (Gordon & Huber, 2007)."

"The reverse causality problem requires immediate attention because successful corruption prosecutions generate political pressure for continued resource allocation, media attention that sustains legislative interest, and institutional momentum that affects future detection patterns (Carpenter, 2001)."

"FBI appropriations are determined through congressional processes that explicitly consider agency performance metrics, including corruption prosecution statistics (Wildavsky & Caiden, 1989)"

"The FBI budget measure using proportional differences between enacted and requested appropriations is problematic because it conflates multiple distinct mechanisms: congressional priorities, executive branch requests, and budget negotiation dynamics, while being applied at the jurisdiction level when budget decisions occur nationally (Fenno, 1966)."

LEARNING FROM PROBLEM TRACES: HOW PROXY METRICS RESHAPE FEEDBACK, ATTENTION, AND RESOURCE CYCLES

ABSTRACT

Organizations learn from performance feedback, but in problem-trace domains—where metrics like safety incidents or corruption detections proxy for the underlying goal of reduction—this creates a signal–proxy problem. We theorize two rival learning logics. A deterrence logic suggests detections inform about latent opportunities, and resources build capability to suppress them, yielding a steeper decline in detections after a peak. A proxy-locked logic, rooted in behavioral and institutional theory, posits that metrics become yardsticks for performance, causing organizations to accelerate activity until the metric improves, then pivot attention, shifting the peak earlier without suppressing the underlying problem. Analyzing a monthly panel of U.S. anti-corruption detections, we find that resources advance the timing of saturation peaks but do not steepen post-peak declines, while legislative attention sharpens near-term pullbacks without altering downturn curvature. These results support the proxy-locked dynamic, revealing a temporal signature whereby feedback ceases to inform mission progress. We advance theory by specifying how proxy error distorts learning and provide a field-based discriminator for adjudicating between genuine capability and metric reactivity.

INTRODUCTION

Organizations adapt by interpreting signals from their environment. Performance feedback theory conceptualizes this process as one of comparing outcomes to aspirations and adjusting behavior when gaps arise (Cyert & March, 1963; Greve, 2003). However noisy or multidimensional they may be, signals are often presumed to point back to the outcomes they represent. Research has extended this logic to contexts where outcomes exceed aspirations on one dimension while falling short on another (Audia & Brion, 2007), or where historical and social comparisons diverge (Joseph & Gaba, 2015). Such tensions complicate evaluation and trigger cycles of shifting attention, recalibration, and sensemaking (Audia & Greve, 2006; Jordan & Audia, 2012; Blagoeva, Mom, & Jansen, 2020; Chambers, Alves, & Aceves, 2024). Across these accounts, signals remain tethered to what they represent.

Yet, in many consequential domains, this assumption breaks down. Oversight, compliance, and safety rely on *trace metrics*: counts of organizational contact with a latent harm process, such as accidents (Perrow, 1984; Vaughan, 1996; Hopkins, 2000), detected violations (Gray & Silbey, 2014; Huising & Silbey, 2011), or corruption prosecutions (Power, 1997; Bevan & Hood, 2006). Trace metrics invert the usual signal–goal relation: rising counts may reflect intensified vigilance rather than a worsening condition, while falling counts may indicate declining vigilance rather than genuine progress (Power, 1997; Bevan & Hood, 2006; Merry, 2011). Their level reflects both the underlying opportunity field and the organization's own vigilance. When organizations adapt to such signals, it becomes unclear whether they are suppressing harm or merely managing its representation (March & Sutton, 1997; Espeland & Stevens, 1998). This problem, long recognized in sociology and political science (Goodhart, 1975; Strathern, 1997; Espeland & Sauder, 2007), has yet to be specified as a dynamic learning process.

We theorize learning under endogenous signal invalidity. Two perspectives illustrate the problem. A deterrence–capability logic, rooted in economics and criminology, treats trace metrics as noisy but ultimately informative (Becker, 1968; Nagin, 2013). Here, additional resources expand monitoring capacity: counts initially rise with vigilance but then fall as opportunities diminish, with resources deepening and prolonging decline. By contrast, a reactivity–attention logic, grounded in bounded rationality and institutional theories of metrics (Simon, 1971; Ocasio, 1997; Espeland & Sauder, 2007), emphasizes how organizations become locked into managing what is most visible. Vigilance escalates until the metric improves, then shifts elsewhere, producing inverted-U detection curves that reflect attention cycles rather than problem suppression. In this view, greater resources accelerate the turning point instead of extending the decline.

We examine these dynamics in U.S. federal corruption enforcement, a domain where the underlying phenomenon is latent and trace metrics are the primary yardstick of oversight. Using a monthly panel of prosecutions across fifty-four jurisdictions over two decades, linked to resource appropriations and political attention, we identify three patterns. First, detections display a consistent inverted-U relationship with their own lags. Second, additional resources shift the turning point earlier rather than deepening the decline. Third, legislative attention sharpens short-term responsiveness without altering long-run collapse. Taken together, these findings are consistent with reactivity–attention and difficult to reconcile with deterrence–capability.

This study contributes to extant research in three ways. First, we show that temporal horizon is not background but constitutive of oversight dynamics. Research on deterrence, enforcement, and monitoring has typically treated time as a neutral backdrop (Becker 1968; Stigler 1970; Polinsky & Shavell 2007). Our findings show that this assumption conceals the most consequential dynamic: short-term and long-term detection trajectories are not only different in degree but qualitatively different in form. In the short term, detections follow a persistence logic, consistent with deterrence–capability models. In the long term, however, detections exhibit a robust inverted-U pattern, initially rising but ultimately collapsing. This shift suggest that theories of oversight cannot be evaluated independently of temporal scope. By embedding horizon directly into theory, we extend deterrence and organizational monitoring research to show that time itself is a structuring force in how oversight functions and fails. This contribution is foundational: it changes not only how oversight is measured but how its very mechanisms are theorized.

Second, we reconcile competing perspectives on the drivers of oversight effectiveness by specifying when each applies. Deterrence-capability logics posit that investment in monitoring capacity yields sustained suppression of misconduct (e.g., Ellickson 1991; Kelling & Coles 1996), whereas behavioral and organizational accounts emphasize attention fatigue, adaptive drift, and the erosion of vigilance (March 1991; Ocasio 1997; Vaughan 1999). Our results demonstrate that both accounts are partially correct, but along different horizons: deterrence explains the short-term rise in detections, while behavioral saturation processes explain the long-term collapse. By anchoring deterrence-capability and reactivity-attention in their respective temporal domains, we reconcile two literatures that have often spoken past one another. This contribution strengthens both perspectives: deterrence gains realism by being bounded, and behavioral logics gain specificity by being temporally anchored.

Third, we reframe corruption oversight (and by extension compliance, safety, and risk regulation) as endogenous learning systems that collapse from within rather than failures imposed from without. Existing accounts often attribute enforcement decline to exogenous shocks (e.g., budget cuts, political constraints, or episodic waves of misconduct; Reuter & Truman 2004; Glaeser & Shleifer 2003). By contrast, we show that collapse is seeded by the very feedback processes that initially generate detection success. Short-term attention to cases and reinforcement of monitoring routines accumulates into long-term saturation, undermining capacity from within. This contribution extends theories of corruption oversight and related boundary contexts, such as compliance regimes, regulatory vigilance, and risk monitoring, by showing how adaptive success can destabilize itself over time. In doing so, we redirect the debate from capacity and resource scarcity to the internal, path-dependent dynamics of oversight systems, highlighting why corruption control proves so difficult to sustain.

THEORY

Performance feedback has become the dominant lens for explaining organizational adaptation (Cyert & March, 1963; Levitt & March, 1988). When outcomes fall below aspirations, organizations search; when outcomes exceed aspirations, they exploit existing routines (Greve, 2003; Lant & Mezias, 1992). Building on this foundation, scholars have extended feedback theory to diverse domains, from innovation search (Posen et al., 2018) to risk taking (Baum & Dahlin, 2007) to competitive interaction (Greve, 1998; Baum, Rowley, & Shipilov, 2005). This work has refined how organizations respond to ambiguity and multidimensionality in signals—showing, for instance, that performance can simultaneously exceed historical benchmarks while lagging social comparisons (Audia & Brion, 2007; Joseph & Gaba, 2015), or that conflicting signals can trigger cycles of recalibration and sensemaking (Jordan & Audia, 2012; Blagoeva, Mom, & Jansen, 2020). Recent research even suggests that signal contestation can sharpen learning (Chambers, Alves, & Aceves, 2024).

Across these accounts, however, a core assumption persists: signals, however noisy, remain valid representations of underlying objectives (Argote & Greve, 2007). The challenge is interpretation, not misdirection. This is the blind spot our study addresses.

Invalid feedback as a learning problem

Organizational learning research typically assumes that performance signals, while noisy or multidimensional, remain tethered to the underlying outcomes they represent (Cyert & March, 1963; Greve, 2003; Argote & Greve, 2007). Even when feedback is ambiguous or conflicting—as when organizations score above aspirations on one dimension but below on another (Audia & Brion, 2007; Chambers, Alves, & Aceves, 2024)—the signals are presumed valid, and the task is one of interpretation.

Yet many consequential domains are governed not by ambiguous signals but by invalid ones. Oversight, compliance, and safety regimes often rely on problem-trace metrics such as accidents, violations, or corruption prosecutions—the very phenomena they are tasked with suppressing. Such measures invert conventional feedback logic: higher counts may reflect heightened vigilance rather than worsening conditions, while lower counts may indicate declining vigilance rather than genuine improvement (Power, 1997; Bevan & Hood, 2006; Merry, 2011). The challenge here is not reconciling multiple valid signals but interpreting a single indicator systematically decoupled from the ultimate objective. The yardstick itself is unreliable.

Hypothesis 1 (H1). *Detection trajectories in problem-trace domains exhibit an inverted-U relationship with past detections: current detections initially increase with past detection intensity, but eventually decrease after a threshold.*

The temporal dilemma of invalid feedback

Invalid feedback creates a recursive learning loop: the signal guiding adaptation is itself shaped by prior vigilance. Because problem-trace metrics are salient and state-dependent, organizations escalate attention until the signal improves, then reallocate effort once visible progress is achieved. This produces a dynamic of escalation and withdrawal—an inverted-U trajectory that mimics substantive improvement while leaving underlying conditions unresolved.

This logic extends theories of the myopia of learning (Levinthal & March, 1993), bounded attention (Ocasio, 1997), and institutional reactivity (Espeland & Sauder, 2007). It shows how signals that are endogenously invalid generate not just misinterpretation but cyclical dynamics—apparent progress in the short term that seeds decline in the long term. Crucially, this embeds temporal horizon into theory: at short lags, detections rise with vigilance, consistent with deterrence–capability logics; at longer horizons, vigilance collapses under its own feedback, consistent with reactivity–attention logics.

Integrating Deterrence and Attention

This framework bridges two perspectives often treated as oppositional. Deterrence–capability views emphasize how resources expand vigilance and suppress opportunities (Becker, 1968; Nagin, 2013). Reactivity–attention accounts, by contrast, stress how organizations chase visible indicators at the expense of underlying conditions (Simon, 1971; Ocasio, 1997; Espeland & Sauder, 2007). Our argument is that both logics hold—but on different horizons. Deterrence dominates in the short term, reactivity in the long term. Horizon thus becomes the key to integrating capability and behavioral perspectives and explaining why oversight systems succeed early but collapse later.

These competing predictions yield two rival hypotheses about how resources reshape detection trajectories:

Hypothesis 2a (H2a). *Higher resources will be associated with steeper post-peak declines in detections.*

Hypothesis 2b (H2b). *Higher resources will be associated with earlier peaks in detections but not with steeper post-peak declines.*

Political attention and the amplification of signals

Finally, political and legislative attention heightens the salience of problem-trace metrics. Prior research on agenda-setting and institutional attention shows how external scrutiny amplifies organizational responsiveness to visible signals (Kingdon, 1984; Baumgartner & Jones, 1993; Ocasio, 1997). In oversight domains, this implies that recent detections will weigh more heavily in shaping current vigilance when legislative attention is high. However, because attention does not resolve the underlying invalidity of the metric, its influence should remain short-lived, sharpening responsiveness in the near term without altering the long-run collapse.

Hypothesis 3 (H3). *Higher legislative attention sharpens short-term responsiveness in detections but does not alter the steepness of long-run declines.*

In sum, we theorize oversight as a case of organizational learning under invalid feedback. By distinguishing invalid from inconsistent signals, embedding temporal horizon into the analysis, and integrating deterrence-capability with reactivity-attention perspectives, we recast oversight collapse as an endogenous learning dynamic. This moves the debate beyond resource scarcity or external shocks to show how success undermines itself through the very feedback cycles meant to sustain it.

METHODS

Data

We assembled a monthly panel that links federal enforcement outputs to two institutional covariates. The outcome series captures observed corruption detections by jurisdiction and month using U.S. federal court filings. Institutional capacity is measured with FBI appropriations. Legislative attention is measured with congressional bill introductions retrieved from the Library of Congress. We merged these sources into a single jurisdiction–month panel and then applied the analysis pipeline described below.

Court filings. We obtained federal dockets from CourtListener and identified filings that include corruption-relevant statutory counts in six established categories: Foreign Corrupt Practices Act, federal bribery, kickbacks, fraud, the False Claims Act, and RICO. We restricted the data to cases in which the defendant is an organization so that detections reflect oversight of organizational actors rather than individuals. Each filing was time-stamped to the filing month, assigned to its federal judicial district, and aggregated to jurisdiction–months. Because detections are "samples of what surfaces," not a census of underlying misconduct, we retain zero months to preserve substantively informative non-events.

FBI appropriations. For each fiscal year between 2002 and 2024 we recorded requested and enacted amounts for the FBI's Salaries and Expenses line and constructed the proportional difference (enacted minus requested, divided by requested). Annual values were expanded to a monthly series to align with the detection counts prior to analysis.

Congressional bills. Legislative attention was assembled from the Library of Congress (Congress.gov). For each month from 2002 to 2024 we retrieved titles and summaries of all bill introductions across both chambers and applied a keyword lexicon capturing corruption-relevant terms (e.g., fraud, accountability, ethics, compliance, bribery, procurement, enforcement). The monthly count of introductions matching this lexicon serves as an indicator of agenda salience, recognizing that many such bills are symbolic rather than operational.[1] Figure 1 summarizes the three smoothed, raw series. The patterns motivate our lagged-feedback design but are not used for identification.

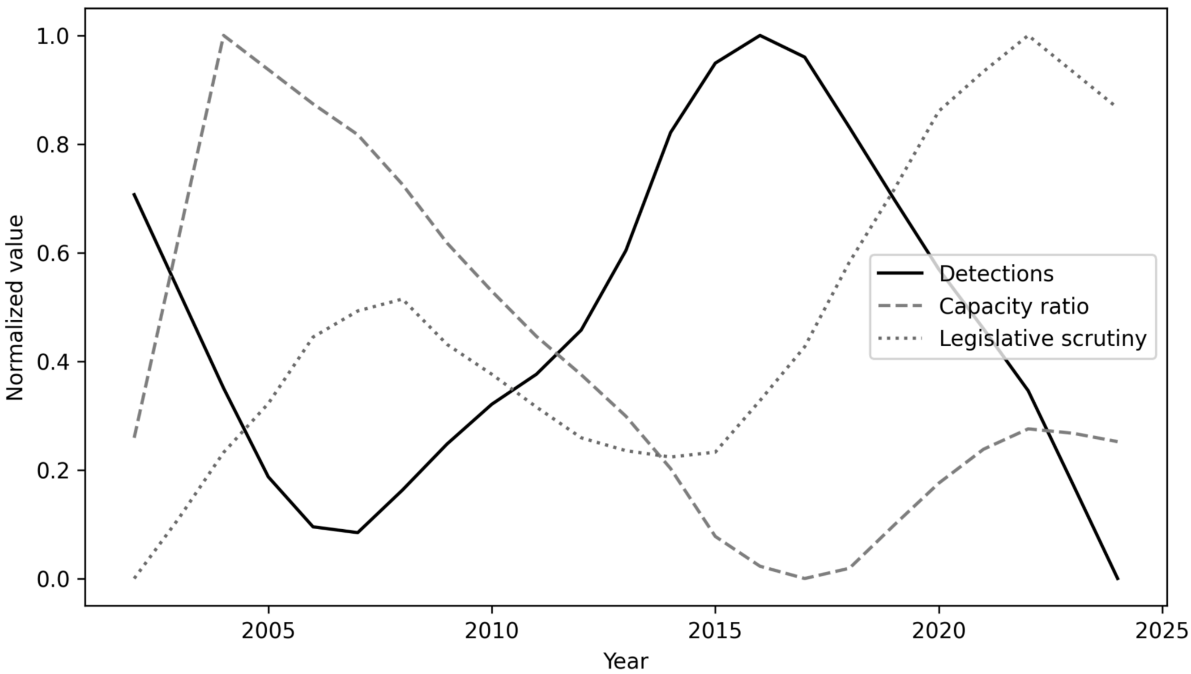
Panel Construction and Estimation

The unit of analysis is the jurisdiction–month. We first aggregated federal judicial districts to 54 federal jurisdictions (the 50 states, the District of Columbia, Puerto Rico, Guam, and the U.S. Virgin Islands). To ensure that each lag corresponds to exactly one prior calendar month, we re-indexed every jurisdiction to a complete monthly sequence and then computed calendar-true lags for k=1,…,12. We explicitly distinguish months with zero detected filings (valid observations that we retain) from months that are missing because the underlying docket information is unavailable or a jurisdiction's coverage starts later. Zeros are part of the constructed monthly series and can enter lag windows; truly missing months remain missing and are not imputed as zero. To estimate models that use the full lag history, we require a complete twelve-month lag window for each observation. Practically, this drops the first

[1] General Corruption/Integrity: corruption, bribery, fraud, integrity, ethics, transparency, misconduct; Corporate and Financial Misconduct: corporate fraud, securities fraud, accounting fraud, insider trading, market manipulation, racketeering, embezzlement, financial crime; Regulatory Oversight/Enforcement: whistleblower, oversight, audit reform, compliance, enforcement, accountability, Sarbanes-Oxley, Dodd-Frank; Conflict of Interest/Lobbying/Revolving Door: conflict of interest, lobbying, revolving door, undue influence, campaign finance, regulatory capture; Procurement and Public Sector Fraud: procurement fraud, contract fraud, kickbacks, false claims; Institutional Control Mechanisms: ethics commission, inspector general, public integrity, anti-corruption office, ombudsman, independent counsel; Foreign Corruption/Multinational Regulation: FCPA, foreign bribery, transnational corruption, cross-border compliance, OECD convention. We also include specific legal and statutory triggers such as Title 18 §§ 201, 666, 1341, 1343, 1951, 1961 (RICO), Foreign Corrupt Practices Act, and Honest Services Fraud. All terms were regex coded to ensure comprehensive matching, with multi-word phrases matched with spacing flexibility to capture variations in bill language.

twelve months after a jurisdiction's coverage begins and any month whose twelve-month window crosses a missing value. We also require non-missing values for the FBI budget series; for the legislative-moderation specification, we additionally require the transformed legislative series. These mechanically induced requirements, combined with continuous monthly coverage after reindexing, yield the final estimation sample of 3,045 territory–months across 32 jurisdictions.

FIGURE 1. Overall Data Trends



Notes. The figure represents the locally weighted scatterplot smoothed (i.e., LOWESS) annual totals for the main variables of interest. The three functions are normalized for scale. The solid line highlights a mid-2010s hump that later declines. The dashed line represents the evolving capacity ratio, calculated as (enacted budget – requested budget)/requested budget, aggregated by year. Positive values represent over-funding periods; negative values, periods experiencing cuts. The pattern mirrors the rise-and-fall pattern seen with detections, albeit at an earlier stage. The dotted line represents the evolution of legislative scrutiny, which rises during periods of capacity increases, but peaks after detection plateaus.

Measurements and Transformations

The dependent variable is the monthly count of organizational corruption detections per jurisdiction. For each lag $k \in \{1,\ldots,12\}$ we transform the lagged count as $z_k = \log(1+\text{detections}_{t-k})$, then center z_k within jurisdiction and include the centered linear term and its square. This within-jurisdiction centering ensures that lag effects are identified off deviations from a jurisdiction's own baseline rather than cross-sectional differences.

The FBI budget measure is standardized to mean zero and unit variance before forming interactions, due to its construction. The legislative series is constructed as a national log(1+x) transformation that is mean-centered by month; because month×year fixed effects absorb national month-specific shocks, the legislative series enters only through interactions with lag terms in the legislative-moderation model (see below).

Modeling Strategy

We estimate negative binomial, generalized linear models to accommodate overdispersed counts (Hausman, Hall, & Griliches, 1984). All specifications include jurisdiction fixed effects and month×year fixed effects, absorbing time-invariant jurisdiction heterogeneity and common month-specific shocks. Standard errors are clustered by jurisdiction. The modeling sequence follows four nested specifications: Model A includes the twelve centered lag terms z_k, Model B adds the twelve squared terms z_k^2 to allow inverted-U relationships, testing Hypothesis 1. Model C adds budget moderation by interacting the standardized budget measure with both z_k and z_k^2, which tests whether resources reshape the curve's anatomy in line with Hypotheses 2a–2b. Model D adds legislative moderation by interacting the centered legislative series with z_k and z_k^2.

Because the legislative series is national and month×year fixed effects are saturated, we identify legislative moderation via orthogonalized interactions: each interaction is residualized against the full right-hand side of Model C (including all budget interactions and fixed effects) so that the resulting terms capture only variation not already explained by resources, lag structure, or common month shocks. Appendix XXX reports the steps of the orthogonalization approach.

Turning Point Computation

For lags where the inverted-U is active in the data, the turning point is computed as

$$z^* = -\frac{\beta_1 + B\gamma_1}{2(\beta_2 + B\gamma_2)}$$

where β_1 and β_2 are the linear and quadratic coefficients on the centered log(1+detections$_{t-k}$) term for lag k, and γ_1, γ_2 are their budget interactions. B denotes the standardized budget variable used in estimation (mean 0, SD 1). We compute SEs and CIs for z^* via the delta method and map it back to an implied count by adding the median jurisdiction's lag-mean in log space and exponentiating.

Summary Statistics and Correlations

Table 1 reports summary statistics and pairwise correlations for the estimation sample of 3,045 jurisdiction–months across 32 jurisdictions. Detections and legislative attention are expressed as centered log transforms of monthly counts due to overdispersion, so their means are near zero by construction (Detections mean = 0.074, SD = 0.558; Legislative attention mean = −0.015, SD = 0.568). Budget is the raw proportional difference between enacted and requested FBI appropriations, presented in its original units for interpretability (mean = 0.02, SD = 0.098) and standardized only in the interaction terms used in the regression models.

TABLE 1. Summary Statistics and Correlation Matrix

	Mean	Standard Deviation	(1)	(2)	(3)
(1) Detections	0.074	0.558	1.000	—	—
(2) Budget	0.020	0.098	-0.204*	1.000	—
(3) Legislative Attention	-0.0154	0.568	-0.005	-0.043*	1.000

Notes. Initial assembled panel: 8,926 jurisdiction–months across 54 jurisdictions (2002–2024). Final estimation sample (calendar-true lags and complete covariates): 3,045 jurisdiction–months across 32 jurisdictions. "Detections" and "Legislative attention" are centered log(1+ ·)

measures; “Budget” is the raw proportional difference (Enacted−Requested)/Requested. Stars indicate $p < .05$ for two-sided Pearson tests on the estimation sample. Correlations are descriptive and combine within- and between-jurisdiction variation; the regression models identify effects from within-jurisdiction dynamics with territory and month×year fixed effects.

The correlations are modest and behave as expected given our design. Detections is negatively correlated with Budget ($r = -0.204$, $p < .05$), while its correlation with Legislative attention is near zero ($r = -0.005$, n.s.); Budget and Legislative attention are weakly negatively correlated ($r = -0.043$, $p < .05$). Because these are pooled associations, they mix within and between jurisdiction variation and do not speak to dynamic learning; their main value is diagnostic. First, the centered transforms behave as intended and indicate meaningful dispersion in detections, consistent with using a negative binomial specification. Second, the small correlations among regressors reduce concerns about collinearity in the polynomial and moderation terms that follow. Third, the absence of a strong contemporaneous association between detections and national legislative attention aligns with our identification strategy, which relies on within jurisdiction dynamics, calendar-true lags, fixed effects, and orthogonalized interactions rather than on pooled cross-sectional patterns. Taken together, the table increases confidence that the subsequent tests will be driven by the temporal anatomy of detection curves rather than by mechanical covariation in levels.

Results

Table 2 reports a parsimonious specification that retains only the terms significant in the fully moderated model; the complete A–D lag architecture (twelve linear lags, their squares, and all budget and legislative interactions) appears in the Appendix. Model fit improves monotonically across the sequence of nested models. Adding the squared lag terms (Model B) improves fit relative to a linear‑lags baseline (Model A vs. B: $\chi^2(12)=113.8$, $p<.001$). Adding budget moderation (Model C) further improves fit (B vs. C: $\chi^2(24)=130.7$, $p<.001$), as does adding orthogonalized legislative moderation (Model D) (C vs. D: $\chi^2(24)=80.3$, $p<.001$).

Pseudo-R^2 rises from .465 to .475 across the sequence, AIC falls from 17,705 to 17,500, variance inflation remains low (1.03–1.51), and Ljung–Box tests show no residual autocorrelation at lags 1–12 once fixed effects and the lag architecture are included.

The first result (Models A/B) establishes clear evidence for H1. Current detections rise with the centered log(1+detections$_{t-11}$) and fall with its square, while short-horizon lags (i.e., $t-1$ through $t-4$) are positive and largely linear. Because identification comes from within-jurisdiction variation under territory and month×year fixed effects, the pattern cannot be attributed to cross-sectional differences or common calendar shocks. The inverted-U at eleven months is visible in the full specification and survives in the trimmed model, which anchors the tests that follow.

TABLE 2. Parsimonious Results

	Model 1 Detections (Linear)	**Model 2** Detections (Quadratic)	**Model 3** FBI Budget	**Model 4** Legislative Attention
Intercept	**1.119***** (0.173)	**1.090***** (0.193)	**1.106***** (0.202)	**1.170***** (0.151)
Z_{t-1}	**0.310***** (0.022)	**0.255***** (0.027)	**0.238***** (0.030)	**0.237***** (0.032)
$L\times Z^2_{t-1}$	—	—	—	**-0.080*** (0.032)
Z_{t-2}	**0.108***** (0.031)	**0.090***** (0.023)	**0.100***** (0.025)	**0.100***** (0.023)
Z_{t-3}	**0.098***** (0.024)	**0.081***** (0.021)	**0.076***** (0.018)	**0.076***** (0.018)
$B\times Z^2_{t-3}$	—	—	-0.015 (0.015)	**-0.023*** (0.012)
$L\times Z^2_{t-3}$	—	—	—	**-0.057*** (0.029)
Z_{t-4}	**0.111***** (0.028)	**0.100***** (0.025)	**0.116***** (0.032)	**0.119***** (0.032)
$B\times Z^2_{t-5}$	—	—	**-0.058**** (0.020)	**-0.063***** (0.019)
Z_{t-8}	**0.064**** (0.024)	**0.065**** (0.023)	**0.080**** (0.027)	**0.083**** (0.028)
$B\times Z_{t-10}$	—	—	**-0.048*** (0.021)	**-0.043*** (0.021)

$Z_{t\text{-}11}$	**0.065** (0.033)	**0.082*** (0.035)	**0.092**** (0.034)	**0.093**** (0.033)
$Z^2_{t\text{-}11}$	—	**-0.030** (0.016)	**-0.047*** (0.022)	**-0.046*** (0.022)
$B\times Z_{t\text{-}11}$	—	—	**-0.044*** (0.022)	**-0.046*** (0.019)
$Z_{t\text{-}12}$	**0.104***** (0.021)	**0.121***** (0.027)	**0.136***** (0.032)	**0.134***** (0.031)
Observations	3,045	3,045	3,045	3,045
Fixed Effects	Territory Month×Year	Territory Month×Year	Territory Month×Year	Territory Month×Year
VIF Range	1.37–1.48	1.27–1.50	1.14–1.50	1.03–1.51
Pseudo-R²	0.4649	0.4685	0.4726	0.4751
AIC	17705.01	17615.18	17532.46	17500.20
Log-likelihood	-8555.51	-8498.59	-8433.23	-8393.10
Likelihood Ratio Tests	**B** vs **A**: $\chi^2(12)$=113.8, p=0.00000 | **C** vs **B**: $\chi^2(24)$=130.7, p=0.00000 | **D** vs **C**: $\chi^2(24)$=80.3, p=0.00000			

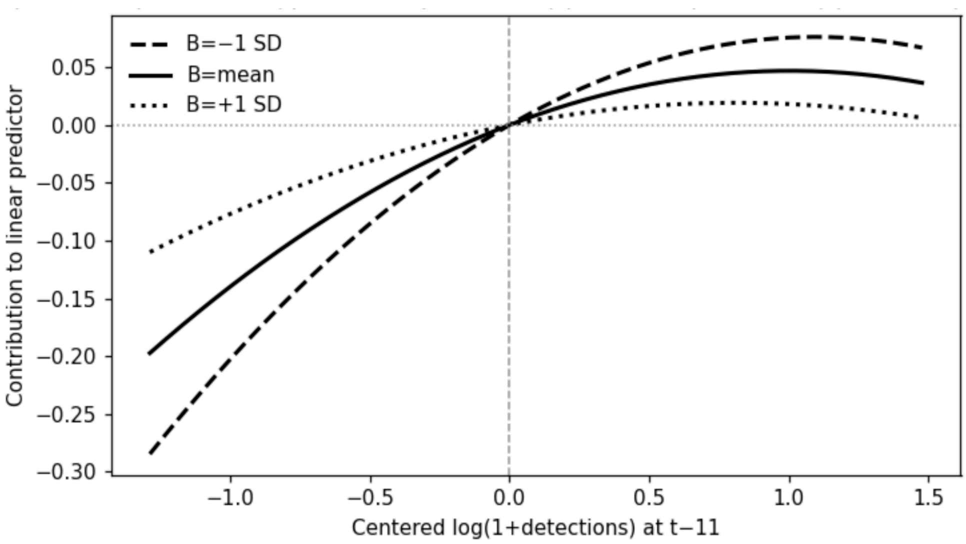
Notes. * p < .1, ** p < .05, *** p < .01. Robust standard errors in parentheses are clustered by territory. Dependent variable is monthly corruption detection count (negative binomial). All models include territory, year, and month fixed effects. Pseudo-R² calculated using McFadden's method. Budget share excluded from Model 5 due to multicollinearity.

We adjudicate between H2a/H2b by asking how budget reshapes the lag-11 curve. In Table 2, budget significantly moderates the *linear* lag-11 term (negative sign) but not the *quadratic* term, implying a shift in the location of the vertex without a change in the curvature. Figure 2 illustrates this anatomy, by plotting the the partial contribution of the lag-11 polynomial at $B = -1$ SD, mean B, and $B = +1$ SD, based on the better fitting Model D.

Two observations stand out: first, the turning point moves leftward as the budget increases. The estimated vertex $z^* = 1.095$ (SE 0.371, 95% CI [0.368, 1.822]) at $B = -1$ SD; 1.005 (SE 0.529, 95% CI [−0.033, 2.043]) at B = mean; and 0.810 (SE 1.128, 95% CI [−1.400, 3.021]) at $B = +1$ SD. Mapping back to counts by adding the median jurisdiction's lag mean in log space and exponentiating, the implied prior detection levels at the peak are ~12.3, ~11.2, and ~9 cases, respectively. Substantively, moving from low- to high-budget advances the peak by about a quarter (twenty-seven percent) fewer prior cases. Second, the downturn does not steepen across budget specifications. This follows mechanically from the coefficient pattern at $t-11$: $B\times zt-11$ is negative and significant, which shifts the numerator of

$z^*= -(\beta 1 + B\gamma 1)/[2(\beta 2 + B\gamma 2)]$, while $B \times z2t-11$ is not different from zero, which leaves curvature in the denominator largely intact. The figure reproduces this anatomy: curves pivot left with higher budget but do not display a sharper downturn. These two features support H2b's proxy-managed prediction, where resources shift the timing of the peak without deepening the decline, rather than H2a's deterrence prediction of a steeper capability-driven downturn.

Figure 2. Budget Moderation at the 11th Month Lag



Notes. Curves plot the partial contribution of the lag-11 term to the negative binomial linear predictor under three budget slices: B=−1 SD (dashed), B=mean (solid), and B=+1 SD (dotted). The x-axis is the within-jurisdiction centered log(1+detections$_{t-11}$); the y-axis is the contribution of the lag-11 polynomial to the linear predictor. Estimates come from Model D with territory fixed effects and month×year fixed effects; budget is standardized before interaction; legislative interactions are included in orthogonalized form; all other lags and fixed effects are omitted from the figure to isolate this mechanism. The vertical line marks the centered lag mean (0); the horizontal line marks zero contribution. The leftward movement of the vertex as B increases and the similar post-peak slopes across slices visualize the earlier turning point without steepening, consistent with proxy-managed adaptation rather than capability-driven deterrence.

Legislative attention provides a boundary test (H3). In the fully moderated model, legislative interactions are orthogonalized to Model C's right hand side and saturated time effe

enter Model D only in orthogonalized form, so any estimated effect is incremental to the full Model C right-hand side and to saturated time fixed effects. In the parsimonious table, the retained legislative terms concentrate at short horizons and amplify near-term adjustments around the metric. Notably, they neither move the lag-11 vertex nor alter curvature, which indicates that political salience modulates the tempo of vigilance without changing the shape of the learning cycle. This is consistent with an attention mechanism that raises the visibility stakes of the proxy but does not supply the durable capability that would deepen the downturn.

Several checks support the interpretation that these patterns reflect temporal learning rather than specification artifacts. Month-of-year linear combinations do not overturn the lag-11 inverted-U and order-1 Fourier terms do not yield seasonal explanations for the curve. As a specification check on serial dependence, we estimated a first-differenced Arellano–Bond model for $y_{it} = \log(1+\text{detections}_{it})$ with one lag of y and monthly time dummies (no level effects; collapsed instrument matrix). This diagnostic assesses the AR structure, not budget/legislative moderation. In the present report the AB routine did not execute (software dependency unavailable), so we do not report AR test statistics. The core inferences therefore rely on the fixed-effects negative-binomial models reported in the main tables and figures.

Taken together, the temporal anatomy of detection trajectories in this panel is most consistent with a proxy-managed learning dynamic. Both logics predict an inverted-U, but only the proxy-managed account predicts that added resources shift the peak earlier without deepening the decline. The results show exactly that pattern, and they do so under fixed effects, orthogonalized attention terms, and dynamic-panel robustness.