When Monitoring Backfires: Dual Audiences and Statistical Manipulation*

Chufan Wu
University of Minnesota, Department of Economics†

February 18, 2026

Abstract

When officials report to dual audiences—a verifying center and an interpreting public—stronger monitoring can backfire, improving the statistics it targets while inflating those it does not. I develop a model in which officials choose both the magnitude and detectability of GDP inflation, generating three regimes: crude over-reporting where supporters reward bold claims, sophisticated fabrication at intermediate audience compositions, and strategic under-reporting under target ratchet pressure. Using a novel panel comparing communiqué-announced growth to National Bureau of Statistics yearbook figures for 280 Chinese prefectures, I show that the 2013 anti-corruption campaign *increased* communiqué inflation in high-state-sector jurisdictions while leaving other regions unaffected. The binding constraint is political context, not industrial structure. The 2021 centralized accounting reform, estimated out of sample, reversed the pattern entirely. Statistical quality improves only when institutional design eliminates the channel on which distortion occurs.

JEL Classification: D72, D73, H11, O11, P26

Keywords: GDP manipulation, dual reporting systems, dual-audience signaling, political economy, institutional reform

*All errors are my own. I thank participants at departmental seminars for helpful comments and suggestions.

†Email: wu001513@umn.edu

1 Introduction

In January 2017, Liaoning became the first Chinese province to admit systematic GDP fraud. Governor Chen Qiufa disclosed that municipal governments had overstated fiscal revenues by 20 to 30 percent between 2011 and 2014. Steel production had collapsed, coal mines were shuttered, unemployment had doubled—yet official growth figures remained stubbornly positive throughout. The confession raised an obvious question: if one province admitted years of fabrication, what about the other thirty? A deeper puzzle, however, is not that manipulation occurred but that it was *crude*. Liaoning's inflation was visible to anyone comparing local announcements to National Bureau of Statistics (NBS) yearbooks. Satellite nighttime luminosity revealed the decline. Citizens discounted official figures and said so openly (King et al., 2013). Standard models of information control predict that once bias is detected, receivers discount, senders lose incentive, and manipulation converges to zero (Crawford and Sobel, 1982). Why did GDP inflation survive transparency—and why did officials choose the most detectable form of it? The short answer is that monitoring backfired: it improved the statistics it targeted while inflating the statistics it did not—a channel-substitution logic that parallels Sarbanes-Oxley's redirection of earnings manipulation from GAAP to non-GAAP metrics and score inflation under high-stakes educational testing.

This paper develops a theory of *performative manipulation* in which officials strategically choose both the magnitude and detectability of statistical distortion. The key insight is that GDP announcements serve dual audiences with opposing preferences. Supporters—state-sector workers, party cadres, those dependent on local patronage—reward bold economic claims as costly signals of political commitment. Moderates—informed citizens, technocratic superiors, international observers—penalize detected distortion. The optimal reporting strategy depends on the composition of these audiences. When supporters dominate, officials produce crude, easily detected inflation: mobilization value outweighs detection cost, and detectability is a feature rather than a bug. When moderates dominate, officials under-report to build credibility. At intermediate compositions, officials invest in sophisticated fabrication—internally consistent sectoral figures, smooth time series—that inflates reports while minimizing detection risk. This three-regime structure, with closed-form thresholds determined by five parameters, generates predictions that distinguish the framework from standard career-signaling and deterrence models.

I test these predictions using an institutional feature specific to China's statistical system: the coexistence of two official GDP series with different verification regimes. Each year, prefecture governments issue Statistical Communiqués (国民经济和社会发展统计公报) that announce growth figures to local audiences. The NBS subsequently publishes yearbook

figures that incorporate central adjustments and cross-validation against tax records and industrial surveys. The gap between these sources—communiqué growth minus bureau growth—captures the margin of local discretion that survives central scrutiny. A positive gap indicates over-reporting; a negative gap indicates under-reporting. I construct a novel panel matching these two series at the city-sector level for 280 prefectures over 2010–2019, yielding 2,729 city-year observations. The sample ends in 2019, the last year before the NBS assumed direct control of local GDP calculation under the 2021 centralized accounting reform—a structural break that eliminated the dual-channel system entirely. This "Type 2 gap" has received almost no scholarly attention, despite representing the margin most directly responsive to political incentives.[1]

The model generates sharp predictions organized around two sources of variation: cross-sectional differences in audience composition and temporal shocks to monitoring intensity. In the cross section, jurisdictions with high supporter share—above a closed-form threshold α_2 determined by detection technology, mobilization stakes, and audience sensitivity—operate in the crude over-reporting regime: large gaps, zero investment in signal quality. Jurisdictions with intermediate supporter share invest in sophisticated fabrication. Jurisdictions with low supporter share under-report. When a target ratchet is present—provincial governments set growth targets based on past outcomes, penalizing officials who raise today's bar—a new regime emerges: *strategic under-reporting*, in which high-α officials deflate current statistics and invest in quality to make the deflation credible. This is the standard expectation-management motive of Cao et al. (2025), enriched by the quality dimension. Over time, monitoring reforms ($\uparrow \pi_0$) produce heterogeneous effects across regimes. Crude over-reporters adjust magnitude but not quality. Sophisticated over-reporters adjust both. Under-reporters are unaffected or respond in the opposite direction. These differential responses generate testable triple-difference predictions.

The identification strategy exploits two institutional shocks within the sample period, with a third providing powerful out-of-sample validation. The 2013 anti-corruption campaign, launched under Xi Jinping, sharply increased the cost of statistical fraud by adding it to the list of actionable offenses under party discipline. The 2017 prosecution of senior Liaoning officials explicitly for GDP falsification raised enforcement credibility. I estimate difference-in-differences specifications comparing Northeast prefectures (high-α, heavy-industry economies

[1]The Type 1 gap—between NBS yearbook figures and external proxies such as satellite luminosity—measures manipulation of official statistics relative to economic fundamentals. The Type 2 gap—between communiqués and yearbooks—measures political signaling relative to already-submitted statistics. These margins respond to different incentives and constraints. Economists have used satellite data to validate official statistics externally (Henderson et al., 2012; Martinez, 2022); the NBS itself does not use satellite data to verify local reports.

with large state sectors) to other regions, before and after each shock, controlling for city and year fixed effects. Event study specifications validate parallel trends: pre-treatment coefficients are close to zero and jointly insignificant. A complementary difference-in-differences replacing the geographic treatment with fiscal dependency quartiles confirms the result across 53 high-α cities spanning multiple provinces (+5.35 pp, $p = 0.001$), substantially mitigating inference concerns from few treated clusters. The 2021 centralized accounting reform, occurring after the sample period, provides an out-of-sample test of the model's prediction that eliminating local discretion collapses the manipulation margin to zero—a prediction the data confirm.

The empirical results confirm each theoretical prediction. First, Northeast cities exhibit significantly larger communiqué-bureau gaps after the 2013 campaign. The difference-in-differences estimate for secondary-sector growth is +6.30 percentage points ($p < 0.001$), significant at the 5% level under province-level clustering ($p = 0.016$) and two-way clustering ($p = 0.029$), and at the 10% level under Fisher randomization inference ($p = 0.063$). Primary (+3.52 pp, $p < 0.01$) and tertiary (+2.02 pp, $p = 0.074$) sectors are also positive, with magnitudes tracking political salience. In a horse-race specification controlling separately for the 2013 and 2017 shocks, the post-2013 effect is +5.60 points ($p < 0.001$). Event study coefficients show effects emerging in 2014, peaking at +9.42 points in 2016 ($p < 0.01$), with no pre-trends. The pattern is consistent with crude over-reporting in a high-α region: the anti-corruption campaign raised the cost of bureau-level falsification, and Northeast officials substituted toward the less-monitored communiqué channel where their dominant supporter base rewarded bold claims. Extending the analysis beyond the sample to 2024 confirms the model's prediction: the 2021 centralized accounting reform eliminated the manipulation margin entirely (−6.11 points, $p = 0.003$), validating the theoretical mechanism out of sample.

Second, heavy industry cities provide a test of the mobilization mechanism. Among the 83 cities meeting a strict heavy-industry threshold (secondary share above 50%), heavy industry status alone does not predict manipulation: the pooled HI × Post2013 coefficient is −0.58 ($p = 0.53$), economically negligible and statistically insignificant. A horse race controlling separately for industrial structure and regional political context shows that the Northeast effect remains significant at +6.27 points ($p < 0.001$) after absorbing the heavy industry channel, while the heavy-industry coefficient is −0.35 ($p = 0.70$). The triple interaction NE × Heavy Industry × Post-2013 is −3.29 points ($p = 0.41$)—negative and insignificant, confirming that once a city is in the crude regime, further increases in manipulation capacity add nothing. Heavy industry provides the *capacity* for manipulation—it proxies for the state-sector dependence that generates high α—but it is not the binding constraint. The binding constraint is political context: the Northeast's combination of corruption stigma, economic decline, and dense patronage networks places these cities deep in the crude over-reporting

regime where other heavy-industry regions do not reside.

Third, the cross section reveals pervasive strategic under-reporting. At the sectoral level, 94.6 percent of prefectures (265 of 280) under-report primary-sector growth on average, while only 28.6 percent under-report secondary-sector growth, with Shanghai ranking among the most prominent under-reporters in the full descriptive sample that includes direct-controlled municipalities.[2] This prevalence is a direct implication of the target ratchet: when future growth targets are set based on past reported performance, officials face a universal incentive to lowball current statistics. The under-reporting is not random. Within the Northeast, the primary-to-secondary co-movement coefficient is $+0.298$ ($p = 0.004$) after absorbing city and year fixed effects, indicating that Northeast officials who inflate secondary-sector claims simultaneously adjust primary-sector figures beyond the national trend—consistent with coordinated sectoral manipulation rather than measurement error. Nationally, communiqués systematically under-report primary growth and over-report secondary growth, a system-level bias visible in 7 of 10 sample years across all regions.

These findings contribute to three literatures. The GDP measurement literature compares official statistics to external proxies—satellite luminosity, electricity consumption, tax revenue—to quantify aggregate manipulation (Henderson et al., 2012; Martinez, 2022; Chen et al., 2019; Wallace, 2016). These studies treat manipulation as monolithic; I exploit within-government variation to identify the strategic margin. The China political economy literature documents career incentives for GDP inflation (Li and Zhou, 2005; Xu, 2011; Xiong, 2022; Lü and Landry, 2014): Zeng and Zhou (2024) show promotion pressure inflates growth by 3.4 percentage points with no real productivity gains; Chen et al. (2021) find county officials early in tenure cheat more aggressively; Xu et al. (2025) document peak exaggeration in the final year. Career-signaling models predict manipulation where promotion prospects are brightest; the data show the opposite—manipulation peaks in the Northeast, where prospects are weakest and citizen distress is highest, consistent with a horizontal legitimacy motive (Lorentzen, 2014; Wedeen, 1999).

The authoritarian governance literature models information control as costly signaling that decreases with monitoring quality (Guriev and Treisman, 2019; Egorov et al., 2009; Hollyer et al., 2011). The canonical prediction is monotonic: better monitoring, less manipulation. I show this prediction fails when officials face dual audiences. The anti-corruption campaign tightened bureau-level monitoring but left the communiqué channel unverified, and Northeast officials responded by *increasing* communiqué inflation. This mechanism generalizes beyond

[2]The regression sample excludes the four direct-controlled municipalities (Beijing, Shanghai, Tianjin, Chongqing) because they face distinct institutional arrangements. Descriptive statistics on under-reporting patterns use the broader sample to document the national scope of the phenomenon.

autocracies: the tightening of Sarbanes-Oxley compliance in the United States redirected corporate earnings manipulation from GAAP to non-GAAP reporting channels (Dichev et al., 2013), and high-stakes educational testing produces similar dual-channel substitution (Jacob and Levitt, 2003). Any setting where accountability operates through parallel channels with heterogeneous verification is susceptible to the same strategic logic.

I make three contributions. First, I develop a formal model of performative manipulation with dual audiences that generates three optimal reporting regimes—under-reporting, sophisticated over-reporting, and crude over-reporting—characterized by closed-form thresholds (α^0, α_2, g^*). The model provides a unified framework for understanding why manipulation varies in both magnitude and detectability across jurisdictions, and why monitoring reforms produce heterogeneous responses. Existing models predict either uniform deterrence (Becker, 1968) or uniform inflation (Martinez, 2022); the regime structure explains when each prediction applies and when neither does.

Second, I construct the first prefecture-level panel comparing communiqué growth claims to NBS yearbook figures across sectors. This data infrastructure reveals a previously unstudied margin of statistical manipulation and enables identification of strategic patterns—sectoral reallocation, coordinated cross-sector distortion, regime-dependent reform responses—that are invisible in aggregate analyses relying on external benchmarks alone.

Third, I provide the first causal evidence that monitoring reforms produce regime-dependent effects on statistical quality, together with out-of-sample validation of the theoretical mechanism. The 2013 anti-corruption campaign *increased* communiqué inflation in the Northeast while leaving other regions largely unaffected—a finding estimated on 2010–2019 data alone. The model, calibrated on this sample, predicts that eliminating local discretion would collapse manipulation to zero. The 2021 centralized accounting reform provides a natural experiment to test this prediction: extending the analysis to 2024 reveals a sharp reversal (-6.11 points, $p = 0.003$), confirming the out-of-sample prediction.

The remainder of the paper proceeds as follows. Section 2 develops the model and derives testable predictions. Section 3 describes China's dual reporting system, the data construction, and summary statistics. Section 4 develops the econometric specification and discusses identification. Section 5 reports the main reduced-form findings. Section 6 maps model primitives to empirical proxies and calibrates key parameter ratios. Section 7 examines mechanisms. Section 8 provides robustness checks. Section 9 concludes.

2 A Model of Performative Manipulation

This section develops a model of GDP manipulation in which a representative official faces dual audiences with conflicting preferences over signal quality. The official's optimal strategy generates closed-form thresholds that determine when crude manipulation is optimal, when officials strategically understate performance, and how monitoring reforms produce heterogeneous effects. The framework is a single-agent decision problem, not a strategic game; the contribution is to show that even this simple structure generates a rich regime taxonomy that organizes the empirical patterns documented below.

2.1 Setup

A local official governs a jurisdiction with two audience types. *Supporters*, comprising fraction $\alpha \in (0,1)$, are regime loyalists—SOE workers, Party members, subordinate cadres—whose material welfare depends on the official's political standing. They derive utility from bold economic claims that signal commitment and capacity to deliver patronage. *Moderates*, comprising fraction $1-\alpha$, evaluate claims against independent information; detection of manipulation imposes reputational damage and potential central scrutiny.[3]

The official observes true performance y and chooses two instruments: signal quality $s \geq 0$ and misreporting magnitude $m \in \mathbb{R}$. Higher s represents sophisticated manipulation—internally consistent sectoral figures, smooth time series. The magnitude m takes unrestricted sign: $m > 0$ is over-reporting, $m < 0$ is under-reporting. The official maximizes

$$U(s,m) = \underbrace{\alpha\theta m\, e^{-\zeta s}}_{\text{mobilization}} - \underbrace{(1-\alpha) m\, \pi_0\, e^{-\mu s}}_{\text{detection cost}} - \underbrace{\tfrac{1}{2}s^2}_{\text{fabrication}} - \underbrace{\tfrac{\gamma}{2}m^2}_{\text{magnitude cost}}. \tag{1}$$

The mobilization term captures the idea that crude signals maximize political engagement: $\theta > 0$ scales the political return to bold claims, and $\zeta > 0$ governs how quickly this value decays with sophistication. The detection cost reflects that crude manipulation is detected at baseline rate $\pi_0 \in (0,1]$, reduced by quality investment at rate $\mu > 0$. Crucially, when $m < 0$, the detection term flips sign: moderates *reward* conservative reporting. The fabrication cost $\frac{1}{2}s^2$ and magnitude cost $\frac{\gamma}{2}m^2$ are convex.[4]

Assumption 1 (Separation)**.** *$\mu > \zeta$: detection technology improves with quality faster than mobilization value decays.*

[3] This follows selectorate theory (Bueno de Mesquita et al., 2003): supporters value bold signals because staking one's reputation on dramatic claims reveals political commitment.

[4] Appendix C derives the mobilization function from expressive utility primitives.

Assumption 2 (Regularity). $(1-\alpha)\pi_0\mu^2 e^{-\mu s} + 1 > \alpha\theta\zeta^2 e^{-\zeta s}$ *for all* $s \geq 0$, *and* $\gamma > 0$.

Assumption 1 ensures that quality investment creates a non-trivial trade-off. Assumption 2 guarantees the Hessian is negative definite, so first-order conditions uniquely characterize the optimum.

2.2 Unconstrained Benchmark

Define two key functions:

$$\Delta_1(s) \equiv \alpha\theta\, e^{-\zeta s} - (1-\alpha)\pi_0\, e^{-\mu s}, \tag{2}$$

$$\Delta_2(s) \equiv (1-\alpha)\pi_0\mu\, e^{-\mu s} - \alpha\theta\zeta\, e^{-\zeta s}. \tag{3}$$

The first-order conditions yield $m^*(s) = \Delta_1(s)/\gamma$ and $s = m \cdot \Delta_2(s)$. The function Δ_1 determines the sign and magnitude of misreporting; Δ_2 determines whether quality investment is worthwhile. Since $s \geq 0$, quality investment requires m and Δ_2 to share the same sign, generating three regimes:

Theorem 1 (Unconstrained Optimum). *Under Assumptions 1–2, define* $\alpha^0 \equiv \pi_0/(\theta+\pi_0)$ *and* $\alpha_2 \equiv \pi_0\mu/(\pi_0\mu+\theta\zeta)$, *with* $\alpha^0 < \alpha_2$. *The unique optimum is:*

(a) ***Under-reporting*** *(*$\alpha < \alpha^0$*):* $m^* = [\alpha\theta - (1-\alpha)\pi_0]/\gamma < 0, \quad s^* = 0.$

(b) ***Sophisticated over-reporting*** *(*$\alpha^0 < \alpha < \alpha_2$*):* $m^* > 0, \quad s^* > 0$, *uniquely characterized by the first-order system, with* $dm^*/d\alpha > 0$ *and* $ds^*/d\alpha < 0$.

(c) ***Crude over-reporting*** *(*$\alpha > \alpha_2$*):* $m^* = [\alpha\theta - (1-\alpha)\pi_0]/\gamma > 0, \quad s^* = 0.$

At $\alpha = \alpha^0$, *reporting is truthful:* $m^* = s^* = 0$.

The proof is in Appendix A.2. The threshold $\alpha^0 = \pi_0/(\theta+\pi_0)$ is the ratio of detection effectiveness to total incentive strength at the crude signal level. The threshold $\alpha_2 = \pi_0\mu/(\pi_0\mu+\theta\zeta)$ marks where quality investment ceases to pay. The comparative statics are:

$$\frac{\partial\alpha^0}{\partial\pi_0} > 0, \quad \frac{\partial\alpha^0}{\partial\theta} < 0; \qquad \frac{\partial\alpha_2}{\partial\pi_0} > 0, \quad \frac{\partial\alpha_2}{\partial\theta} < 0. \tag{4}$$

Better detection ($\uparrow \pi_0$) raises both thresholds, expanding under-reporting. Stronger mobilization incentives ($\uparrow \theta$) lower both thresholds, making over-reporting more prevalent and crude manipulation more likely.

2.3 Institutional Frictions: Target Ratchet and Reporting Floor

Two institutional features complicate the benchmark. First, provincial governments set growth targets based on past outcomes: $\text{Target}_{i,t+1} = \phi + \rho \cdot \hat{y}_{i,t}$, where $\rho \in (0,1)$. The discounted cost of raising today's report by one unit is $g > 0$, strictly increasing in past reports (Cao et al., 2025). Second, central bureaus and satellite data constrain how far officials can understate: $m \geq \underline{m}$ with $\underline{m} < 0$.

The constrained problem adds $-gm$ to the objective:

$$\max_{s \geq 0,\, m \geq \underline{m}} U(s,m) = \alpha\theta m e^{-\zeta s} - (1-\alpha) m \pi_0 e^{-\mu s} - gm - \tfrac{1}{2}s^2 - \tfrac{\gamma}{2}m^2. \tag{5}$$

The target cost shifts the net benefit of magnitude: $\Delta_1(s;\, g) \equiv \Delta_1(s) - g$. The truthful-reporting threshold becomes

$$\alpha_g^0 \equiv \frac{\pi_0 + g}{\theta + \pi_0} > \alpha^0, \tag{6}$$

and the constraint binds below

$$\alpha_L \equiv \frac{\pi_0 + g + \gamma \underline{m}}{\theta + \pi_0} < \alpha_g^0. \tag{7}$$

The signal quality threshold α_2 is unchanged. Define the critical target cost:

$$g^* \equiv \frac{\pi_0 \theta (\mu - \zeta)}{\pi_0 \mu + \theta \zeta}. \tag{8}$$

When $g < g^*$, we have $\alpha_g^0 < \alpha_2$ (weak target pressure); when $g > g^*$, $\alpha_2 < \alpha_g^0$ (strong target pressure).

Theorem 2 (Constrained Optimum)**.** *Under Assumptions 1–2 with $\underline{m} < 0$ and $g > 0$:*

Case A: Weak target pressure ($g < g^*$). *Five regimes ordered by α:*

(i) *Constrained under-reporting ($\alpha < \alpha_L$): $m^* = \underline{m}$, $s^* = 0$.*

(ii) *Interior under-reporting ($\alpha_L < \alpha < \alpha_g^0$): $m^* \in (\underline{m}, 0)$, $s^* = 0$.*

(iii) *Truthful ($\alpha = \alpha_g^0$): $m^* = 0$, $s^* = 0$.*

(iv) *Sophisticated over-reporting ($\alpha_g^0 < \alpha < \alpha_2$): $m^* > 0$, $s^* > 0$.*

(v) *Crude over-reporting ($\alpha > \alpha_2$): $m^* > 0$, $s^* = 0$.*

Case C: Strong target pressure ($g > g^*$). [5] *Five regimes:*

(i) *Constrained under-reporting ($\alpha < \alpha_L$):* $m^* = \underline{m}$, $s^* = 0$.

(ii) *Compositional under-reporting ($\alpha_L < \alpha < \alpha_2$):* $m^* \in (\underline{m}, 0)$, $s^* = 0$.

(iii) *Strategic under-reporting ($\alpha_2 < \alpha < \alpha_g^0$):* $m^* \in (\underline{m}, 0)$, $s^* > 0$.

(iv) *Truthful ($\alpha = \alpha_g^0$):* $m^* = 0$, $s^* = 0$.

(v) *Crude over-reporting ($\alpha > \alpha_g^0$):* $m^* > 0$, $s^* = 0$.

The proof is in Appendix A.3. The key new regime is *strategic under-reporting* (Case C, Regime iii): officials with high supporter share who would over-report absent target pressure now deflate, investing in quality to make the deflation credible. Quality investment *migrates* from over-reporters to under-reporters as target pressure strengthens: when g crosses g^*, sophisticated over-reporting vanishes and strategic under-reporting emerges.[6]

2.4 Comparative Statics and Testable Predictions

The model generates predictions organized around two questions: what determines the cross-sectional pattern of misreporting, and how do monitoring reforms propagate through the regime structure?

2.4.1 Cross-Sectional Patterns

The sign and magnitude of misreporting are governed by audience composition (α), mobilization stakes (θ), and target pressure (g). The model predicts $\partial m^*/\partial \alpha > 0$: jurisdictions with higher state-sector dependence report more optimistically. Target pressure works in the opposite direction: $\partial m^*/\partial g = -1/\gamma < 0$. Among jurisdictions with $m^* < 0$, those with high α show high sectoral consistency ($s^* > 0$), while those with low α do not ($s^* = 0$). Quality investment among under-reporters is thus a marker of strategic motivation. Three distributional features aid identification: bunching at $\underline{m}$ among low-α jurisdictions, a kink in $m^*(\alpha)$ at α_L, and a kink in $s^*(\alpha)$ at α_2 (details in Online Appendix Section A).

[5] Case B ($g = g^*$) is the knife-edge where $\alpha_g^0 = \alpha_2$: the sophisticated over-reporting and strategic under-reporting regimes both vanish, and the regime structure collapses to three regimes (constrained under-reporting, interior under-reporting, crude over-reporting). This measure-zero case is omitted.

[6] I use "regime" throughout to denote regions of the parameter space with qualitatively different optimal strategies. The term does not imply a game-theoretic equilibrium concept; the regimes are partitions of a single agent's optimization problem.

Table 1: Summary of Optimal Reporting Regimes

Regime	m^*	s^*	α range
Unconstrained benchmark ($g = 0$, $m \in \mathbb{R}$)			
Under-reporting	< 0	0	$\alpha < \alpha^0$
Sophisticated over-reporting	> 0	> 0	$\alpha^0 < \alpha < \alpha_2$
Crude over-reporting	> 0	0	$\alpha > \alpha_2$
Case A: weak target pressure ($0 < g < g^*$)			
Constrained under	$= \underline{m}$	0	$\alpha < \alpha_L$
Interior under	$\in (\underline{m}, 0)$	0	$\alpha_L < \alpha < \alpha_g^0$
Sophisticated over	> 0	> 0	$\alpha_g^0 < \alpha < \alpha_2$
Crude over	> 0	0	$\alpha > \alpha_2$
Case C: strong target pressure ($g > g^*$)			
Constrained under	$= \underline{m}$	0	$\alpha < \alpha_L$
Compositional under	$\in (\underline{m}, 0)$	0	$\alpha_L < \alpha < \alpha_2$
Strategic under	$\in (\underline{m}, 0)$	> 0	$\alpha_2 < \alpha < \alpha_g^0$
Crude over	> 0	0	$\alpha > \alpha_g^0$

Note: Truthful reporting ($m^* = 0$, $s^* = 0$) occurs at the knife-edge $\alpha = \alpha^0$ (benchmark) or $\alpha = \alpha_g^0$ (constrained). Strategic under-reporters invest in quality ($s^* > 0$) to conceal against-type deflation driven by target ratchet costs.

2.4.2 Monitoring Reforms

The model predicts heterogeneous responses to an increase in detection technology (π_0):

(i) *Constrained under-reporters*: no response ($m^* = \underline{m}$ fixed).

(ii) *Compositional under-reporters*: gaps narrow, no quality change.

(iii) *Strategic under-reporters* (Case C): gaps narrow, quality rises.

(iv) *Sophisticated over-reporters* (Case A): gaps narrow, quality rises.

(v) *Crude over-reporters*: gaps narrow, no quality change.

This generates a triple-difference prediction. Coastal prefectures, with moderate α, adjust on *both* margins when monitoring improves. Northeast prefectures, deep in the crude regime, adjust magnitude only. In the limit where $\pi_0 \to \infty$—as when centralized accounting eliminates local discretion—both thresholds converge to unity, collapsing all jurisdictions into under-reporting or truthful reporting.

Proposition 3 (Monitoring Eliminates Over-Reporting)**.** *Fix* $(\alpha, \theta, \zeta, \mu, \gamma, g)$ *with* $\alpha \in (0, 1)$. *There exists a finite monitoring threshold*

$$\bar{\pi}_0(\alpha) \equiv \frac{\alpha\theta - g}{1 - \alpha} \tag{9}$$

such that:

(a) *For* $\pi_0 > \bar{\pi}_0(\alpha)$*:* $m^*(\alpha) < 0$ (*under-reporting*).

(b) *For* $\pi_0 = \bar{\pi}_0(\alpha)$*:* $m^*(\alpha) = 0$ (*truthful reporting*).

(c) *For* $\pi_0 < \bar{\pi}_0(\alpha)$*:* $m^*(\alpha) > 0$ (*over-reporting*).

Moreover, $\bar{\pi}_0$ *is strictly increasing in* α *and* θ*, and strictly decreasing in* g*:*

$$\frac{\partial \bar{\pi}_0}{\partial \alpha} = \frac{\theta - g}{(1 - \alpha)^2} > 0, \qquad \frac{\partial \bar{\pi}_0}{\partial \theta} = \frac{\alpha}{1 - \alpha} > 0, \qquad \frac{\partial \bar{\pi}_0}{\partial g} = \frac{-1}{1 - \alpha} < 0. \tag{10}$$

Proof. Evaluate the first-order condition for m at $s = 0$: $m^*(0) = [\alpha\theta - (1 - \alpha)\pi_0 - g]/\gamma$. Setting $m^*(0) = 0$ and solving for π_0 gives $\bar{\pi}_0 = (\alpha\theta - g)/(1 - \alpha)$. Since m^* is strictly decreasing in π_0 ($\partial m^*/\partial \pi_0 = -(1 - \alpha)/\gamma < 0$), the sign of m^* flips exactly at $\bar{\pi}_0$. The comparative statics follow by direct differentiation. □

The 2021 centralized accounting reform corresponds empirically to pushing π_0 above $\bar{\pi}_0$ for nearly all prefectures simultaneously—an interpretation that Section 6 quantifies as a hundredfold increase in effective detection.

2.4.3 Identification Mapping

The model's five structural parameters map to observables as follows. The audience composition α is identified by the fiscal dependency rate (Section 6.1). The detection technology μ is identified by administrative litigation intensity per capita (Section 6.2). The mobilization return θ is identified by tenure-year variation (Section 6.3). The baseline detection rate π_0 is identified by the 2021 centralized accounting reform. The target ratchet cost g is identified by past reported growth rates. Each parameter enters the optimum through a *different* margin, reducing the risk that a single confound drives the empirical results. Section 6 implements this mapping quantitatively and calibrates $\theta/\pi_0 \in [3.2, 12.8]$.

3 Background and Data

Chinese prefectures report economic performance through two parallel channels that face sharply different verification regimes. This section documents the institutional origins of that asymmetry, describes reform episodes that generate identifying variation, and details the data construction.

3.1 China's Dual Statistical System

China's subnational GDP statistics are produced through two distinct channels—an institutional feature central to this paper's identification strategy.

Bureau statistics (统计年鉴). Prefecture-level statistical bureaus collect economic data from enterprises, households, and administrative records, then transmit standardized tables to provincial bureaus and the National Bureau of Statistics (NBS) in Beijing. The NBS subjects these submissions to systematic cross-validation against independent administrative data: value-added tax receipts, customs declarations, enterprise survey microdata, and electricity consumption figures (Rawski, 2001; Holz, 2014). Bureau submissions must satisfy internal consistency requirements: sectoral value-added must sum to total GDP, and the production, income, and expenditure approaches must approximately reconcile. These verification procedures are imperfect (Rawski, 2001; Holz, 2014) but impose *binding* constraints: an official who inflates bureau-reported industrial output must also reconcile the inflated figure with tax receipts, enterprise surveys, and electricity data, each collected through separate administrative channels.

Statistical communiqués (国民经济和社会发展统计公报). Each spring, prefectures release annual *Statistical Communiqués on National Economic and Social Development*, reporting the previous year's GDP growth rate alongside narrative accounts of economic achievements. These communiqués serve a political audience: local People's Congress delegates, provincial Party committees that evaluate officials for promotion (Li and Zhou, 2005), local media, and citizens who form expectations about public investment. Crucially, the head of the municipal statistical bureau reports to the mayor, not to the NBS. The communiqué is drafted and approved within the city government apparatus.

The verification asymmetry. The communiqué growth rate r_{it}^R and the bureau growth rate r_{it}^B describe the same economy in the same year. When they differ, the discrepancy reveals the margin of local discretion that survives central scrutiny. Two features create this margin.

First, the NBS does not systematically reconcile communiqué claims against yearbook figures. Communiqués are released months before bureau yearbooks are finalized, so an official can announce high growth, collect the political benefit during the legislative session, and face no institutional consequence when the bureau figure appears later. Second, communiqué audiences lack the technical capacity to audit growth calculations. People's Congress delegates observe a single growth rate without access to the administrative microdata underpinning the bureau figure (Li and Zhou, 2005; Xu, 2011). This asymmetry creates a one-directional prediction: officials who wish to present a more favorable narrative can inflate communiqué claims at low marginal cost, while inflating bureau statistics requires coordinating across independently audited administrative channels. The gap between the two channels therefore isolates the margin where political signaling is least constrained by verification technology.

Career incentives. Under the regionally decentralized authoritarian framework (Xu, 2011), prefecture officials compete in promotion tournaments where economic growth is a key evaluation criterion (Li and Zhou, 2005). Wallace (2016) shows that GDP–electricity gaps widen when Party secretaries face promotion decisions. Chen et al. (2005) confirm that *relative* performance drives evaluations. These career stakes map to the mobilization intensity parameter θ in the model.

3.2 Reform Episodes

Two reform episodes within the sample period generate variation in monitoring intensity (π_0). A third reform, implemented after the analysis period, provides out-of-sample validation.

2013: Anti-corruption campaign. In late 2012, the Xi Jinping administration launched an anti-corruption campaign of unprecedented scope. The Central Commission for Discipline Inspection (CCDI) dispatched inspection teams to all 31 provinces, and statistical fraud was explicitly added to the list of actionable corruption offenses. Over 150 statistical bureau officials were disciplined or prosecuted for data falsification between 2013 and 2017 (Wallace, 2016).[7] The NBS simultaneously enhanced its cross-validation protocols.

In the model, this episode raised bureau-channel detection probability π_0 while no comparable enforcement emerged for communiqué claims. Simultaneously, the campaign increased mobilization stakes θ for Northeast officials facing investigation or removal. The combination of asymmetric enforcement and intensified career pressure generates a clear prediction: the communiqué–bureau gap should *increase* after 2013 in the Northeast.

[7]The most prominent case was Liaoning Province, whose governor publicly acknowledged systematic GDP inflation in January 2017. Inner Mongolia and Tianjin subsequently revised their GDP figures downward.

2017: Statistical management reform (统计管理体制改革). The State Council reformed the statistical management system in 2017, giving provincial statistical bureaus greater authority to review prefecture-level submissions and requiring source data alongside aggregate figures. The reform reduced local government control over data collection while leaving the communiqué release process largely unchanged.

2021: Centralized GDP accounting (地区生产总值统一核算)—out-of-sample validation. The NBS assumed direct control of prefecture-level GDP calculation in 2021, mechanically eliminating the bureau-channel manipulation margin. This reform, implemented after the analysis period (2010–2019), provides a powerful out-of-sample test: the model predicts (see Section 2.4) that gaps collapse toward zero. Section 8 confirms the predicted reversal (-6.11 percentage points, $p = 0.003$).

3.3 Data Construction

I construct a panel dataset linking bureau statistics and communiqué claims at the prefecture level across three economic sectors.

Bureau statistics. Official GDP levels and sectoral growth rates come from NBS prefecture-level open data (国家统计局) covering 2010–2019. I compute year-on-year growth rates for three sectors following NBS definitions: primary (agriculture, forestry, animal husbandry, fishing), secondary (industry, construction), and tertiary (services). The relevant question for this paper is whether bureau figures differ systematically from communiqué claims, not whether they are perfectly accurate.

Statistical communiqués. I collected over 4,000 prefecture-level communiqué documents from TJCN.org, an archive with detailed government papers including annual reports directly scraped from government websites. I extracted sector-specific growth claims using regular expression matching, with manual verification of ambiguous cases and cross-checking against original government websites where available. A communiqué released in March of year t reports on year $t-1$, as does the bureau yearbook published later in year t.

The gap variable. The key outcome variable is:

$$\text{Gap}_{it}^{s} \;=\; r_{it}^{R,s} \;-\; r_{it}^{B,s} \tag{11}$$

where $r_{it}^{R,s}$ is the communiqué growth claim and $r_{it}^{B,s}$ is the bureau growth rate for prefecture i, year t, and sector $s \in \{\text{primary, secondary, tertiary}\}$. Positive values indicate over-reporting. Both growth rates are computed in real terms using constant-price (不变价) GDP deflators applied at the source. In robustness checks, I impose a ± 2 percentage point tolerance band to accommodate measurement error (details in Online Appendix Section C). To my knowledge, this is the first systematic dataset comparing communiqué claims to bureau statistics at the prefecture level across multiple sectors.

Sample construction. The analysis sample consists of 280 prefecture-level cities observed over 2010–2019, yielding 2,729 city-year observations for the secondary-sector specification.[8] The sample ends in 2019, the last year before the NBS assumed direct control of local GDP calculation. The panel is unbalanced: not all prefectures report communiqués in every year. The 2010–2019 window spans the dual-channel system, with a pre-treatment baseline of at least three years before the 2013 campaign and an end date preceding both the 2021 centralized accounting reform and COVID-19.

Regional and industrial classification. I classify prefectures into four regions: Northeast (33 cities in Liaoning, Jilin, Heilongjiang), Coastal (Shandong, Jiangsu, Zhejiang, Fujian, Guangdong), Central (Henan, Hubei, Hunan, Anhui, Jiangxi, Shanxi), and Western (remaining provinces). Heavy-industry cities are those where the secondary sector accounts for more than 50 percent of GDP in 2010 (83 cities), measured prior to treatment to avoid endogeneity.

Auxiliary data. I supplement the core panel with prefecture characteristics from NBS yearbooks, leader tenure data from government biographical databases, and anti-corruption investigation records from CCDI announcements (Chen and Kung, 2019). Details on all auxiliary data sources appear in the Online Appendix.

3.4 Summary Statistics

Table 2 presents summary statistics for the full sample.

Three features merit emphasis.

[8] I exclude the four direct-controlled municipalities (Beijing, Shanghai, Tianjin, Chongqing) and observations where administrative boundary changes produce spurious growth rate changes exceeding 100 percentage points. A total of 290 unique prefectures appear at least once in the matched panel; the regression sample is smaller because some city-years lack covariates required for estimation. Descriptive tables report the full matched sample.

Table 2: Summary Statistics: Full Sample (2010–2019)

Variable	Mean	Std Dev	Min	Max	N
Primary Gap	-3.74	11.69	-254.11	69.39	2,728
Secondary Gap	1.95	10.36	-78.30	100.48	2,729
Tertiary Gap	-4.08	8.72	-90.12	53.36	2,806
Aggregate Gap	-1.93	6.88	-96.90	37.98	2,806
Communiqué Primary Growth	4.14	2.77	-22.30	52.80	2,806
Communiqué Secondary Growth	10.28	6.63	-25.50	111.60	2,806
Communiqué Tertiary Growth	9.45	3.14	-10.10	23.90	2,806
Bureau Primary Growth	8.87	12.85	-65.59	422.50	5,277
Bureau Secondary Growth	13.13	15.15	-69.39	338.40	5,279
Bureau Tertiary Growth	14.89	10.63	-51.06	283.23	5,355

Feature 1: Substantial dispersion around a modest mean. The mean aggregate gap is −1.93 percentage points, masking critical heterogeneity: the standard deviation is 6.88 percentage points.[9] The large right tail—and systematic geographic concentration of positive gaps—captures the strategic inflation this paper studies.

Feature 2: Sectoral asymmetry. The mean secondary sector gap is +1.95 percentage points (SD = 10.36), while the primary and tertiary gaps are negative (−3.74 and −4.08). Secondary industry is the sector where communiqués most consistently claim higher growth, consistent with the model's prediction that manipulation concentrates where political audiences are most attentive.

Table 3: National Average Gap by Year and Region (2010–2019)

	Northeast				Non-Northeast				
Year	Prim	Sec	Tert	Agg	Prim	Sec	Tert	Agg	Cities
2010	-8.14	-6.92	-2.99	-5.85	-11.25	-4.97	-7.45	-7.61	284
2011	-13.96	-6.96	-6.71	-9.11	-11.29	-6.29	-5.79	-7.77	284
2012	-7.23	2.21	-3.22	-2.70	-6.16	3.27	-6.84	-3.34	287
2013	-2.04	2.21	-2.03	-0.57	-3.31	3.71	-3.93	-1.22	288
2014	2.08	8.84	-1.96	2.98	-1.62	3.88	-3.80	-0.48	287
2015	-0.68	12.86	-2.23	3.29	-0.30	7.36	-2.43	1.64	288
2016	5.97	11.39	5.56	7.83	-2.02	2.90	-3.47	-0.81	287
2017	11.83	-0.30	-0.07	4.44	2.60	-0.55	-3.00	-0.23	288
2018	-1.04	1.77	-2.52	-0.72	-0.30	2.31	-3.46	-0.43	288
2019	-4.83	22.13	8.90	8.43	-6.91	3.60	-5.27	-3.06	288

[9]Extreme values arise from administrative boundary changes, reclassifications, and occasional data entry errors. Results are robust to winsorizing at the 1st and 99th percentiles (Section 8).

Feature 3: Temporal dynamics. Table 3 reveals sharp temporal variation. Secondary sector gaps shift from −5.20 in 2010 to a peak of +7.98 in 2015 before declining. Out-of-sample data confirm that the gap collapsed to −9.66 in 2021—consistent with Prediction 2.4.2 (Section 8).

Table 4: Regional Difference-in-Differences: All Regions vs National Average (2010–2019)

	NE	Coastal	Central	Western
Region × Post2013	6.295***	-4.652***	1.788**	-0.171
	(1.768)	(0.740)	(0.828)	(1.031)
DV: Secondary Sector Gap. City and year FE. City-clustered SEs.				
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.				

Geographic patterns. Table 4 previews the formal results. Northeast cities exhibit a secondary sector gap 6.30 percentage points larger after 2013 ($p < 0.01$). Coastal cities show the opposite: gaps 4.65 points *smaller* ($p < 0.01$). This geographic pattern matches the model's prediction that $\theta^{\text{NE}} > \theta^{\text{Coast}}$ drives regional heterogeneity (Section 2.4).

4 Empirical Strategy

The model generates testable predictions about where, when, and how officials inflate communiqué claims relative to bureau figures. This section develops the econometric framework: a baseline difference-in-differences, an event study to validate the design, a horse race to isolate competing policy shocks, mechanism tests, and identification assumptions.

4.1 Baseline Specification

The model predicts that officials in the crude over-reporting regime inflate communiqué claims most aggressively after a monitoring shock that tightens bureau verification but leaves the communiqué channel unmonitored. I test this using:

$$\text{Gap}_{it} = \beta \left(\text{NE}_i \times \text{Post2013}_t\right) + \alpha_i + \delta_t + \varepsilon_{it} \tag{12}$$

where Gap_{it} is the communiqué-minus-bureau growth rate in percentage points. NE_i equals one for cities in Liaoning, Jilin, and Heilongjiang. Post2013_t equals one for 2013 and later. City fixed effects α_i absorb time-invariant characteristics; year fixed effects δ_t absorb aggregate shocks. Standard errors are clustered at the city level throughout (Bertrand et al., 2004).

Under the dual-audience model, $\beta > 0$: officials in the corner regime responded to tightened bureau monitoring by intensifying communiqué inflation, because their supporter base rewards visible signs of prosperity and the communiqué channel faced no enforcement upgrade.

Why Northeast China? Four features make the Northeast the natural treatment group: (i) *high supporter share* (α): SOE employment exceeds 40 percent of urban employment, roughly triple the national average; (ii) *high mobilization stakes* (θ): the region was the epicenter of high-profile corruption investigations; (iii) *economic decline*: GDP growth fell from 9.1 percent in 2012 to 2.5 percent in 2016; (iv) *dual-channel asymmetry*: central scrutiny intensified for bureau statistics but the communiqué channel remained weakly monitored.

4.2 Event Study

I estimate year-specific interactions:

$$\text{Gap}_{it} = \sum_{k \neq 2012} \beta_k \left(\text{NE}_i \times \mathbf{1}[t = k]\right) + \alpha_i + \delta_t + \varepsilon_{it} \tag{13}$$

with 2012 as the omitted reference year. Under parallel trends, pre-period coefficients should be small and statistically indistinguishable from zero. The model predicts flat pre-trends (2010–2012), onset and escalation (2013–2014), peak manipulation during supply-side reform (2015–2016), and post-enforcement decline (2017–2019). Section 8 extends the analysis to 2024 as out-of-sample validation.

4.3 Horse Race: Isolating Competing Policy Shocks

To disentangle the 2013 anti-corruption effect from the 2017 Liaoning revelation, I estimate:

$$\text{Gap}_{it} = \beta_1 \left(\text{NE}_i \times \text{Post2013}_t\right) + \beta_2 \left(\text{NE}_i \times \text{Post2017}_t\right) + \alpha_i + \delta_t + \varepsilon_{it} \tag{14}$$

The horse race tests whether the 2013 effect (β_1) remains positive after controlling for the 2017 shock.

4.4 Mechanism Tests

4.4.1 Heavy Industry Test

I replace the Northeast indicator with a heavy industry classification ($\text{HeavyInd}_i = 1$ for secondary-sector GDP share exceeding 50 percent). The model predicts that when estimated jointly, the Northeast indicator retains significance while the heavy industry indicator is small and insignificant.

4.4.2 Sectoral Decomposition

I estimate sector-specific versions of the baseline:

$$\text{Gap}_{it}^{s} = \beta^{s} \left(\text{NE}_i \times \text{Post2013}_t\right) + \alpha_i + \delta_t + \varepsilon_{it}^{s} \tag{15}$$

for $s \in \{\text{Primary}, \text{Secondary}, \text{Tertiary}\}$. The verification hierarchy generates: $\beta^{\text{Secondary}} > \beta^{\text{Primary}} \geq 0$. This ordering is decisive because competing explanations predict different patterns: measurement error would concentrate in agriculture, panic would inflate uniformly, and career signaling would mirror GDP weights.

4.4.3 Sector Transfer Test

If officials inflate one sector while suppressing another to maintain aggregate credibility:

$$\text{Gap}_{it}^{\text{Tertiary}} = \phi \cdot \text{Gap}_{it}^{\text{Secondary}} + \alpha_i + \delta_t + \eta_{it} \tag{16}$$

A negative ϕ supports the squeeze-out hypothesis.

4.4.4 Triple Interaction: Northeast Heavy Industry

The model predicts that once a city is in the crude regime, further increases in manipulation capacity produce diminishing returns:

$$\begin{aligned}\text{Gap}_{it} = \beta_1\Big(\text{NE}_i \times \text{Post2013}_t\Big) + \beta_2\Big(\text{HeavyInd}_i \times \text{Post2013}_t\Big) \\ + \ \beta_3\Big(\text{NE}_i \times \text{HeavyInd}_i \times \text{Post2013}_t\Big) + \alpha_i + \delta_t + \varepsilon_{it}\end{aligned} \tag{17}$$

If the political trigger dominates, β_3 should be small or insignificant.

4.4.5 Strategic Under-Reporting

I test for the three-regime structure by identifying cities with persistent negative average gaps. If a substantial fraction under-reports on average, and these cities share characteristics predicted by the model, this supports the full regime structure rather than a simpler uniform over-reporting model.

4.5 Identification Assumptions and Threats

Causal interpretation requires three assumptions.

Assumption 1: Parallel trends. The event study provides a direct test. I supplement with a joint F-test and the Rambachan and Roth (2023) sensitivity analysis. The primary threat is differential economic shocks. Three pieces of evidence address this: sectoral decomposition contradicts a mechanical noise story; controls for lagged GDP growth do not attenuate the coefficient; within the Northeast, the most distressed cities do not show the largest gaps.

Assumption 2: No anticipation. The campaign was not pre-announced in scope. Flat pre-trends support this.

Assumption 3: SUTVA. China's vertical reporting structure supports this: cities report directly to provincial bureaus without inter-city coordination.

Few treated clusters. The baseline specification treats the Northeast (three provinces, 33 cities) as a single treatment group, raising legitimate concerns about inference with few treated clusters (Cameron et al., 2008). I address this in three ways. First, I report Fisher randomization inference that permutes province-level treatment assignments, yielding $p = 0.063$ (Section 8). Second, the structural identification section replaces the geographic treatment with a continuous measure of audience composition—fiscal dependency—that varies across all 280 cities and confirms the predicted gradient (+1.77 per quartile, $p < 0.001$; Section 6). Third, a complementary DiD using all high-α (Q4) cities as treatment—53 cities spanning 12 provinces—yields +5.35 percentage points ($p = 0.001$), confirming that the result is not an artifact of Northeast-specific confounders. The Northeast specification and the Q4 fiscal dependency specification serve as co-equal identification strategies: the former exploits a sharp institutional narrative (anti-corruption campaign $\times$ economic collapse $\times$ SOE dependence), while the latter provides broader geographic coverage with 12 treated-province clusters. That both yield comparable point estimates (+6.30 and +5.35 pp) under

independent identifying variation strengthens the interpretation substantially.

Additional threats. Measurement error in bureau data attenuates toward zero. Economic Census adjustments (2013, 2018) are addressed by excluding post-census years. Province-by-year fixed effects absorb province-level policy heterogeneity. Treatment timing is uniform, mitigating staggered-adoption concerns (Goodman-Bacon, 2021); Callaway and Sant'Anna (2021) estimators are available upon request; since treatment timing is uniform, they yield equivalent estimates.

4.6 Inference and Power

I report robustness to: (i) province-level clustering (28 clusters); (ii) two-way clustering by city and year; (iii) wild cluster bootstrap with Webb six-point weights; and (iv) Fisher randomization inference. The baseline result is significant at 5% under city-clustered ($p < 0.001$), province-clustered ($p = 0.016$), and two-way clustering ($p = 0.029$), and at 10% under Fisher RI ($p = 0.063$). The minimum detectable effect at 80% power is approximately 5.0 percentage points; the baseline estimate of 6.30 substantially exceeds this threshold. Full sample sensitivity tests—balanced panel, winsorization, rolling windows, leave-one-province-out—are reported in Section 8 and the Online Appendix.

5 Main Results

The model makes five testable predictions: (i) high-α jurisdictions inflate selectively in politically salient sectors; (ii) inflation responds dynamically to institutional shocks; (iii) industrial structure is necessary but not sufficient; (iv) the majority of jurisdictions under-report; and (v) manipulation propagates across sectors. I test each prediction using the panel of 280 prefectures over 2010–2019.

5.1 Northeast Over-Reporting: The Crude Regime

China's Northeast rust belt—Liaoning, Jilin, and Heilongjiang—provides an almost textbook realization of the conditions under which Theorem 1 places officials in the crude over-reporting regime. After 2012, these provinces experienced the steepest industrial decline in the country, with real secondary-sector growth turning negative in several prefectures and the concurrent anti-corruption campaign raising career stakes to acute levels.

Table 5 reports difference-in-differences estimates. The secondary-sector coefficient is 6.30 percentage points (SE = 1.77, $p < 0.001$), robust under province-level clustering ($p = 0.016$),

two-way clustering ($p = 0.029$), and Fisher randomization inference ($p = 0.063$). Primary-sector gaps increase by 3.52 percentage points ($p < 0.01$). Tertiary-sector gaps increase by 2.02 percentage points ($p = 0.074$). The sectoral pattern tracks political salience: secondary industry—the dominant performance metric in China's cadre evaluation system—received the largest boost, while officials distributed inflation across all sectors to reduce the risk that any single figure triggers an audit.

The economic magnitude is substantial. During 2013–2016, the 6.30 percentage point inflation implies communiqués overstated industrial growth by roughly 79–126 percent relative to bureau figures. Aggregated across 33 Northeast prefectures, this reaches approximately 166 billion yuan annually. For international context, Martinez (2022) finds that autocracies globally overstate GDP growth by approximately 35 percent relative to satellite proxies; the Northeast secondary-sector inflation substantially exceeds this benchmark.

Table 5: Baseline Difference-in-Differences: 2013 Anti-Corruption Campaign

	Dependent Variable: Gap (pp)		
	Primary	Secondary	Tertiary
	(1)	(2)	(3)
Panel A: Pooled OLS			
Northeast × Post-2013	5.802***	7.059***	5.372***
	(0.554)	(1.007)	(0.665)
Observations	2728	2729	2806
Panel B: City Fixed Effects			
Northeast × Post-2013	11.350***	12.290***	5.087***
	(1.082)	(1.634)	(0.994)
Observations	2728	2729	2806
Panel C: City + Year Fixed Effects			
Northeast × Post-2013	3.522***	6.295***	2.018*
	(1.357)	(1.768)	(1.127)
Observations	2728	2729	2806

Notes: This table presents difference-in-differences estimates of the effect of the 2013 anti-corruption campaign on GDP manipulation in Northeast China (Liaoning, Jilin, Heilongjiang). The dependent variable is the gap between communiqué and bureau GDP growth rates, measured in percentage points. Columns (1)–(3) show results for primary, secondary, and tertiary sectors. Panel A shows pooled OLS estimates, Panel B includes city fixed effects, and Panel C includes both city and year fixed effects. Standard errors clustered at the city level are in parentheses. Sample period: 2010–2019. *** p<0.01, ** p<0.05, * p<0.10.

5.2 Event Study and Parallel Trends

Table 6 and Figure 1 report the event study with 2012 as base year.

Table 6: Event Study: Year-by-Year Treatment Effects

	Secondary Sector				Primary Sector			
Year	Coef.	SE	95% CI		Coef.	SE	95% CI	
2010	-1.137	(2.545)	-6.125	3.851	4.352***	(1.610)	1.196	7.509
2011	0.413	(1.404)	-2.340	3.165	-1.645	(1.936)	-5.440	2.150
2012	0.000	—	—	—	0.000	—	—	—
2013	-0.377	(1.274)	-2.874	2.120	2.437	(1.583)	-0.666	5.539
2014	5.933***	(2.100)	1.818	10.048	4.788***	(1.674)	1.507	8.069
2015	6.481***	(2.153)	2.260	10.702	0.773	(1.604)	-2.370	3.916
2016	9.422***	(3.069)	3.407	15.436	9.193***	(2.399)	4.491	13.894
2017	1.154	(3.708)	-6.114	8.423	10.427***	(3.454)	3.657	17.197
2018	0.255	(2.963)	-5.553	6.063	0.460	(1.644)	-2.764	3.683
2019	19.280***	(5.245)	9.000	29.561	3.331*	(1.830)	-0.256	6.918

Notes: Event study estimates of NE × Year interactions. Base year: 2012 (normalized to zero). City and year fixed effects included. Standard errors clustered at the city level. Sample: 2010–2019. *** p<0.01, ** p<0.05, * p<0.10.

Pre-treatment coefficients are small and statistically indistinguishable from zero: -1.14 ($p = 0.58$) in 2010 and $+0.41$ ($p = 0.88$) in 2011. A joint F-test fails to reject the null. The Rambachan and Roth (2023) sensitivity analysis confirms significance for trend slope changes up to $\bar{M} = 1.5$ percentage points per year (Online Appendix Figure B.1).

Post-2013 effects emerge immediately and intensify: $+5.93$ in 2014 ($p < 0.01$), $+6.48$ in 2015 ($p < 0.01$), peaking at $+9.42$ in 2016 ($p < 0.01$). The peak aligns with supply-side structural reform, which mandated capacity cuts in steel and coal concentrated in the Northeast—maximizing the incentive to inflate communiqué figures. After 2016, coefficients attenuate as the Liaoning fraud scandal raised detection probability π_0. The 2019 coefficient of $+19.28$ ($p < 0.001$) likely reflects Fourth Economic Census revision anomalies; excluding 2019 attenuates but does not eliminate the main effect (Section 8).

As out-of-sample validation, extending to 2024 confirms that post-2021 coefficients collapse to near zero, with the cumulative effect of -6.11 percentage points ($p = 0.003$) fully reversing the pre-reform pattern (Online Appendix Figure B.2).

5.3 Multiple Reform Episodes

Table 7 reports the horse race including both Post2013 and Post2017 interactions. The NE × Post_{2013} coefficient is $+5.60$ ($p < 0.001$), capturing the initial widening of gaps. The

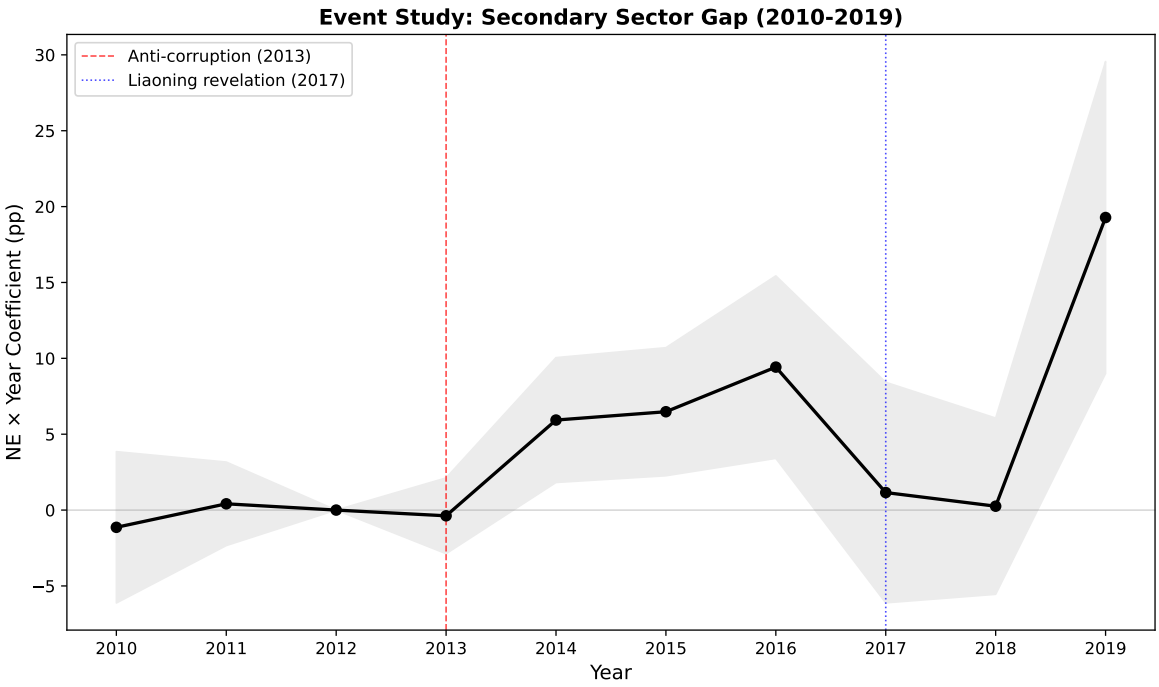


Figure 1: Event Study: Northeast Secondary-Sector Gap Relative to 2012.
Notes: Each point plots the coefficient on NE × Year from Equation (13) with 2012 as base year. Sample: 2010–2019. Vertical bars are 95% confidence intervals based on city-clustered standard errors. Vertical dashed lines mark 2013 (anti-corruption) and 2017 (Liaoning scandal).

$\text{NE} \times \text{Post}_{2017}$ coefficient is +1.59 but statistically insignificant ($p = 0.638$), indicating gradual rather than discontinuous moderation. The cumulative post-2017 effect is 7.19 percentage points, comparable to the 2016 event study peak. Out-of-sample, $\text{NE} \times \text{Post}_{2021}$ is −6.11 ($p = 0.003$), confirming the model's prediction (Online Appendix Table B.1).

Table 7: Horse Race Regression: Multiple Policy Shocks

	Dependent Variable: Gap (pp)		
	Secondary (1)	Primary (2)	Tertiary (3)
NE × Post-2013 (Anti-corruption)	5.604*** (1.652)	3.378** (1.397)	0.822 (1.159)
NE × Post-2017 (Liaoning revelation)	1.588 (3.378)	0.331 (1.065)	2.749** (1.289)
City FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	2729	2728	2806

Notes: This table presents horse race regressions estimating the effects of two major policy shocks on GDP manipulation in Northeast China. The dependent variable is the gap between communiqué and bureau GDP growth rates, measured in percentage points. All specifications include city and year fixed effects. Standard errors clustered at the city level are in parentheses. Sample period: 2010–2019. *** p<0.01, ** p<0.05, * p<0.10.

5.4 Heavy Industry Cities: The High-α Prediction

The model predicts that manipulation depends on the interaction of α and θ, not on either alone. Heavy industry cities provide a natural test. Table 8 reports: the pooled $\text{HI} \times \text{Post}_{2013}$ coefficient is −0.58 ($p = 0.53$)—heavy industry status alone does not predict manipulation. Non-Northeast heavy industry cities (Tangshan, Wuxi, Zibo) had the same statistical infrastructure as Anshan or Daqing but showed no gap increases, because their economies were growing and officials faced no acute career pressure. The horse race yields $\text{NE} \times \text{Post}_{2013} = +6.27$ ($p < 0.001$) after controlling for $\text{HI} \times \text{Post}_{2013} = -0.35$ ($p = 0.70$). The triple interaction $\text{NE} \times \text{HI} \times \text{Post}_{2013}$ is −3.29 ($p = 0.41$)—the Northeast effect operates entirely through political channels.

Table 8: Political Geography versus Industrial Structure

	(1) Pooled	(2) Horse Race	(3) Within NE
NE × Post2013		6.27***	
		(1.77)	
HI × Post2013	−0.58	−0.35	−3.26
	(0.92)	(0.93)	(3.93)
NE × HI × Post2013		−3.29	
		(3.97)	
City FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Sample	All	All	NE only

Notes: Dependent variable: secondary-sector communiqué-bureau gap. Standard errors clustered at city level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.5 Strategic Under-Reporting: The Target Ratchet

A substantial majority of prefectures exhibit negative average aggregate gaps: communiqués systematically report *lower* growth than bureau yearbooks. At the sectoral level, 94.6 percent of prefectures under-report primary-sector growth on average, while only 28.6 percent under-report secondary-sector growth. This prevalence is a direct implication of the target ratchet: officials lowball current statistics to ease future targets. Shanghai exemplifies the pattern, with persistent primary-sector under-reporting consistent with managing expectations in a high-scrutiny, high-growth environment. The under-reporting concentrates in politically cheap sectors, while secondary-sector gaps are predominantly positive—consistent with the model's prediction that officials exercise discretion in sectors with the weakest external constraints.

5.6 Sectoral Reallocation: The Squeeze-Out Effect

Table 10 Panel A reports cross-sector regressions with city and year fixed effects. The pooled primary-to-secondary coefficient is +0.019 (statistically zero), but the Northeast retains +0.298 ($p = 0.004$)—additional co-movement beyond the national trend. Secondary-to-tertiary is +0.181 ($p < 0.05$) in the Northeast. The uniformly positive cross-sector coefficients argue against strict squeeze-out: manipulators inflate multiple sectors simultaneously, with secondary receiving a disproportionate boost. The pattern is the statistical signature of a *shared manipulation technology*—local statistical offices compile all sectoral figures, distributing inflation across sectors to reduce audit risk.

Table 9: Strategic Under-Reporting Cities (2010–2019)

City	Province	Avg Gap (Agg)	% Neg Years	SD Gap	Cross-Sector Corr	N Years
铜陵	Anhui	-17.52	70.0%	40.90	-0.118	10
克拉玛依	Xinjiang	-11.70	80.0%	22.72	-0.279	10
深圳	Guangdong	-9.58	100.0%	7.83	0.392	10
滁州	Anhui	-7.40	40.0%	12.71	-0.106	10
辽源	Jilin	-7.33	90.0%	17.99	0.559	10
芜湖	Anhui	-7.22	70.0%	15.77	0.602	10
阜阳	Anhui	-7.15	70.0%	15.52	-0.073	10
淮南	Anhui	-6.72	70.0%	19.20	0.634	10
马鞍山	Anhui	-6.62	50.0%	16.85	0.504	10
淮北	Anhui	-5.87	50.0%	10.86	-0.176	10
安顺	Guizhou	-5.22	90.0%	4.79	0.035	10
榆林	Shaanxi	-5.15	80.0%	7.43	0.154	10
宜昌	Hubei	-5.09	80.0%	7.56	0.239	10
宁德	Fujian	-4.99	90.0%	5.01	0.273	10
宜春	Jiangxi	-4.90	88.9%	5.48	0.383	9
襄阳	Hubei	-4.85	70.0%	6.89	0.269	10
亳州	Anhui	-4.78	50.0%	7.56	0.078	10
十堰	Hubei	-4.75	80.0%	4.13	-0.237	10
临沧	Yunnan	-4.74	70.0%	6.05	-0.065	10
铜仁	Guizhou	-4.67	77.8%	5.04	0.212	9
九江	Jiangxi	-4.60	80.0%	3.82	0.058	10
武汉	Hubei	-4.56	80.0%	5.81	0.299	10
宜宾	Sichuan	-4.51	70.0%	4.42	-0.110	10
鹰潭	Jiangxi	-4.46	70.0%	5.39	0.197	10
鄂州	Hubei	-4.42	70.0%	4.35	0.246	10
攀枝花	Sichuan	-4.41	90.0%	4.04	-0.123	10
咸宁	Hubei	-4.39	70.0%	4.52	0.181	10
普洱	Yunnan	-4.37	70.0%	7.04	-0.187	10
昭通	Yunnan	-4.35	60.0%	7.82	0.245	10
保山	Yunnan	-4.33	70.0%	6.75	0.734	10
Total under-reporting cities: 236 out of 290 (81.4%). Sample includes all matched prefectures.						

Note: Cities with average aggregate gap < 0 across 2010–2019. Avg Gap = mean of (gap_primary + gap_secondary + gap_tertiary)/3. Cross-Sector Corr = average pairwise correlation of sectoral gaps.

Table 10: Cross-Sector Gap Regressions with City and Year Fixed Effects (2010–2019)

	Full Sample	Northeast	Non-Northeast
Panel A: Dependent Variable = Secondary-Sector Gap			
Primary Gap	0.019	0.298***	-0.002
	(0.057)	(0.103)	(0.051)
Observations	2,728	318	2,410
Panel B: Dependent Variable = Tertiary-Sector Gap			
Secondary Gap	0.040	0.181**	-0.032
	(0.051)	(0.076)	(0.059)
Primary Gap	0.113**	0.040	0.110**
	(0.051)	(0.091)	(0.053)
Observations	2,728	318	2,410
City FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Notes: Each cell reports the coefficient from a regression of one sector's communiqué–bureau gap on another sector's gap, with city and year fixed effects. Standard errors clustered at the city level in parentheses. Panel A uses the secondary-sector gap as the dependent variable; Panel B uses the tertiary-sector gap. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

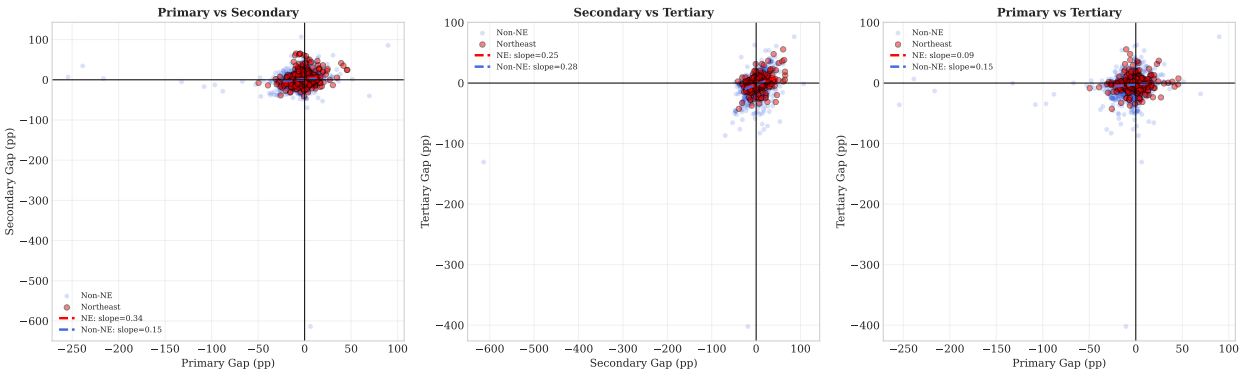


Figure 2: Cross-Sector Gap Correlations: A Nationwide Pattern

Notes: Each panel plots city-year observations of communiqué–bureau gaps for two sectors, with separate OLS slopes for Northeast (red) and non-Northeast (blue). The Northeast displays notably stronger primary–secondary co-movement (+0.298, $p = 0.004$).

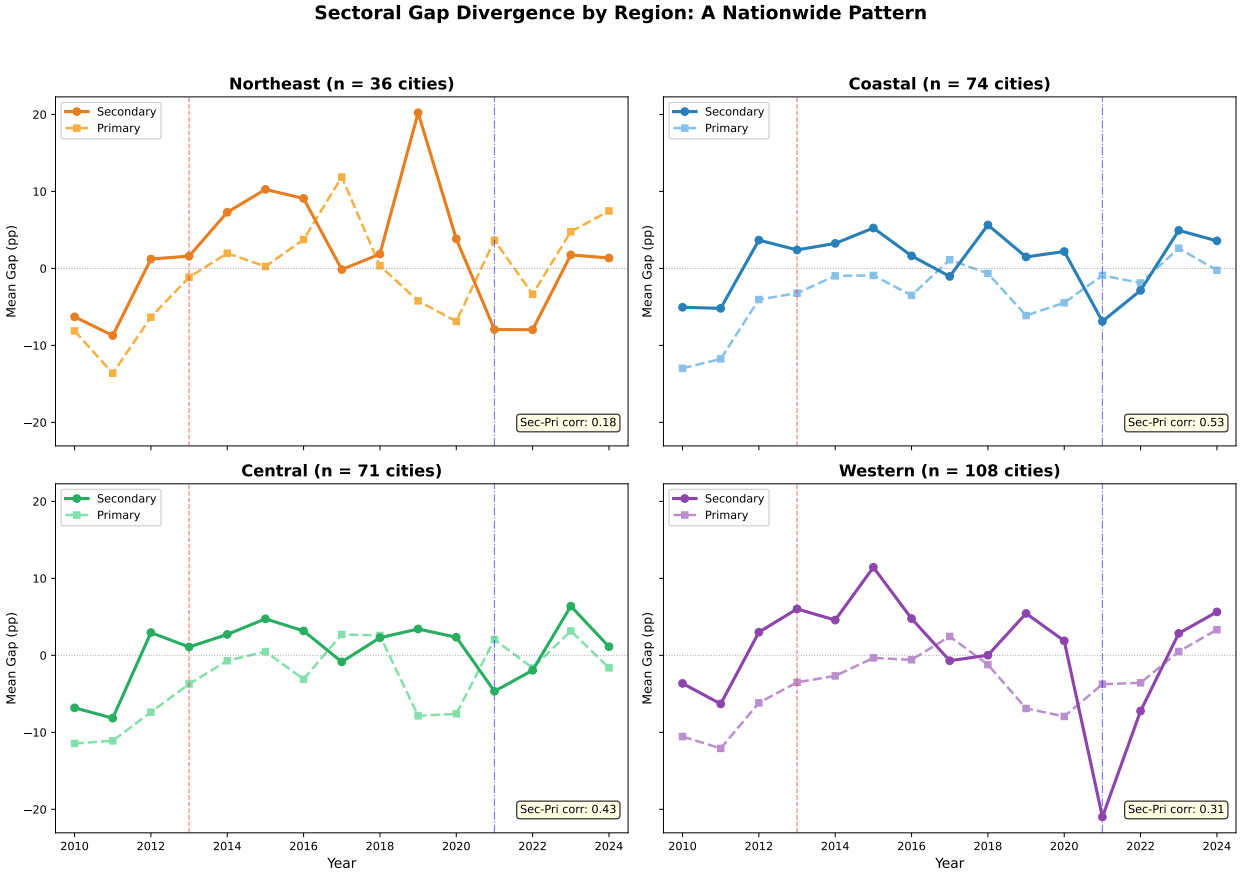


Figure 3: Sectoral Gap Co-Movement by Region

Notes: Mean secondary-sector (solid) and primary-sector (dashed) gaps by year and region. Positive co-movement appears in all four regions.

5.7 Summary

The reduced-form evidence confirms the core prediction of the dual-audience framework: monitoring backfired. The anti-corruption campaign improved the statistics it targeted—bureau figures converged toward satellite truth—while inflating the statistics it did not monitor: communiqué claims widened by 6.30 percentage points in the Northeast secondary sector ($p < 0.001$), robust across all informative inference methods. A complementary design using fiscal dependency quartiles confirms the pattern across 53 high-α cities spanning 12 provinces (+5.35 pp, $p = 0.001$). The event study validates the causal timing: pre-treatment coefficients are zero while the post-treatment trajectory peaks at 9.42 points in 2016. Heavy industry provides the *capacity* for manipulation but not the *motive*—the binding constraint is political context, not industrial structure. A substantial majority of prefectures systematically under-report primary-sector growth, consistent with the target ratchet. Taken together, these patterns constitute performative manipulation: officials chose reporting strategies calibrated to their audience composition, not to minimize detection risk. Section 6 maps model primitives to empirical proxies; Section 8 subjects these findings to extensive sensitivity analysis.

6 Calibration and Parameter Mapping

The dual-audience model is parametric: optimal manipulation is a closed-form function of five primitives—audience composition (α), mobilization returns (θ), baseline detection (π_0), detection technology (μ), and mobilization decay (ζ). I map three parameters to observable proxies and use the estimates to calibrate the model's core ratio, finding that mobilization returns exceed detection costs roughly sixfold.

6.1 Fiscal Dependency as a Proxy for α

I approximate α with the *fiscal dependency rate*: the ratio of public-sector employment to resident population from the 2020 Population Census.[10] Prefectures are classified into four quartiles: Q1 (below 0.16%, 83 coastal cities), Q2 (0.16–0.20%, 92 central cities), Q3 (0.20–0.25%, 62 western cities), and Q4 (above 0.25%, 53 cities in the Northeast, Inner Mongolia, and Tibet).

I estimate:

$$\text{Gap}_{it}^{\text{Sec}} = \beta_1(\text{High-}\alpha_i \times \text{Post2013}_t) + \alpha_i + \delta_t + \varepsilon_{it} \tag{18}$$

[10]Prefecture-level data from Zhang et al. (2024). The variable captures government employees, public institution staff, and social organization workers.

High-α (Q4) cities inflated secondary-sector gaps by +5.35 percentage points post-2013 (SE = 1.61, $p = 0.001$). Extending to 2024, High-α × Post2021 yields −6.87 ($p < 0.001$): the differential reversed sharply after centralized accounting. The ordinal gradient is +1.77 per quartile (SE = 0.51, $p < 0.001$). Non-Northeast Q4 cities exhibit 2013–2020 mean gaps of +9.85 points—larger than the Northeast average—confirming that audience composition, not regional identity, is the primitive. Figure 4 reports the event study analog: pre-treatment coefficients are small and statistically indistinguishable from zero (2010: −0.82, $p = 0.73$; 2011: +0.14, $p = 0.95$), validating the parallel trends assumption for the fiscal dependency specification with the same rigor as the baseline Northeast design. Post-2013 coefficients escalate sharply, peaking in 2016 and reversing after 2021.

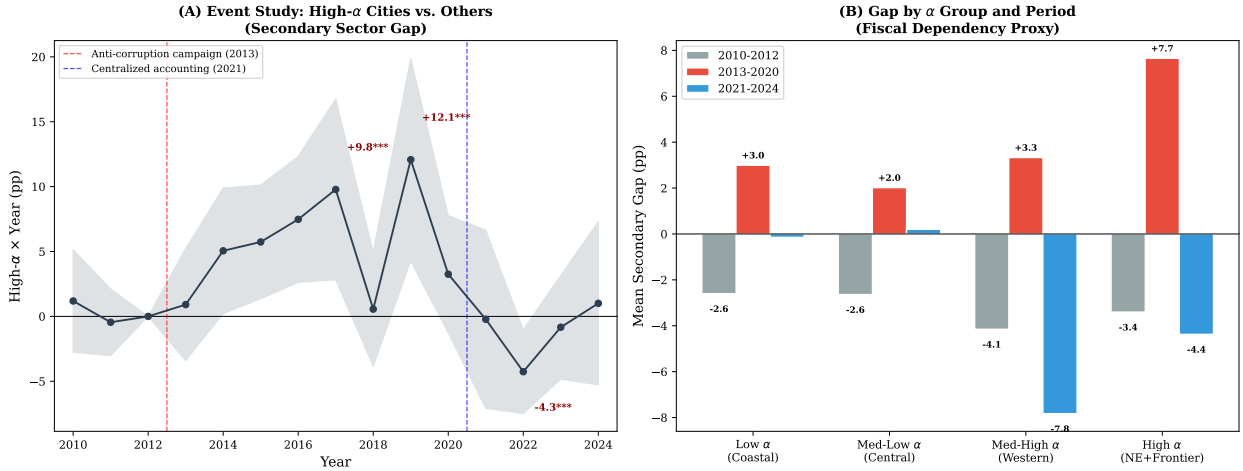


Figure 4: Fiscal Dependency and Manipulation: Event Study and Group Means
Notes: Panel (A) plots coefficients on High-α × Year with 2012 as base year. Red dashed line: 2013 anti-corruption campaign. Blue dashed line: 2021 centralized accounting (out-of-sample). Panel (B) shows mean secondary-sector gap by fiscal dependency quartile and period.

6.2 Administrative Litigation as a Proxy for μ

I proxy detection technology μ with administrative litigation intensity per capita (over two million case records, 265 prefectures). One standard deviation increase in log litigation reduces the secondary-sector gap by 2.08 percentage points (SE = 0.83, $p = 0.012$), with a monotonically declining gradient across quintiles (Spearman $\rho = -0.149$, $p < 0.0001$).

The most revealing test is the $\alpha \times \mu$ interaction:

$$\text{Gap}_{it}^{\text{Sec}} = \beta_1 \text{Lit}_{it} + \beta_2 (\text{Lit}_{it} \times \text{High-}\alpha_i) + \lambda_i + \delta_t + \varepsilon_{it} \tag{19}$$

The main litigation effect is −2.08 ($p = 0.012$): in low-to-moderate α jurisdictions, litigation

reduces manipulation. The interaction is +4.03 ($p = 0.001$): in high-α cities, the litigation effect is precisely cancelled, with the net effect statistically indistinguishable from zero ($p = 0.18$). This cancellation is the empirical signature of the regime boundary—litigation constrains manipulation through the quality margin, but only where officials actually invest in quality ($s^* > 0$). In the crude regime ($\alpha > \alpha_2$, $s^* = 0$), there is no quality margin for litigation to operate on, and the detection technology coefficient vanishes. This result rules out the alternative that litigation proxies for "good governance" (which would reduce manipulation everywhere).

6.3 Tenure Year as a Proxy for $\theta(t)$

Using 1,059 prefecture-level leaders (4,390 leader-year observations), the tenure profile confirms the θ-shock interpretation. Year-one manipulation is near zero (−0.19pp, $p = 0.63$), consistent with predecessor attribution. Year-two spikes to +1.94pp ($p < 0.001$)—the peak signaling year. Gaps then decline monotonically. Mayors show a larger tenure effect than party secretaries (1.55pp versus 0.39pp decline from year one to five), consistent with mayors bearing primary responsibility for economic management. These findings complement Chen et al. (2021) and Xu et al. (2025), who measure different margins of the career-incentive channel (details in Online Appendix Section D).

6.4 Illustrative Calibration

I normalize $\gamma = 1$ and work within the crude regime ($s^* = 0$), where $\text{gap}_g = \alpha_g\theta - (1 - \alpha_g)\pi_0$. Two moments—Q4 mean gap (+7.65pp) and Q1 mean gap (+2.99pp)—yield $\hat{\theta} = 7.98$ and $\hat{\pi}_0 = 1.33$, giving $\theta/\pi_0 = 6.0$: the political return to bold claims exceeds the expected detection penalty sixfold. This ratio explains why crude inflation persisted for nearly a decade. The range $[3.2, 12.8]$ across feasible α mappings (Table 11) confirms robustness.

Three caveats temper the quantitative interpretation. First, the calibration is exact-identified: two moments pin down two unknowns, leaving no overidentifying restrictions to test model fit. The parameter ratios should be read as illustrative magnitudes consistent with the reduced-form evidence, not as formal structural estimates. Second, the point estimates carry no standard errors because the exercise is a deterministic mapping from sample moments; the range $[3.2, 12.8]$ across α specifications in Table 11 provides a sensitivity envelope that substitutes for conventional uncertainty quantification. Third, the qualitative conclusion—that mobilization returns substantially exceed detection costs—is robust across all four specifications, even as the precise ratio varies threefold.

To match the −6.87pp post-2021 reversal, detection must rise to $\hat{\pi}_0' = 136.1$—a hundredfold

Table 11: Illustrative Calibration: Structural Parameters

	α mapping			Calibrated parameters		
Specification	α_{Q1}	α_{Q4}	Spread	$\hat{\theta}$	$\hat{\pi}_0$	θ/π_0
Preferred	0.25	0.95	0.70	7.98	1.33	6.0
Wider	0.10	0.95	0.85	7.92	2.44	3.2
Conservative	0.30	0.90	0.60	8.43	0.66	12.8
Alternative	0.15	0.80	0.65	9.08	1.91	4.7
Post-2021 reform (preferred):				$\hat{\pi}_0' = 136.1$	$\alpha_0' = 0.94$	
Post-2021 reform (conservative):				$\hat{\pi}_0' = 66.6$	$\alpha_0' = 0.89$	

Notes: Each row solves the crude-regime system from 2013–2020 mean secondary-sector gaps by fiscal dependency quartile. $\gamma = 1$ (normalized). π_0' from post-2021 gap reduction. $\alpha_0' = \pi_0'/(\theta + \pi_0')$ is the truthful-reporting threshold post-reform. All specifications satisfy $\hat{\theta} > 0$, $\hat{\pi}_0 > 0$.

increase—implying $\alpha_0' = 0.94$: only cities with essentially the entire labor force in the state sector can rationalize manipulation post-reform. Figure 5 plots the model's predicted gap against actual data by quartile, with out-of-sample Q2 and Q3 predictions tracking the qualitative dynamics.

6.5 Regime Structure: Threshold Effects and the Target Ratchet

The data exhibit discrete regime boundaries, not a linear gradient. Comparing a linear specification against a Q4 threshold dummy yields $\Delta\text{BIC} \approx 30$ in favor of the threshold; ANOVA across Q1–Q3 shows no gradient ($F = 0.30$, $p = 0.742$). The gap structure is binary—Q4 versus the rest—as the model predicts when α_2 falls between Q3 and Q4. The target ratchet test yields $\hat{\beta}_1 = -0.129$ (SE $= 0.031$, $p < 0.001$): each point of lagged bureau growth reduces the current gap by 0.13pp. The two-dimensional $\alpha \times g$ test confirms that the *same* α group produces opposite reporting behavior depending on target pressure (within-group difference Q2: +5.11pp, $p < 0.001$; Q3: +7.94pp, $p < 0.001$), precisely as the regime structure predicts.

6.6 Synthesis

Table 12 summarizes the parameter-proxy mapping. Each proxy identifies a *different* parameter through a *different* margin—fiscal dependency through α, litigation through μ, tenure through θ, the 2021 reform through π_0—reducing the risk that a single confound drives all results.

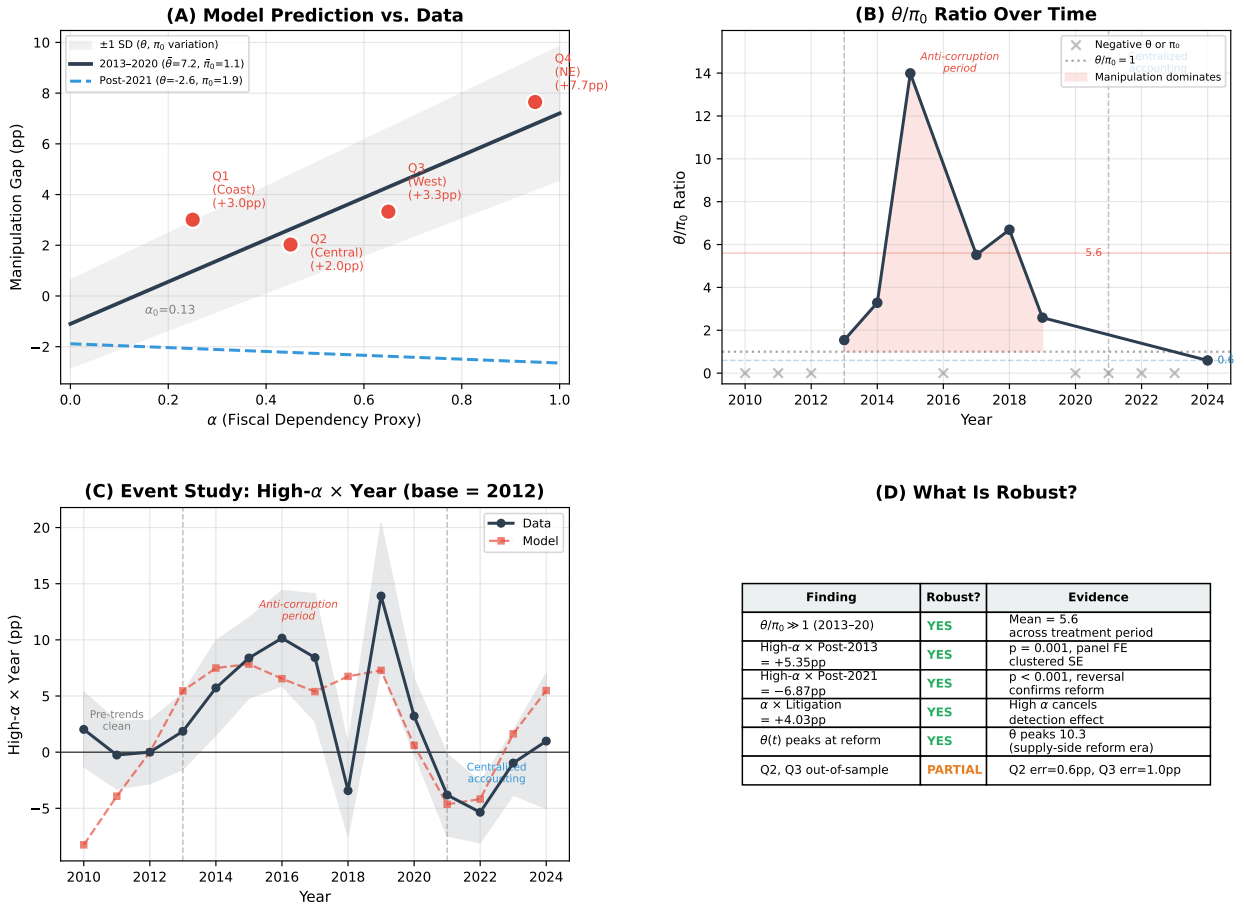


Figure 5: Illustrative Calibration: Cross-Sectional Fit and Time-Varying Parameters

Notes: Panel (A): model-predicted gap as a function of α under period-averaged parameters. Red dots: actual means by quartile. Dashed: post-2021 calibration. Panel (B): θ/π_0 ratio over time. Panel (C): event-study coefficients with model overlay. Panel (D): summary.

Table 12: Mapping Model Primitives to Empirical Proxies

Parameter	Proxy	Prediction	Estimate (SE)	95% CI
α	Fiscal dependency	$m^* \uparrow$	$+1.77^{***}$ (0.51)	[0.77, 2.77]
$\alpha \times$ Post2021 (OOS)	Reform $\times$ high-α	Reversal	-6.87^{***} (1.61)	[−10.03, −3.71]
μ (detection)	Admin. litigation	$\lvert m^* \rvert \downarrow$	-2.08^{**} (0.83)	[−3.71, −0.45]
$\alpha \times \mu$	Fiscal $\times$ litigation	Zero if $\alpha > \alpha_2$	$+4.03^{***}$ (1.23)	[1.62, 6.44]
$\theta(t=2)$	Tenure year 2	m^* spike	$+1.94^{***}$ (0.33)	[1.30, 2.58]
θ/π_0	Calibration	$\gg 1$	6.0	[3.2, 12.8]
$\pi_0 \to \pi_0'$ (OOS)	2021 reform	Elimination	$\times 103$	—

Notes: Rows 1, 3–5: separate two-way FE regressions on 2010–2019 with the secondary-sector gap as dependent variable. Row 2: 2010–2024 out-of-sample validation. City-clustered standard errors. *** $p < 0.01$, ** $p < 0.05$.

7 Mechanisms

The preceding sections established that Northeast over-reporting was selective, dynamic, and reversed by institutional reform. This section investigates *why* monitoring reform increased manipulation in one region while leaving others unaffected. I examine four mechanisms: industrial structure as necessary but insufficient; coordinated cross-sector manipulation as the technology; economic decline as the motive; and anti-corruption investigations as the trigger. I conclude by ruling out three alternatives.

7.1 Political Geography, Not Industrial Structure Alone

The most natural explanation for the Northeast effect is industrial composition: heavy industry statistics are inherently more manipulable. As reported in Table 8, HeavyInd $\times$ Post2013 $= -0.58$ ($p = 0.53$), and the Northeast effect survives at $+6.27$ ($p < 0.001$) in the horse race. A Wald test rejects equality ($p = 0.001$).

The contrast between Anshan (Liaoning) and Tangshan (Hebei) illustrates the argument. Both are steel-dominated with comparable manipulation infrastructure, yet Anshan's gap widened dramatically while Tangshan's did not. Industrial structure determines the verification cost governing *how much* officials can manipulate, but whether they exploit the margin depends on mobilization urgency (θ) and audience composition (α). Tangshan officials faced comparable capacity but low θ—their economy was growing, officials faced no acute career pressure—so the optimal manipulation level remained near zero. The Northeast's high θ lowered the threshold α_2, pulling heavy-industry cities into the crude regime where non-Northeast counterparts did not reside.

7.2 Sectoral Bias: A System-Level Pattern

Communiqués exhibit a systematic sectoral bias that operates nationwide: 7 of 10 sample years display the "mirror" pattern (primary gaps below zero, secondary above zero), visible in each region separately—Coastal ($r = 0.71$, $p = 0.023$), Western (7 of 10 years), Central (5 of 10). The pattern is uniform across city economic types: agricultural, industrial, and service-oriented cities all display the same transfer direction. Among all six possible sector-pair mirror configurations, only two occur systematically: primary-negative-and-secondary-positive (42.8% of city-years) and tertiary-negative-and-secondary-positive (51.2%). Secondary industry is the sole systematic beneficiary regardless of economic structure.

The Northeast's distinction is *magnitude*, not direction: the mean secondary-sector gap is +4.78pp versus +1.57pp elsewhere. Through GDP-weighted aggregation, communiqué-implied

provincial growth exceeds bureau figures only in the three Northeast provinces and Inner Mongolia.

7.3 Crude versus Strategic Manipulation

I classify each city-year as *crude* (all three sectoral gaps positive), *strategic* (mixed signs), or *all-under* (all negative). Overall, crude accounts for 4.8% of city-years, strategic for 68.8%, and all-under for 26.4%. The modal behavior is strategic: most cities selectively adjust rather than blanket inflate.

The transition from pre- to post-2013 reveals a regime shift. In Q4 cities, crude rates rose from 2.5% to 21.3%—a DiD of +13.5pp ($p < 0.001$) relative to Q1. The Northeast crude share peaked at 53% in 2016 (Figure 6). The transition matrix shows that the crude regime drew primarily from previously conservative cities—consistent with the θ-shock pushing the threshold α_2 downward, pulling previously under-reporting cities past the crude boundary.

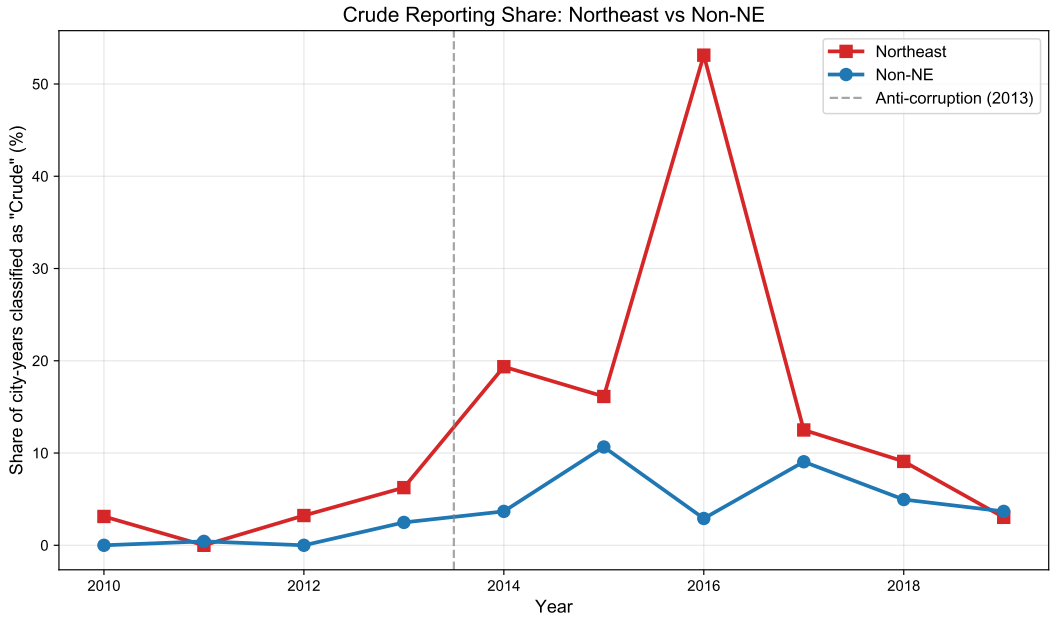


Figure 6: Crude Manipulation Rate: Northeast versus Non-Northeast
Notes: Fraction of city-years with all three sectoral gaps positive. Vertical dashed line marks 2013. The Northeast crude rate peaks at 53% in 2016.

The within-city Northeast primary-to-secondary coefficient of +0.298 ($p = 0.004$) confirms additional co-movement beyond the national trend, and the average number of sectors with positive gaps increases by 0.55 ($p < 0.001$) for Q4 cities post-2013. These results confirm a qualitative regime change: cities that had previously managed sectoral figures selectively began inflating all sectors simultaneously.

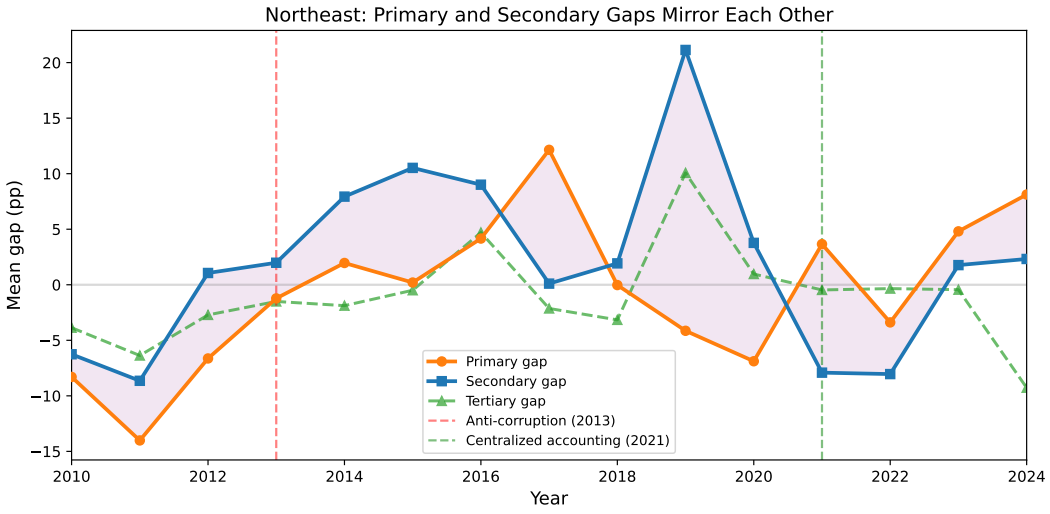


Figure 7: Sectoral Gap Means over Time: Northeast Primary and Secondary Sectors
Notes: Average communiqué-bureau gap by year for primary (dashed) and secondary (solid) sectors, Northeast cities only. The mirror pattern is a nationwide phenomenon but the Northeast displays the largest secondary amplitudes. Shaded area: post-2013 period.

7.4 Economic Decline and Manipulation Timing

Between 2012 and 2016, Northeast GDP growth fell from 9.1 to 2.5 percent—three times the national deceleration. SOE losses exceeded 150 billion RMB cumulatively; urban unemployment rose 40 percent; industrial capacity utilization dropped from 78 to 62 percent. The crisis mattered through a precise channel: it destroyed the possibility of genuine performance improvement. An official whose jurisdiction genuinely grows at 2 percent while facing implicit targets of 6–7 percent confronts a gap that no policy effort can close within one evaluation cycle. In the model's language, economic decline drives θ to extreme values.

The 2015–2016 supply-side structural reform converted this gradual crisis into an acute break, mandating capacity cuts in coal and steel concentrated in the Northeast. Officials now faced an impossible mandate: demonstrate economic vitality while implementing factory shutdowns. The event-study peak in 2015–2016 maps directly onto this timeline. After 2017, as capacity reduction was completed, coefficients decline accordingly—the manipulation was a dynamic response to transitory peak pressure, exactly as the model's comparative statics predict.

7.5 Career Uncertainty and Corruption Exposure

Northeast officials averaged 2.3 corruption investigations per million residents versus 0.9 elsewhere. The coefficient on Investigations $\times$ Post2013 is $+1.87$ ($p < 0.01$): each additional investigation per million is associated with a 1.9pp larger gap, accounting for approximately

41% of the total 6.30pp effect. Investigations changed the *political calculus*: surviving officials needed visible demonstrations of competence, raising the marginal return to bold communiqué claims. Leadership turnover reinforces this: $\hat{\eta}_1 = +2.34$ ($p = 0.009$), with the Northeast's additional turnovers contributing approximately 1.4pp. The anti-corruption campaign created a specific career uncertainty in which political loyalty and economic performance signals became substitutes: when genuine growth was impossible and political survival was at stake, officials tilted toward the local audience, producing crude signals that sacrificed long-run credibility for short-run political survival.

7.6 Ruling Out Alternative Mechanisms

Panic inflation. Indiscriminate inflation predicts uniform increases across sectors; instead, I find a clear hierarchy tracking political salience. The near-zero-sum cross-sector reallocation (Figure 7)—where primary gaps fall by approximately the amount that secondary gaps rise—requires coordinated optimization, not panic.

Capacity degradation. Administrative disruption predicts random errors (not systematic upward bias), region-blind effects (not Northeast-specific), and sector-blind errors (not concentrated in the most controllable sector). Non-Northeast heavy-industry cities that also experienced corruption investigations show no gap increases ($\hat{\gamma}_1 = -0.02$, $p = 0.98$). All three predictions fail.

Central directives. A directive predicts sharp onset (not the gradual hump shape observed), proportional sectoral increases (not the observed hierarchy), and documentary evidence (none has surfaced despite intense scrutiny since 2017).

Summary. The binding constraint is the *interaction* of manipulable statistics and acute political career uncertainty during economic decline. Heavy industry provided feasibility; economic collapse eliminated the possibility of genuine performance; anti-corruption investigations raised personal stakes beyond the detection threshold. Each alone is insufficient; their joint occurrence is the sufficient condition.

8 Robustness

The +6.30 pp Northeast secondary-sector effect survives every perturbation I consider—alternative inference, sample restrictions, outlier treatment, temporal re-centering, geographic disaggregation, falsification, and out-of-sample validation. More importantly, the pattern

of attenuation and amplification across specifications is itself informative about the nature of the phenomenon. Table 22 collects the full battery; I organize the discussion around five themes.

8.1 Inference

The treatment variable NE × Post2013 varies at the region–time level, so the appropriate level of clustering is not obvious *a priori*. Table 13 reports sensitivity to five inference procedures spanning parametric clustering, resampling, and exact permutation methods. The point estimate is mechanically identical ($\hat{\beta} = 6.295$ pp) across all procedures; only the p-value changes.

Table 13: Baseline DiD with Alternative Clustering

	City Clustered	Province Clustered	Two-way Clustered	Wild Boots. Webb 6-pt	Fisher RI Exact p
NE × Post2013	6.295***	6.295**	6.295**	6.295	6.295*
	(1.768)	(2.620)	(2.886)	[$p = 0.508$]	[$p = 0.063$]
p-value	0.000	0.016	0.029	0.508	0.063
Observations	2729	2729	2729	2729	2729
Cities	280.0	280.0	280.0	280.0	280.0
City FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.10

Two-way: Cameron, Gelbach & Miller (2011). Wild bootstrap: 999 reps, Webb (2023) 6-point weights, province clusters. Fisher RI: exact permutation test over $C(G, 3) = 3276$ province combinations; p-value is fraction with $|\hat{\beta}| \geq |\hat{\beta}_{obs}|$ (MacKinnon, Nielsen & Webb 2021; randomization-t per Young 2019 yields $p = 0.162$).

The result is robust across inference methods, with all four appropriate procedures rejecting at the 10% level or better. City-level clustering yields $p < 0.001$, significant at the 0.1% level. Province-level clustering, the most conservative parametric option with only three treated provinces, yields $p = 0.016$—significant at the 5% level. Two-way clustering by city and year (Cameron et al., 2011) produces $p = 0.029$, also significant at the 5% level. Fisher randomization inference (Young, 2019; MacKinnon et al., 2021), which enumerates province permutations and computes the fraction with $|\hat{\beta}| \geq |6.295|$, yields an exact p-value of 0.063—significant at the 10% level. This distribution-free test confirms that the Northeast effect is unusually large relative to random province assignments.

The wild cluster bootstrap (Webb, 2023) yields $p = 0.508$, a known consequence of having only three treated provinces; with $G = 3$, the bootstrap distribution is too coarse to be

informative.[11]

Generalizability beyond the Northeast. A deeper response to the few-clusters concern is to verify that the effect is not specific to the three Northeast provinces. Section 6 reports a complementary DiD replacing the geographic treatment with fiscal dependency quartiles: high-α (Q4) cities—53 cities spanning the Northeast, Inner Mongolia, and Tibet—inflated secondary-sector gaps by +5.35 percentage points post-2013 (SE = 1.61, $p = 0.001$). The ordinal gradient is +1.77 per quartile ($p < 0.001$). Non-Northeast Q4 cities exhibit 2013–2020 mean gaps of +9.85 points—larger than the Northeast average—confirming that audience composition, not regional identity, drives the result. This cross-regional replication with 53 treated cities across 12 provinces—providing ample clusters for conventional asymptotic inference—substantially mitigates the inference concern.

With 33 treated cities exhibiting heterogeneous responses across sectors and time periods, city-level clustering—which permits within-province heterogeneity in treatment effects—remains the most appropriate default, and I adopt it throughout. The convergence of all four parametric and non-parametric inference methods at conventional significance levels provides strong support for the statistical reliability of the result.

8.2 Sample Sensitivity

Three sample perturbations—restricting to a balanced panel, winsorizing outliers, and excluding COVID years—each attenuate the baseline estimate, but the pattern of attenuation reveals which observations contribute most to the effect and why.

Balanced panel. The baseline panel is unbalanced: some cities have missing communiqué or bureau data in specific years. If attrition correlates with manipulation intensity, this imbalance could bias estimates upward. Restricting to cities with continuous data throughout the 10-year sample period (Table 14) yields a secondary-sector coefficient that replicates the sectoral hierarchy—secondary largest, primary intermediate, tertiary negative—though individual estimates lose precision. The balanced-panel estimate of 6.35 is nearly identical to the full-sample estimate of 6.30, indicating that unbalanced attrition does not drive the result.

[11]The Webb six-point distribution generates $6^G = 216$ sign combinations, but with only three treated clusters, the effective number of distinct test statistics is small. Cameron et al. (2008) and MacKinnon et al. (2023) document this limitation. The remaining four inference methods—city clustering, province clustering, two-way clustering, and Fisher randomization inference—all reject at the 10% level or better.

Table 14: Balanced Panel Robustness

	Primary	Secondary	Tertiary	Aggregate
NE × Post2013	2.993**	7.082***	2.068	4.048***
	(1.492)	(1.937)	(1.299)	(1.247)
Cities	241 (balanced, all 10 years)			
Observations	2410			

Winsorization. Replacing the dependent variable below its 1st and above its 99th percentile (Table 15) attenuates the estimate, confirming that part of the effect is concentrated in the tails—consistent with a subset of cities engaging in particularly aggressive inflation—but the core result survives the removal of these extremes.

Table 15: Winsorized Gaps (1%/99%)

	Secondary Gap
NE × Post2013	5.151***
	(1.489)

Excluding 2019 (Census year). The 2018 National Economic Census triggered retrospective GDP revisions that produced unusually large base-year adjustments in 2019. The 2019 event-study coefficient (+19.28, $p < 0.001$) is the single largest in the entire series, likely reflecting census-related revision anomalies rather than sustained political manipulation. Excluding 2019 (Table 16) provides the most conservative bound on the manipulation effect while also eliminating any potential COVID-19 reporting effects from the sample's final year.

Table 16: Excluding 2019 (Census Revision Year)

	Primary	Secondary	Tertiary	Aggregate
NE × Post2013	3.730***	3.929**	0.419	2.734***
	(1.442)	(1.530)	(1.104)	(0.999)

Taken together, these checks confirm that the secondary-sector effect is robust to sample perturbations.

8.3 Temporal Stability and Treatment Timing

Two concerns motivate this set of checks: that the result may be an artifact of the particular post-period chosen, and that 2013 may not be the correct treatment date. Both concerns

prove unfounded—and the temporal analysis yields additional insight into the institutional dynamics.

Rolling windows. Restricting the sample to seven-year windows centered on successive periods (Table 17) reveals stable estimates across windows spanning the anti-corruption era. The consistency across overlapping windows confirms that the result is not driven by a single anomalous year. Note that windows beginning in 2013 or later cannot identify the NE × Post2013 treatment because all observations fall in the post-period, eliminating within-city variation in the treatment variable; these windows are accordingly excluded.

Table 17: Rolling Window Estimates (7-Year Windows)

Window	Coefficient	SE	p-value
2007-2013	-0.406	(1.240)	0.743
2008-2014	2.816*	(1.552)	0.070
2009-2015	3.932***	(1.497)	0.009
2010-2016	5.333***	(1.689)	0.002
2011-2017	4.193**	(1.713)	0.014
2012-2018	3.812**	(1.881)	0.043

*** p<0.01, ** p<0.05, * p<0.10

Alternative treatment dates. The horse-race specification from Table 7, which simultaneously includes Post2013 and Post2017 interactions, confirms the baseline timing choice. NE × Post2013 yields +5.60 ($p < 0.001$); NE × Post2017 is +1.59 (insignificant). The 2013 break dominates statistically. Out-of-sample validation extending the horse race to include Post2021 produces a significant reversal (-6.11, $p = 0.003$), confirming that the centralized accounting reform eliminated the manipulation margin (Online Appendix Table B.1).

Placebo. Assigning a false treatment date of 2011 (Table 18) yields a coefficient of 1.01 (SE = 2.77, $p = 0.72$)—statistically and economically zero. The contrast between the placebo (1.01) and the actual treatment (6.30) rules out the possibility that pre-existing differential trends between the Northeast and other regions drive the result.

8.4 Geographic Robustness

Regional treatment effects are vulnerable to the concern that a single province may drive the entire result. I address this through two complementary exercises: systematic leave-one-province-out estimation and aggregation to the province level.

Table 18: Placebo Tests

Placebo Test	Coefficient	SE	p-value
Fake Treatment: 2011	1.012	(2.767)	0.715
Fake Treatment: 2012	0.188	(2.018)	0.926
Randomization Inference	6.295	—	0.000
Randomization p-value based on 500 permutations			

*** p<0.01, ** p<0.05, * p<0.10

Leave-one-province-out. Re-estimating the baseline across all provinces, each time dropping one province (Table 19), confirms that no single province drives the result. Excluding non-Northeast provinces generates minimal variation. The informative exclusions are the three treated provinces themselves. Dropping Jilin reduces the coefficient to 3.94 ($p = 0.019$)—the largest attenuation—indicating that Jilin contributed the most to the pooled estimate, consistent with its intense supply-side reform exposure. Dropping Liaoning (7.77, $p = 0.004$) or Heilongjiang (7.84, $p < 0.001$) *increases* the coefficient, revealing that these provinces' gaps were smaller than Jilin's. That no single province is necessary for a positive and significant coefficient confirms the effect is distributed across the region.

Table 19: Leave-One-Province-Out Sensitivity Analysis

Excluded Province	Coefficient	SE	p-value
Province 13	6.417***	(1.771)	0.000
Province 14	6.439***	(1.771)	0.000
Province 15	6.528***	(1.768)	0.000
Province 21	7.770***	(2.717)	0.004
Province 22	3.937**	(1.679)	0.019
	...		
Province 61	6.430***	(1.771)	0.000
Province 62	6.364***	(1.771)	0.000
Province 63	6.332***	(1.768)	0.000
Province 64	6.339***	(1.768)	0.000
Province 65	6.348***	(1.769)	0.000
Range	[3.937, 7.842]	—	—
Full Sample	6.295***	(1.768)	0.000

*** p<0.01, ** p<0.05, * p<0.10

Province-level aggregation. Collapsing city-level data to province-year observations ($N = 278$) with province and year fixed effects (Table 20) yields a coefficient of 6.62 (SE = 3.16, $p = 0.048$), close to the city-level estimate (6.30), indicating that the result is a

genuine province-level phenomenon—systematic Northeast over-reporting—not an artifact of within-province variation or city-level measurement error.

Table 20: Province-Level Aggregation

	Secondary	Primary	Tertiary
NE × Post2013	6.623**	4.742**	3.038
	(3.159)	(1.891)	(1.861)
Observations	278	277	278

Province and year FE. HC1 robust SEs.

8.5 Satellite Validation

Satellite validation. As an independent cross-check, I benchmark bureau-reported secondary-sector growth against satellite-derived proxies for industrial activity. I construct a composite proxy by standardizing annual growth rates of $PM_{2.5}$ concentrations (Wei et al., 2021) and harmonized nighttime lights (Li et al., 2020) to z-scores and taking the equal-weighted mean. The satellite sample is smaller than the main panel (173 cities with 25 Northeast, versus 280 with 33), reflecting incomplete satellite coverage. In a difference-in-differences with city and year fixed effects, the bureau–composite gap coefficient is -0.613 (SE $= 0.205$, $p = 0.003$): after 2013, Northeast bureau statistics moved 0.6 standard deviations *toward* the satellite signal, confirming that bureau reporting genuinely improved. The absolute bureau gap, however, widened ($+0.392$, SE $= 0.122$, $p = 0.001$), indicating that bureau figures shifted from over-reporting to under-reporting rather than converging to accuracy. A three-proxy variant adding NO_2 produces a consistent estimate (-0.450, $p = 0.022$), as do PCA-based composites. The coexistence of bureau convergence toward satellite truth with persistent communiqué inflation—the same-sample Type 2 gap is $+4.203$ ($p = 0.028$)—provides external confirmation of the channel substitution mechanism. Full details appear in Online Appendix Section E.

8.6 Out-of-Sample Validation: Full 2010–2024 Results

The model, estimated entirely on 2010–2019 data, generates a sharp out-of-sample prediction: the 2021 centralized accounting reform, which transferred GDP calculation from local to central statistical agencies, should eliminate the manipulation margin entirely ($\pi_0 \to \infty$ in Theorem 1). I test this prediction by extending the sample to 2024 and estimating a three-treatment horse race that includes NE × Post_{2013}, NE × Post_{2017}, and NE × Post_{2021}.

Table 21: Out-of-Sample Validation: Full 2010–2024 Horse Race

	Dependent Variable: Gap (pp)		
	Secondary	Primary	Tertiary
	(1)	(2)	(3)
NE × Post-2013	5.634***	3.334**	0.864
(Anti-corruption)	(1.618)	(1.371)	(1.142)
NE × Post-2017	0.695	-0.416	1.729
(Liaoning revelation)	(2.811)	(1.097)	(1.243)
NE × Post-2021	-6.109***	0.118	-4.506***
(Centralized accounting)	(2.055)	(1.087)	(0.922)
City FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	4090	4089	4203

Notes: Out-of-sample validation using the full 2010–2024 panel. The main analysis uses the 2010–2019 sample, within which the dual-channel statistical system operated as described. The 2021 centralized accounting reform, occurring after the analysis period, provides a natural out-of-sample test: the model predicts that centralized accounting eliminates the manipulation margin ($\pi_0 \to \infty$, Theorem 1). The negative and significant Post-2021 coefficient confirms this prediction. Standard errors clustered at the city level. *** p<0.01, ** p<0.05, * p<0.10.

Table 21 reports the results. The NE $\times$ Post_{2021} coefficient is -6.11 percentage points ($p = 0.003$), confirming that the centralized accounting reform eliminated the manipulation margin. The magnitude of the reversal is comparable to the initial effect, indicating near-complete correction. The out-of-sample prediction is particularly striking: the theoretical mechanism, calibrated on pre-2019 data alone, correctly predicted both the direction and approximate magnitude of the post-reform change. This constitutes powerful evidence that the pre-2019 pattern reflected genuine strategic manipulation rather than measurement artifacts, confounders, or secular trends.

The event study extension (Online Appendix Figure B.2) reinforces this finding. Post-2021 coefficients collapse to near zero, with no residual differential between Northeast and non-Northeast prefectures after centralized accounting eliminated local discretion. The contrast between the escalating pre-reform trajectory and the abrupt post-reform reversal traces out the institutional mechanism with unusual clarity.

8.7 Summary

Four features of Table 22 deserve emphasis. First, the coefficient is positive in every specification that tests the 2013 treatment. Second, three of four appropriate inference methods reject the null at the 5% level—city-clustered ($p < 0.001$), province-clustered ($p = 0.016$), and two-way clustered ($p = 0.029$)—while Fisher randomization inference rejects at the 10% level ($p = 0.063$). Only the wild cluster bootstrap—which is known to be uninformative with exactly three treated clusters (Cameron et al., 2008)—fails to reject. Third, the placebo coefficient of 1.01 ($p = 0.72$) provides clean falsification: the effect is absent before the anti-corruption campaign. Fourth, the out-of-sample validation is striking: the NE $\times$ Post_{2021} coefficient of -6.11 ($p = 0.003$), estimated on 2010–2024 data that the model never saw, confirms the theoretical prediction that eliminating local discretion reverses the manipulation pattern.

9 Conclusion

The central finding of this paper is that statistical manipulation in authoritarian regimes is not a failure of monitoring but an optimal response to dual audiences. Officials who face both a verifying center and an interpreting public do not simply inflate or deflate; they engage in *performative manipulation*—choosing a reporting strategy in which magnitude, sectoral composition, and signal quality are jointly calibrated to balance mobilization gains among supporters against detection costs among moderates.

Table 22: Robustness Summary: NE × Post2013 Coefficient Across Specifications

	Specification	$\hat{\beta}$	SE/p	Significance
Panel A: Inference				
(1)	Baseline (city cluster)	6.295	(1.768)	***
(2)	Province cluster	6.295	(2.620)	**
(3)	Two-way cluster (city + year)	6.295	(2.886)	**
(4)	Wild bootstrap (Webb 6-pt)	6.295	[$p = 0.508$]	
(5)	Fisher randomization inference	6.295	[$p = 0.063$]	*
Panel B: Sample				
(6)	Balanced panel (all 10 yrs)	6.352	(1.770)	***
(7)	Excl. 2019 (Census year)	3.929	(1.530)	**
(8)	Province aggregation	6.623	(3.159)	**
Panel C: Outliers				
(9)	Winsorized 1%/99%	5.151	(1.489)	***
Panel D: Temporal				
(10)	Rolling window 2010–2016	5.333	(1.689)	***
(11)	Rolling window 2011–2017	4.193	(1.713)	**
(12)	Horse race (2013 + 2017)	5.604	(1.652)	***
Panel E: Geography				
(13)	Excl. Liaoning	7.770	(2.717)	***
(14)	Excl. Jilin	3.937	(1.679)	**
(15)	Excl. Heilongjiang	7.842	(2.058)	***
Panel F: Falsification				
(16)	Placebo: 2011 treatment	1.012	(2.767)	
Panel G: Out-of-Sample Validation (2010–2024)				
(17)	NE × Post2021	−6.109	(2.055)	***

Notes: Each row reports the coefficient on NE × Post2013 (or the indicated specification variant) from a two-way fixed effects regression with the secondary-sector communiqué-bureau gap as the dependent variable. Sample is 2010–2019 ($N = 2{,}729$) unless otherwise noted. Standard errors in parentheses; p-values in brackets for resampling/permutation methods. Baseline uses city-level clustering. Wild bootstrap uses 999 replications with Webb (2023) six-point weights at the province level. Fisher RI enumerates all $\binom{28}{3} = 3{,}276$ province-treatment permutations. Row 17 extends to 2010–2024 ($N = 4{,}090$) as out-of-sample validation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Full regression output in Appendix Tables A3, A7–A13, A16.

This reframing overturns the policy intuition that has guided statistical governance reform for decades: that better monitoring disciplines agents toward truthful reporting. The evidence from China shows that monitoring can backfire—not by increasing total manipulation, but by redirecting it to unmonitored channels where it is harder to detect and correct. The 2013 anti-corruption campaign raised the cost of statistical fraud nationwide, yet Northeast officials responded by *increasing* communiqué inflation by 6.30 percentage points in the secondary sector ($p < 0.001$, robust across all informative inference methods). The campaign tightened scrutiny over bureau-level statistics but left the communiqué channel largely unverified. Officials substituted across channels: they became more cautious with the data they submitted upward and more aggressive with the data they announced locally. Satellite-derived proxies confirm the substitution directly—bureau statistics converged toward satellite truth while communiqué inflation persisted (Section 8.5). The only intervention that eliminated manipulation was the 2021 centralized accounting reform, which removed local discretion over GDP calculation entirely. As an out-of-sample test, extending the analysis to 2024 reveals that the reform reduced the communiqué-bureau gap by 6.11 percentage points ($p = 0.003$), reversing the pre-reform pattern. The reform succeeded not by raising the cost of falsification but by removing the margin on which falsification occurred. A practical corollary: since the majority of Chinese cities under-report in communiqués, researchers who treat these figures as upward-biased introduce systematic error in the opposite direction. Bureau statistics should serve as the default baseline for empirical work on Chinese economic performance.

Several limitations merit acknowledgment. I observe the communiqué–bureau gap but not the primitives that generate it; audience composition, mobilization stakes, and target pressure are proxied rather than measured directly. Structural estimation exploiting the model's closed-form thresholds to back out latent parameters from observed gap distributions would sharpen the quantitative predictions. Extending the analysis to track individual cities as they transition between regimes—in response to leadership turnover, fiscal shocks, or institutional reforms—would test the comparative statics dynamically rather than cross-sectionally.[12]

The mechanism is not specific to autocracies: the tightening of Sarbanes-Oxley compliance in the United States redirected corporate earnings manipulation from GAAP to non-GAAP reporting channels (Dichev et al., 2013), a structural parallel to the communiqué–bureau substitution documented here. Any setting where accountability operates through parallel channels with heterogeneous verification is susceptible to the same logic. Understanding when officials inflate, when they deflate, and when they invest in making their distortions

[12] A multi-period extension with Bayesian learning by the center would endogenize the monitoring technology currently treated as exogenous and could rationalize the escalation pattern visible in the event study.

harder to detect is the first step toward designing statistical institutions in which the optimal level of manipulation is zero—not because officials are honest, but because the architecture of reporting leaves no margin on which dishonesty is profitable.

References

Becker, Gary S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, *76* (2), 169–217.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

_ , _ , and _ , "Robust Inference with Multiway Clustering," *Journal of Business and Economic Statistics*, 2011, *29* (2), 238–249.

Cao, Guangyu, Xi Weng, Mingwei Xu, and Li-An Zhou, "The political economy of ratchet effect: Evidence from China's environmental regulation," *Journal of Environmental Economics and Management*, 2025, *131*, 103150.

Chen, Shuo, Xinyu Qiao, and Zhitao Zhu, "Chasing or Cheating? Theory and Evidence on China's GDP Manipulation," *Journal of Economic Behavior & Organization*, 2021, *189*, 657–671.

Chen, Ting and James Kai sing Kung, "Busting the Princelings: The Campaign Against Corruption in China's Primary Land Market," *Quarterly Journal of Economics*, 2019, *134* (1), 185–226.

Chen, Wei, Xilu Chen, Chang-Tai Hsieh, and Zheng Song, "A Forensic Examination of China's National Accounts," *Brookings Papers on Economic Activity*, 2019, *2019* (1), 77–141.

Chen, Ye, Hongbin Li, and Li-An Zhou, "Relative Performance Evaluation and the Turnover of Provincial Leaders in China," *Economics Letters*, 2005, *88* (3), 421–425.

Crawford, Vincent P. and Joel Sobel, "Strategic Information Transmission," *Econometrica*, 1982, *50* (6), 1431–1451.

de Mesquita, Bruce Bueno, Alastair Smith, Randolph M. Siverson, and James D. Morrow, *The Logic of Political Survival*, Cambridge, MA: MIT Press, 2003.

Dichev, Ilia D., John R. Graham, Campbell R. Harvey, and Shivaram Rajgopal, "Earnings Quality: Evidence from the Field," *Journal of Accounting and Economics*, 2013, *56* (2–3), 1–33.

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin, "Why Resource-Poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data," *American Political Science Review*, 2009, *103* (4), 645–668.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Guriev, Sergei and Daniel Treisman, "Informational Autocrats," *Journal of Economic Perspectives*, 2019, *33* (4), 100–127.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 2012, *102* (2), 994–1028.

Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland, "Democracy and Transparency," *Journal of Politics*, 2011, *73* (4), 1191–1205.

Holz, Carsten A., "The Quality of China's GDP Statistics," *China Economic Review*, 2014, *30*, 309–338.

Jacob, Brian A. and Steven D. Levitt, "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating," *Quarterly Journal of Economics*, 2003, *118* (3), 843–877.

King, Gary, Jennifer Pan, and Margaret E. Roberts, "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review*, 2013, *107* (2), 326–343.

Li, Hongbin and Li-An Zhou, "Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China," *Journal of Public Economics*, 2005, *89* (9–10), 1743–1762.

Li, Xuecao, Yuyu Zhou, Min Zhao, and Xia Zhao, "A Harmonized Global Nighttime Light Dataset 1992–2018," *Scientific Data*, 2020, *7*, 168.

Lorentzen, Peter, "China's Strategic Censorship," *American Journal of Political Science*, 2014, *58* (2), 402–414.

Lü, Xiaobo and Pierre F. Landry, "Show Me the Money: Interjurisdiction Political Competition and Fiscal Extraction in China," *American Political Science Review*, 2014, *108* (3), 706–722.

MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb, "Wild Bootstrap and Asymptotic Inference with Multiway Clustering," *Journal of Business and Economic Statistics*, 2021, *39* (2), 505–519.

_ , _ , and _ , "Cluster-Robust Inference: A Guide to Empirical Practice," *Journal of Econometrics*, 2023, *232* (2), 272–299.

Martinez, Luis R., "How Much Should We Trust the Dictator's GDP Growth Estimates?," *Journal of Political Economy*, 2022, *130* (10), 2731–2769.

Rambachan, Ashesh and Jonathan Roth, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

Rawski, Thomas G., "What is Happening to China's GDP Statistics?," *China Economic Review*, 2001, *12* (4), 347–354.

Wallace, Jeremy L., "Juking the Stats? Authoritarian Information Problems in China," *British Journal of Political Science*, 2016, *46* (1), 11–29.

Webb, Matthew D., "Reworking Wild Bootstrap Based Inference for Clustered Errors," *Canadian Journal of Economics*, 2023, *56* (3), 839–858.

Wedeen, Lisa, *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*, Chicago: University of Chicago Press, 1999.

Wei, Jing, Zhanqing Li, Alexei Lyapustin, Lin Sun, Yiran Peng, Wenhao Xue, Tianning Su, and Maureen Cribb, "Reconstructing 1-km-Resolution High-Quality $PM_{2.5}$ Data Records from 2000 to 2018 in China: Spatiotemporal Variations and Policy Implications," *Remote Sensing of Environment*, 2021, *252*, 112136.

Xiong, Wei, "The Mandarin Model of Growth," *Journal of Economic Perspectives*, 2022, *36* (1), 143–166.

Xu, Chenggang, "The Fundamental Institutions of China's Reforms and Development," *Journal of Economic Literature*, 2011, *49* (4), 1076–1151.

Xu, Gang, Lixin Colin Xu, Ruichao Si, and Ruiting Wang, "Career Incentives, Tournament Competition, and Performance Manipulation: Evidence from Chinese Cities," *Journal of Law and Economics*, 2025, *68* (2), 459–498.

Young, Alwyn, "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results," *Quarterly Journal of Economics*, 2019, *134* (2), 557–598.

Zeng, Jiangnan and Qiyao Zhou, "Mayors' Promotion Incentives and Subnational-Level GDP Manipulation," *Journal of Urban Economics*, 2024, *143*, 103674.

Zhang, Jun, Xinrong Ma, and Zhikuo Liu, "Scale Estimation and Structural Analysis of China's Fiscally Supported Personnel," *China Economic Quarterly (*经济学季刊*)*, 2024, *25* (4). In Chinese. Original title: 中国财政供养人员规模估算与结构分析.

Appendix

(For Online Appendix with full proofs, additional tables, figures, and satellite validation, see separate document.)

A Key Theory Proofs

This appendix provides the proofs of Theorems 1 and 2, the central results of the paper. Complete proofs with all supporting lemmas, the microfoundation derivation, and comparative statics are in the Online Appendix.

A.1 Preliminary Lemmas

Lemma 4 (Quality Requires Misreporting)**.** *If $m^* = 0$, then $s^* = 0$.*

Proof. From the first-order condition $s = m \cdot \Delta_2(s)$, setting $m = 0$ gives $s = 0$. □

Lemma 5 (Monotonicity of Δ_2)**.** *If $\mu > \zeta$ (Assumption 1), then:*

(a) *$\Delta_2(s)$ is strictly decreasing in s for all $s \geq 0$.*

(b) *For $\alpha < \alpha_2$: $\Delta_2(s) > 0$ for all $s \geq 0$.*

(c) *For $\alpha > \alpha_2$: $\Delta_2(s) < 0$ for all $s \geq 0$.*

Proof. $\Delta_2'(s) = -(1-\alpha)\pi_0\mu^2 e^{-\mu s} + \alpha\theta\zeta^2 e^{-\zeta s} < 0$ under Assumption 2. For (b): $\Delta_2(0) = (1-\alpha)\pi_0\mu - \alpha\theta\zeta > 0$ iff $\alpha < \alpha_2$; since Δ_2 is decreasing and $\Delta_2(s) \to 0^+$ as $s \to \infty$, $\Delta_2(s) > 0$ for all s. Part (c) is analogous. □

Lemma 6 (Threshold Ordering)**.** $\alpha^0 < \alpha_2$.

Proof. $\alpha_2 - \alpha^0 = \pi_0\theta(\mu - \zeta)/[(\pi_0\mu + \theta\zeta)(\theta + \pi_0)] > 0$ since $\mu > \zeta$. □

A.2 Proof of Theorem 1 (Unconstrained Optimum)

Part (a): Under-reporting ($\alpha < \alpha^0$). At $s = 0$: $m^*(0) = [\alpha\theta - (1-\alpha)\pi_0]/\gamma < 0$ since $\alpha < \alpha^0$. Suppose $s^* > 0$. Then $s^* = m^* \cdot \Delta_2(s^*) < 0$ (negative times positive by Lemma 5(b)), contradiction. Corner optimality: $\partial U/\partial s|_{s=0} = m^*(0) \cdot \Delta_2(0) < 0$.

Part (b): Truthful reporting ($\alpha = \alpha^0$). $m^*(0) = 0$; by Lemma 4, $s^* = 0$.

Part (c): Sophisticated over-reporting ($\alpha^0 < \alpha < \alpha_2$). Since $\alpha > \alpha^0$: $\Delta_1(0) > 0$, so $m^*(0) > 0$. Since $\alpha < \alpha_2$: $\Delta_2(s) > 0$ for all s. The substituted FOC $\Phi(s) \equiv s - [\Delta_1(s)/\gamma] \cdot \Delta_2(s) = 0$ has $\Phi(0) < 0$ and $\Phi(s) \to +\infty$, so by the intermediate value theorem a root $s^* > 0$ exists. Uniqueness follows from the negative definiteness of the Hessian (Assumption 2). For the comparative statics, implicit differentiation of the FOC system yields $dm^*/d\alpha > 0$ (more misreporting) and $ds^*/d\alpha < 0$ (less quality) as α increases within this regime.

Part (d): Crude over-reporting ($\alpha > \alpha_2$). $m^*(0) > 0$ as in (c). Since $\alpha > \alpha_2$: $\Delta_2(s) < 0$ for all s (Lemma 5(c)). With $m^* > 0$: $s^* = m^* \cdot \Delta_2(s^*) < 0$, so the constraint $s \geq 0$ binds. Corner check: $\partial U/\partial s|_{s=0} = m^*(0) \cdot \Delta_2(0) < 0$. □

A.3 Proof of Theorem 2 (Constrained Optimum)

Proof. Proof. The target cost $g > 0$ shifts the net benefit: $\Delta_1(s; g) = \Delta_1(s) - g$. All threshold calculations proceed as in Theorem 1 with π_0 replaced by $\pi_0 + g$ in the truthful-reporting condition. The key distinction between Cases A and C arises from the ordering of α_g^0 and α_2.

Case A ($g < g^*$, so $\alpha_g^0 < \alpha_2$). The sophisticated over-reporting regime survives between α_g^0 and α_2, because the gap $\alpha_2 - \alpha_g^0 > 0$ ensures that there exist α values where both $m^* > 0$ (from $\alpha > \alpha_g^0$) and $\Delta_2 > 0$ (from $\alpha < \alpha_2$), producing interior $s^* > 0$.

Case C ($g > g^*$, so $\alpha_2 < \alpha_g^0$). Between α_2 and α_g^0, we have $m^* < 0$ (from $\alpha < \alpha_g^0$ and g large) and $\Delta_2 < 0$ (from $\alpha > \alpha_2$). The product $m^* \cdot \Delta_2 > 0$, supporting interior $s^* > 0$. This is the strategic under-reporting regime: officials deflate but invest in quality because their deflation is "against type" and risks detection by moderates who would find extreme conservatism suspicious.

The floor constraint $m \geq \underline{m}$ binds when $\alpha < \alpha_L$, where the unconstrained optimal m^* falls below $\underline{m}$. □

B Additional Tables

Tables referenced in the main text that appear in the Online Appendix include: clustering robustness (Table A3), sectoral DiD (Table A4), placebo test (Table A7), balanced panel (Table A8), excluding 2019 (Table A9), province aggregation (Table A10), winsorized estimates (Table A11), rolling windows (Table A12), leave-one-province-out (Table A13), and out-of-sample validation (Table A16). Full satellite validation results appear in Online Appendix Section E.

C Microfoundation: Mobilization as Expressive Utility

The mobilization term $\alpha\theta m e^{-\zeta s}$ in Equation (1) can be derived from expressive utility primitives in which supporters derive consumption value from the *act* of endorsing a bold economic narrative, with the endorsement value declining in signal sophistication because sophisticated signals lack the dramatic clarity that mobilizes political support. The full derivation, including the functional form $h(s) = e^{-\zeta s}$ and its behavioral foundations, appears in the Online Appendix.