Manipulation by Design: Performance Incentives and Adverse Selection in Bureaucracy

Ning He* Jiawei Fu†

August 8, 2026

Abstract

Performance-based promotion is often seen as meritocratic. We show it can do the opposite. When principals value the effort that promotion incentives elicit over the quality of the bureaucrats they elevate, it becomes rational to tolerate bureaucrats' manipulation of performance metrics. Manipulation drives both weaker and stronger candidates to compete harder, raising aggregate effort while corrupting selection. We formalize this logic and test it in the Chinese bureaucracy, developing a method that recalibrates city-level GDP growth to separate actual performance from manipulation. We find that reported growth predicts promotion but recalibrated growth does not. Instead, promotion is positively associated with overreporting: the system rewarded officials not for capability but for cheating. These findings position manipulation in authoritarian bureaucracies as a feature of incentive design rather than a symptom of weak oversight, and suggest how autocracies can grow without compromising loyalty: by eliciting effort rather than identifying the ablest.

*Assistant Professor, School of Public and International Affairs, University of Georgia
†Assistant Professor, Department of Political Science, Duke University

1 Introduction

A defining principle of modern bureaucracy is the insulation of officials from their political principals: secure tenure and advancement by competence safeguard officials' impartiality and expertise (Weber 2019). High-powered, performance-contingent incentives, by contrast, are deliberately kept at the margins in such a system. Authoritarian regimes, where political survival depends on bending the administrative apparatus to the regime's will (Gehlbach and Simpser 2015; Hassan 2020; Wintrobe 1998), often depart from this Weberian model. Rather than insulating bureaucrats, autocrats often tie their careers tightly to measured performance, deploying high-powered incentives to mobilize the bureaucracy toward the regime's goals (Breton and Wintrobe 1986; Gregory 2009; Heinen 2022; Nove 1958).

The trouble with high-powered incentives is that they reward results, and results can be faked: bureaucrats are tempted to manipulate the numbers on which their careers depend (Eckhouse 2022; Harrison 2011; Kalgin 2016). Such manipulation not only diverts effort from productive work; it also corrupts the signals that promotion relies on. The standard explanation traces manipulation to the principal's inability to observe what agents actually do, a difficulty especially acute where accountability mechanisms are weak (Acemoglu et al. 2020; He 2026; Sandefur and Glassman 2015). Manipulation, on this view, is an implementation failure, one that better monitoring can resolve. The promise of such an account is appealing: with enough oversight, a bureaucracy can be mobilized to deliver results on command, yet remain meritocratic in whom it elevates – a system where the Weberian trade-off between responsiveness and merit no longer binds.

This promise, however, rests on a questionable assumption: that the principal wants manipulation gone. The actual root of manipulation, we argue, is often not that the principal cannot detect it, but that he chooses to tolerate it. That choice reflects his priority between two goals that existing accounts have shown to be in tension – motivating effort and selecting the most capable, where pursuing one tends to come at the expense of the other (Höchtl et al. 2015; Mattozzi and Merlo 2015). When a principal prioritizes the effort subordinates exert over the selection of the ablest,

tolerating manipulation is not an oversight failure but a tool for motivating effort – the agents, counterintuitively, compete harder when cheating remains an option. Rather than a flaw to be rooted out, manipulation can be built into the design of a bureaucracy that prioritizes effort, one that wears the appearance of meritocracy without selecting on capability. This logic is not specific to autocracies, or to economic output: it applies wherever a principal evaluates heterogeneous subordinates on a manipulable metric and values their effort over selecting the ablest.

We formalize this argument in a model where the principal's objective – how much he prioritizes effort versus selection – determines whether manipulation is suppressed or sustained. This departs from existing models, in which manipulation is purely a friction that works against the principal's goals (Acemoglu et al. 2020; Eckhouse 2022; Serrato, Wang and Zhang 2019). When the principal prioritizes meritocratic selection, he acts to ensure the accuracy of performance measures through audits and punishment for cheating; this deters manipulation but dampens the effort of less capable agents. When the principal instead prioritizes aggregate effort, he tolerates manipulation and accepts distorted metrics as the basis for promotion – an incentive design that sustains effort at the cost of adverse selection, favoring dishonest agents over more capable ones.

We examine the second scenario in the Chinese bureaucracy, a setting that matches the model's effort-prioritizing condition. Our case is the promotion of city party secretaries, the officials who govern China's prefecture-level cities and whose careers are tied to the economic growth their cities report. The provincial leaders who control their promotion are themselves rewarded for the aggregate output of the jurisdictions they oversee (Bo 1996; Jia, Kudamatsu and Seim 2015; Li and Zhou 2005). This gives them reason to weigh that output more heavily than the quality of the subordinates they elevate. For principals with these incentives, our model predicts tolerated manipulation. Manipulation of economic statistics is, in fact, widespread in the Chinese bureaucracy (Chen et al. 2019; Chen, Qiao and Zhu 2021; Wallace 2016). In such a promotion system, we argue, reported GDP should no longer signal capability: the promoted should be no more able than those passed over; their advancement is driven instead by the other component of the reported figure – manipulation itself.

Testing this prediction requires a measure of actual performance – the signal of capability that manipulation has masked. Building on recent research on GDP manipulation across regimes (Martinez 2022), we develop a new method to recover it. We apply this method to recalibrate the reported GDP growth of China's prefecture-level cities over 2000–2013, when high-powered incentives were at their peak. The recalibration builds on a clear pattern in the data: reported GDP growth fell in years of city party secretary turnover, even as other economic indicators kept growing at a stable rate. We argue, and provide tests to support, that turnover weakened incumbents' incentives to inflate the figures, leaving the growth reported for turnover years less contaminated by over-reporting. Using reported GDP growth and nighttime lights from turnover years, we train a nonparametric model on this less-contaminated data, then use the fitted model to predict GDP growth for all city-years. These predictions serve as our estimates of actual GDP growth. On average, they fall 0.44 percentage points below the official figures, implying that reported growth was overstated by about 4%.

Drawing on these recalibrated figures, we reassess the relationship between economic performance and the promotion of city party secretaries. Our analysis confirms the documented positive correlation between official GDP growth and promotion at this administrative level (e.g., Landry 2008; Yao and Zhang 2015), but the correlation disappears once we substitute the recalibrated figures for the official ones – a contrast that is not just driven by the data-generating process. What did predict promotion was over-reporting: the more officials inflated their figures, the more likely they were to be promoted. The tournament, on this evidence, did not merely fail to sort city party secretaries by their ability to deliver economic growth; it rewarded the dishonest.

This article advances the literature on bureaucratic incentives and selection. Performance incentives are often credited with improving effort as well as selection: by spurring effort, they are thought to identify and elevate the most capable (Lazear and Rosen 1981; Rosen 1986). Modeling the trade-off between effort and selection explicitly, we show this complementarity need not hold – incentives tuned to maximize effort can erode selection, so that a tournament ostensibly rewarding performance may fail to elevate the ablest even as reported performance and promotion remain

correlated. What produces this erosion, as we show, is not a principal who seeks ability but cannot verify and punish manipulation (Serrato, Wang and Zhang 2019); nor is the principal a "mediocrat" who deliberately elevates the less able (Mattozzi and Merlo 2015). In our account, the principal accepts the corrupted signal as the price of effort, while preserving the appearance of meritocracy.

Our findings also have implications for understanding political selection in authoritarian regimes. The influential framework of the loyalty–competence trade-off predicts that autocrats must accept weaker performance as the price of selecting on loyalty (Egorov and Sonin 2011; Reuter and Robertson 2012; Zakharov 2016). China is an apparent exception to this account: its sustained growth under one-party rule has been taken as evidence that autocrats can escape the trade-off and select on competence after all (Bell 2016). Our findings cut against this reading: what has been taken for meritocratic selection may instead be a system that motivates the mediocre and rewards over-reporting. The more consequential implication, however, is that our case points to a general gap the theory leaves open: how a regime might sustain performance without compromising on loyalty.

The claim that an autocrat who staffs the bureaucracy for loyalty must accept poor performance rests on an unstated assumption: that performance depends solely on selecting capable officials. Yet performance can be raised by motivating less capable officials, as our findings suggest, so that a regime can select on loyalty and still obtain results. The performance cost of selecting on loyalty predicted by the existing account may be partly offset by a system of motivation. This aligns with what work on authoritarian purges has shown: autocrats can select on loyalty while still motivating effort (He and Wu 2026; Montagnes and Wolton 2019). Our evidence comes from a single hierarchy and a single measure of performance, but it points to a possible revision to the standard account: authoritarian survival is a matter not only of guarding against capable but dangerous subordinates but also of motivating mediocre but loyal ones. This may help explain why some autocracies prove both durable and effective.

2 Selection and Incentives in Promotion Tournament

This section develops our argument in two steps. We first lay out the central argument that a principal who prioritizes effort may prefer to tolerate manipulation rather than stamp it out; we then formalize this logic in a tournament model whose predictions we take to the data.

2.1 Central Argument

In a performance-based promotion tournament, promotion is awarded based on the candidates' relative rankings on certain performance metrics (Lazear and Rosen 1981). A promotion tournament has two potential functions: incentive provision and selection of the most capable candidates. The prospect of advancing to a higher-ranking, better-compensated position incentivizes bureaucrats to exert effort to maximize their chance of promotion. Promotion tournaments may also help principals distinguish subordinates of different ability and select the more capable ones for upper-level positions. In that case, the promotion is considered to be meritocratic.

A central question about promotion tournaments concerns the relationship between incentive provision and meritocratic selection. Early work on this question assumes that a tournament can simultaneously maximize effort and select more capable candidates (Rosen 1986), a claim that more recent research contests by identifying a trade-off between incentives and selection (Höchtl et al. 2015; Mattozzi and Merlo 2015; Serrato, Wang and Zhang 2019). This trade-off arises because promotion tournaments provide differential incentives to candidates with heterogeneous ability. Candidates with lower ability are often disincentivized to exert effort when they anticipate a low chance of outperforming those candidates with higher ability (Brown 2011). Those differential incentives can amplify performance gaps between candidates, thereby facilitating selection of candidates with higher ability. This selection gain, however, entails an effort loss: the aggregate effort exerted by all candidates is lower than maximal because the tournament discourages effort from not only lower-ability candidates, but higher-ability candidates as well – the latter can win without exerting much effort when their rivals are both less able and discouraged.

Although a principal must weigh selection against effort provision when designing the contest, this does not necessarily rule out promotion tournaments even when effort is the priority. We argue that a promotion tournament can be structured to maximize aggregate effort. When the structure of the competition moderates the heterogeneity between candidates, incentive provision can be improved (Höchtl et al. 2015). One way to achieve that is to allow candidates to manipulate performance metrics. Principals prioritizing total effort can implement a tournament in which manipulated performance figures are accepted. The intuition is that cheating opportunities raise lower-ability candidates' promotion prospects and, when the ability gap is sufficiently large, their real effort as well. Aware that lower-ability candidates' odds of promotion can be increased by cheating, higher-ability candidates are encouraged to exert greater effort to preserve their competitive advantage. The aggregate effort, then, will increase. However, this improvement in effort comes at the cost of meritocratic selection, as cheating can distort performance rankings of the candidates and attenuate the correlation between capability and promotion.

2.2 A Formal Model

We formalize this argument with a tournament-style competition (Catalinac, Bueno de Mesquita and Smith 2020; Lazear and Rosen 1981), in which the winner takes the reward and the loser receives nothing. Promotion contests take precisely this form: only a subset of bureaucrats can advance to the limited positions above. For simplicity, we consider one politician – the principal – and two bureaucrats – the agents. The politician assigns the two bureaucrats the same task, such as tax collection, job creation, or crime reduction, and promotes the one who produces more output.

Each bureaucrat chooses an effort level e_i. Following standard practice, we model ability through the marginal cost of effort: a more able bureaucrat exerts effort at lower cost. The two bureaucrats differ in this cost c_i, with $c_1 < c_2$. For simplicity, we assume output equals effort, so that a bureaucrat's output is e_i. The politician cannot directly observe the bureaucrats' output. He bases the promotion decision on reported output, which the bureaucrats have an incentive to inflate. A reported figure is $r_i = (1 + \rho_i)e_i$, where $\rho_i \in [0, 1]$ is the extent to which bureaucrat i inflates

output.

Manipulation carries a risk of detection. We capture this detection hazard by $\rho_i^2 e_i$, which rises with the degree of manipulation ρ_i and the scale of output e_i.[1] The quadratic term ρ_i^2 means that small manipulations are unlikely to be caught, while aggressive ones are caught with sharply increasing risk. A bureaucrat who is caught is punished, which occurs after the promotion decision and independently of its outcome: a bureaucrat who manipulates bears the expected punishment whether or not she is promoted. We summarize the politician's *cheating intolerance* – combining the intensity of monitoring and the severity of punishment – by a single parameter k. The expected cost of manipulation is then $k\rho_i^2 e_i$.

The expected utility of bureaucrat i,

$$u_i = \Pr(r_i > r_j) - c_i e_i - k\rho_i^2 e_i, \tag{1}$$

is the expected reward from winning the tournament minus the costs of competing. A bureaucrat is promoted by reporting higher output than her rival, which occurs with probability $\Pr(r_i > r_j)$. The associated reward is normalized to one. Competing carries two costs: the cost of real effort, $c_i e_i$, and the expected cost of manipulation, $k\rho_i^2 e_i$. Rewriting in terms of reported output r_i, we have

$$\tilde{c}_i(k) = \frac{c_i + k\rho_i^2}{1 + \rho_i}, \tag{2}$$

where $\tilde{c}_i(k)$ is the effective marginal cost of reported output. Evaluated at each bureaucrat's optimal manipulation $\rho_i^*(k)$, this cost rises with cheating intolerance k. Because each bureaucrat bears these costs whether or not she wins, the tournament is an all-pay contest, and its equilibrium is in mixed strategies. All proofs appear in Appendix A1.

The timing is as follows. First, the politician chooses his cheating intolerance k. Next, observing k, the bureaucrats choose their effort e_i and manipulation ρ_i. Finally, the bureaucrat reporting higher output is promoted. We assume the bureaucrats observe each other's costs, as this simplifies exposition and isolates the core mechanism.

We begin by studying how each bureaucrat manipulates in equilibrium, given the politician's cheating intolerance k. The lower-ability bureaucrat relies more heavily on manipulation, because exerting effort is more costly for her; this heavier manipulation also leaves her at greater risk of detection. The higher-ability bureaucrat, by contrast, substitutes effort for manipulation: with her lower effort cost, competing through real output is cheaper than risking detection. Equilibrium manipulation thus lines up with ability:

Proposition 1. *Given any value of $k > 0$, the bureaucrat with lower ability manipulates the output figure by a larger degree $\rho_2^* \geq \rho_1^*$. Each ρ_i^* is non-increasing in k, and strictly decreasing for $k > c_i/3$.*

This asymmetry, proved in Appendix A1.1, underpins the results that follow. Manipulation compresses the gap in reported output between the two bureaucrats: when manipulation is available, the cost of producing a given reported figure falls by more for the lower-ability bureaucrat, leading her to manipulate more aggressively than her rival is willing to risk. The two therefore appear more alike on paper than they are in fact. That compression, as we show below, drives both the effort gain and part of the selection loss in a cheating-tolerant tournament.

We now turn to the politician's cheating intolerance. Appendix A1.2 characterizes each bureaucrat's expected effort in the unique mixed-strategy equilibrium. Cheating intolerance k shapes total effort through two opposing forces: a higher k raises the effective cost of reported output, so both bureaucrats aim at smaller figures, which lowers real effort; cheating intolerance also curbs inflation, so a larger share of the figure they report must be actually produced, which raises effort. Because the two pull in opposite directions, total effort varies non-monotonically with k.

How, then, does the politician set his cheating intolerance? He chooses a k that maximizes his objective. We consider two such objectives: the bureaucrats' total effort and the quality of selection. The latter is captured by a *promotion gap* – the advantage in promotion probability the contest confers on the higher-ability bureaucrat, which is shaped by both effort and manipulation. The analysis that follows characterizes the politician's optimal tolerance under each objective, and shows that the two are in tension.

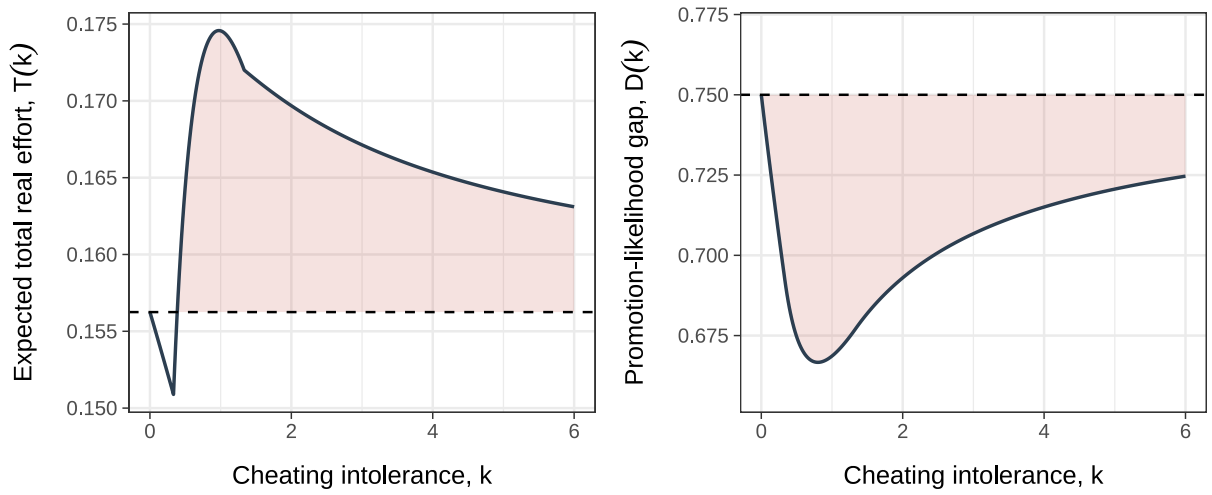

Figure 1: Cheating intolerance, bureaucratic effort, and selection. The left panel plots equilibrium total real effort $T(k)$ against cheating intolerance k; the right panel plots the higher-ability bureaucrat's promotion advantage $D(k)$. In each, the dashed line marks the no-manipulation benchmark (T^{NM} and D^{NM}, respectively). Both panels use the illustrative costs $c_1 = 1$ and $c_2 = 4$.

The left panel of Figure 1 plots equilibrium total effort $T(k)$ against cheating intolerance, with the horizontal dashed line marking the no-manipulation benchmark T^{NM} – the total effort a tournament would elicit if manipulation were impossible. Across almost the entire range of k, the solid curve lies above the dashed line: a tournament that tolerates some manipulation elicits more total effort than one that bans it. This effort gain from cheating tolerance is a general feature of the model, which holds as long as the two bureaucrats differ enough in ability:

Proposition 2 (Cheating intolerance and total effort)**.** *When the ability gap is sufficiently large ($c_2/c_1 > 1 + \sqrt{2}$), there exists a range of cheating intolerance over which equilibrium total effort exceeds the no-manipulation benchmark, $T(k) > T^{\mathrm{NM}}$.*

To see the intuition, consider how each bureaucrat responds to enforcement. When manipulation is banned outright, the lower-ability bureaucrat has little hope of winning – she cannot out-perform a rival who can match any effort at lower cost – and so exerts little effort. The higher-ability bureaucrat, facing a discouraged opponent, can prevail without working hard. Aggregate effort is therefore depressed. Tolerating some manipulation changes this: it lets the lower-ability bureaucrat stay in contention by inflating her figure, which forces the higher-ability bureaucrat to exert real

effort to keep her lead. This effort gain is not confined to the higher-ability bureaucrat. When the ability gap is large enough, the lower-ability bureaucrat also raises effort, as cheating restores her chance of winning.[2] Taken together, because limited manipulation can raise total effort above what a manipulation-free tournament would produce, the effort-maximizing politician would set an intermediate k, tolerating some manipulation rather than stamping it out. Appendix A1.3 shows that his optimal intolerance is finite and strictly positive, so that both bureaucrats manipulate in equilibrium.

Having characterized the effort-maximizing politician's choice, we now turn to the politician who prioritizes selecting the most capable. A meritocratic system maximizes the higher-ability bureaucrat's advantage in promotion probability – that is, promotion reflects the gap in competence between the candidates as closely as possible. Let $D(k) \equiv \Pr(1 \text{ wins} \mid k) - \Pr(2 \text{ wins} \mid k)$ denote this advantage in equilibrium, with D^{NM} its no-manipulation benchmark. Any tolerance of manipulation costs the higher-ability bureaucrat part of this advantage:

Proposition 3 (Cheating intolerance and the promotion gap)**.** *At every finite $k > 0$ the higher-ability bureaucrat's promotion advantage lies strictly below its no-manipulation benchmark, $D(k) < D^{\mathrm{NM}}$. The advantage is lowest at an intermediate intolerance: D falls on $(0, k^{\dagger})$ and rises on $(k^{\dagger}, \infty)$, where $k^{\dagger} = c_2/\left(1 + 2\sqrt{c_2/c_1}\right)$, approaching D^{NM} at both extremes.*

Proposition 3, proved in Appendix A1.4, establishes that a politician who prioritizes meritocracy will not tolerate manipulation.[3] With no manipulation, neither bureaucrat inflates, so promotion depends entirely on actual output. Because effort is more costly for the lower-ability bureaucrat, she produces less, and promotion strictly favors the higher-ability bureaucrat. Tolerating manipulation, as we establish in Proposition 1, shrinks the gap in reported output, so the higher-ability bureaucrat's advantage erodes. But this erosion is not monotonic in k. Tightening enforcement from a lenient baseline affects the higher-ability bureaucrat first: she has the least to gain from inflating, since she can compete on real output, and so is the first to abandon it. Her rival, who relies more heavily on manipulation to remain competitive, continues to inflate at the maximum. As enforcement intensifies enough to deter her rival as well, reported output converges back toward true output,

and selection recovers (right panel of Figure 1). Meritocratic promotion thus requires a thorough crackdown rather than a partial one; the latter disciplines the bureaucrat who was cheating least while leaving the heavier manipulator untouched – and selection deteriorates rather than improves.

Together, the two results define a trade-off. A politician who values effort has reason to tolerate some manipulation, since it keeps the contest competitive and, under certain conditions, raises effort from both bureaucrats; a politician who values selection has reason to suppress it, since manipulation lets the weaker candidate close the gap in reported output and corrupts the ranking. The same tolerance that raises effort degrades selection. Which the politician chooses thus depends on what he values. His choice determines whether the promotion system selects on competence or, despite its meritocratic appearance, on willingness to manipulate. In practice, the two are difficult to distinguish from outside: an effort-maximizing tournament sustains a correlation between reported performance and promotion even as the link between ability and promotion is tenuous – the observable pattern we take to the data.

3 The Case of China

Our theory does not predict which priority a given principal will hold; it identifies the consequences of each for effort and selection. That central trade-off between the two, however, is not something our data can test directly. To test it we would have to observe tournaments arrayed across the range of principals' priorities, from those who prioritize selection to those who prioritize effort. Perhaps no single bureaucracy offers such variation; in practice we observe only one point on the spectrum – a certain type of tournaments guided by a common priority. We therefore do not test the trade-off but characterize one branch of it – the effort-prioritizing one – with our case and show that it carries the selection feature our model predicts. For that we turn to the Chinese bureaucracy, where the officials who decide promotions are themselves evaluated and rewarded on the economic growth of the jurisdictions they oversee, and so have strong reason to prioritize aggregate output.

A vast literature on the Chinese bureaucracy relates local officials' career outcomes to their

success in meeting the state's policy priorities. Among these priorities, GDP growth has drawn the most scholarly attention, with numerous studies documenting a positive association between reported GDP growth and the promotion of local party and government officials (e.g., Bo 1996; Jia, Kudamatsu and Seim 2015; Landry, Lü and Duan 2018; Li and Zhou 2005; Yao and Zhang 2015). These findings have been taken as evidence of tournament-style competition in the Chinese bureaucracy, in which officials who deliver higher growth are more likely to rise. These findings resonate with two influential claims. First, the prospect of promotion gives local officials a powerful motive to pursue growth, an incentive widely cited as a driver of China's rapid development (Xu 2011). Second, by rewarding performance, the CCP is thought to elevate its most capable officials, running a bureaucracy governed by meritocracy (Bell 2016).

Alongside performance-based promotion runs widespread GDP fraud. The production of local economic statistics in China is decentralized: a prefectural city's figures are compiled by its own statistics bureau – an office that answers directly to the very mayor and party secretary whose performance those figures record. Controlling the bureau's budget and its director's career, these leaders effectively control what it reports. GDP growth rates typically cross their desks for approval before publication (Holz 2014, 322), and the numbers that emerge are, unsurprisingly, inflated. That local economic data are routinely overstated is an open secret among CCP officials. Li Keqiang, then party secretary of Liaoning and soon to be vice premier, privately told the U.S. ambassador in 2007 that his own province's GDP figures were "man-made" and "for reference only" (The Economist 2010).

The CCP has done little to curb this fraud. Although centralizing the production of statistics – so that the numbers no longer pass through the officials being judged – would curtail manipulation, no such reform came for decades. The National Bureau of Statistics proposed centralizing provincial GDP accounting in 2004 but did not implement it until 2019 (Xinhua 2017); below the provincial level, reporting is still controlled by officials at the same tier. Nor were lighter remedies used: higher authorities verified the figures only infrequently and punished manipulation even more rarely. Detected fraud drew revision rather than sanction: central inspectors called data falsification

"epidemic" in Liaoning in 2014, yet the province was left to quietly restate its 2011–2014 figures years later with no one punished for the fabrication itself (Caixin Global 2017). The regulatory framework for ensuring accurate reporting, as Holz (2014, 322) concludes, is extraordinarily weak. That weak enforcement holds not only between the center and the provinces but also – and more pervasively – between provinces and prefectures, where even anecdotal cases of enforcement are hard to find.

Why did provincial leaders leave fraud at lower levels unchecked? Their tolerance was deliberate, not a failure of enforcement capacity. A top priority for provincial leaders was raising GDP growth in their province – a component of their own performance review. What advances them is the aggregate output their subordinates produce in pursuit of provincial growth targets, not the discernment to identify more capable subordinates for higher office. Tolerating fraud, as our model predicts, serves exactly that priority. The tolerance is not just inferred: even official newspapers have acknowledged coordinated falsification running across the local hierarchy, from provinces down through prefectures to counties (China Daily 2018). But targets cannot be met through inflation alone; the leaders must combine inflation with real output. That calculus – maximizing aggregate effort, not just inflating figures – feeds into promotion decisions, producing a tournament in which advancement rests on manipulated growth figures.

Performance-based promotion and tolerated fraud thus coexist by design – and that combination calls into question the standard interpretation of the evidence. If promotion is awarded on reported figures that are inflated, then the well-documented correlation between GDP growth and advancement need not reflect meritocratic selection. Officials who reported faster growth may have risen not because they delivered it, but because they inflated it more aggressively. The empirical literature cannot distinguish these possibilities, since it measures performance using the very figures that manipulation contaminates. Whether the correlation reflects selection on competence or selection on manipulation is therefore an open question, one that requires separating real performance from the inflation layered on top.

These considerations translate the selection prediction of our model into testable claims. The

model's selection result, as Proposition 3 lays out, implies that when a principal tolerates manipulation, promotion weakly tracks ability and, in our setting, the real growth that ability produces. If provincial leaders prioritize aggregate output and tacitly accept GDP fraud, then the actual economic performance of city party secretaries should have less bearing on whether they are promoted:

> *Hypothesis 1:* City party secretaries' actual economic performance, measured by recalibrated GDP growth, is less strongly associated with promotion than is reported GDP growth.

If ability only weakly explains promotion, then what kind of official is favored, and what else might the tournament be rewarding? Our model assumes that the candidates compete in the same environment, and that the only meaningful difference across them is ability. In practice, though, city party secretaries differ in their willingness to cheat – whether from personal scruple, professional norms, or reluctance to earn a reputation for dishonesty – and this willingness is not entirely explained by ability. When a principal prioritizes effort, as in our case, that priority shows up not only in weak selection on competence but in how promotion tracks the willingness to manipulate.

High tolerance for manipulation reduces the reward gap between candidates of different *ability* but might widen it between candidates of different *honesty*. Among city party secretaries of equal ability, the one who feels less troubled by falsifying can simply inflate her figures, while a more scrupulous rival must produce the output she claims. When enforcement is uniformly low, a dishonest party secretary inflates more, and the effective cost of her reported figures, as equation (2) suggests, is correspondingly lower. She reports a higher number at a cost her more scrupulous rival cannot match and is therefore the more likely to rise. When fraud goes unchecked, promotion tracks willingness to falsify:

> *Hypothesis 2:* Among city party secretaries of comparable actual performance, those who over-reported GDP growth to a greater degree are more likely to be promoted.

4 Recalibrating GDP Growth Figures

To test the hypotheses above, we develop a method to recalibrate local GDP growth figures and estimate the manipulation embedded in those figures. We apply it to prefectural-level cities over 2000–2013, when performance incentives were most tightly tied to GDP growth. The method improves on the proxy-based approaches common in existing work, which estimate real growth by regressing the reported figures on the proxies of growth. Such approaches, we show, rest on an assumption that is unlikely to hold: that manipulation is uncorrelated with actual growth. When it fails, proxy methods cannot separate the two cleanly: their fitted values capture not only real growth but the manipulation correlated with it. Our approach instead identifies a subset of reported growth figures that are less susceptible to manipulation and uses them to model real growth. This section presents the method, contrasts it with these alternatives, and compares our recalibrated figures with the official series.

4.1 Leader Turnover and Manipulation

Our method exploits variation in local leaders' incentives to overreport GDP growth – variation tied to the turnover of city party secretaries. These incentives weaken in turnover years because of the timing of GDP accounting. A given calendar year's growth figure is not settled until early the following year, so when the party secretary changes, the figure is finalized and released under the incoming leader. That new leader has little reason to inflate it: the growth occurred largely under a predecessor, and with two leaders having held the office during the same year, it cannot readily be claimed as the new leader's own record. Consistent with this, reported growth in turnover years does not predict the incoming secretary's later promotion (Appendix Table A1). The figures reported in turnover years should therefore be less contaminated by manipulation – the premise on which our recalibration rests.

Just as this reasoning predicts, reported GDP growth falls in years of city party secretary turnover.[4] As Figure 2 shows, average growth in turnover years is lower than in the years immediately

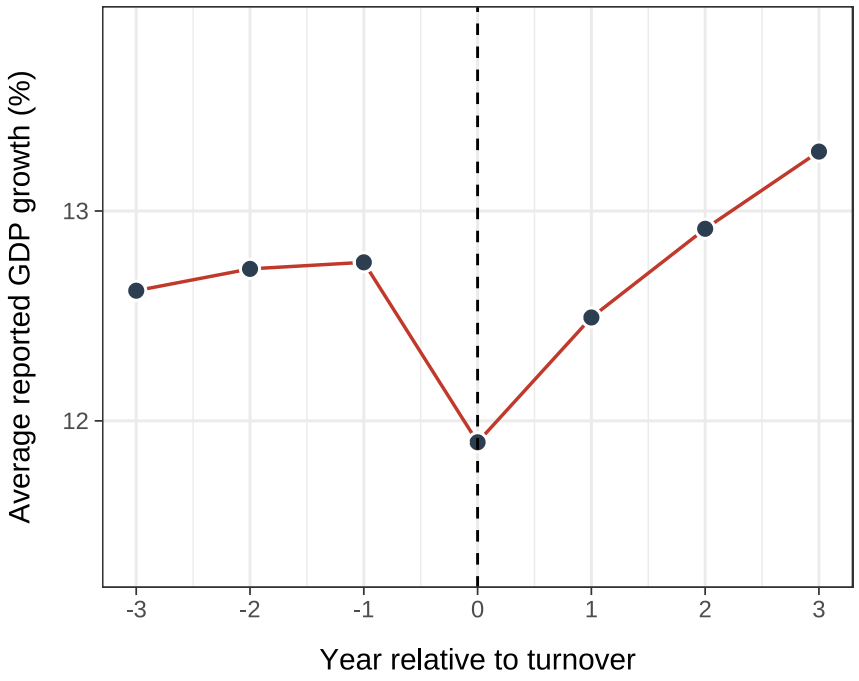

Figure 2: City party secretary turnover and reported GDP growth.

before and after, by about 0.5 percentage points relative to years without turnover (hereafter *incumbent years*). To confirm that this gap is statistically significant, we regress reported growth on turnover with city and year fixed effects, and estimate a decline of virtually the same magnitude (Appendix Table A2).[5]

The turnover gap is open to an alternative interpretation: that real economic activity slows when leaders change. The analyses that follow rule out this possibility and confirm the opposite: that the gap reflects reduced manipulation, not a real downturn. We first test whether real growth actually slowed in turnover years. Regressing a range of growth proxies – nighttime lights, value-added tax, bank lending, and others – on party secretary turnover, we find no significant difference between turnover and incumbent years for any of the proxies (Figure 3); reported GDP growth, by contrast, is the only indicator that shifts in turnover years. Real growth thus shows no turnover-year slowdown, leaving reduced manipulation as the more likely source of the gap.

A second test looks for signs of manipulation in the reported figures. Manipulated numbers often heap at round values, leaving artificial bunching in the distribution of reported growth (McCrary

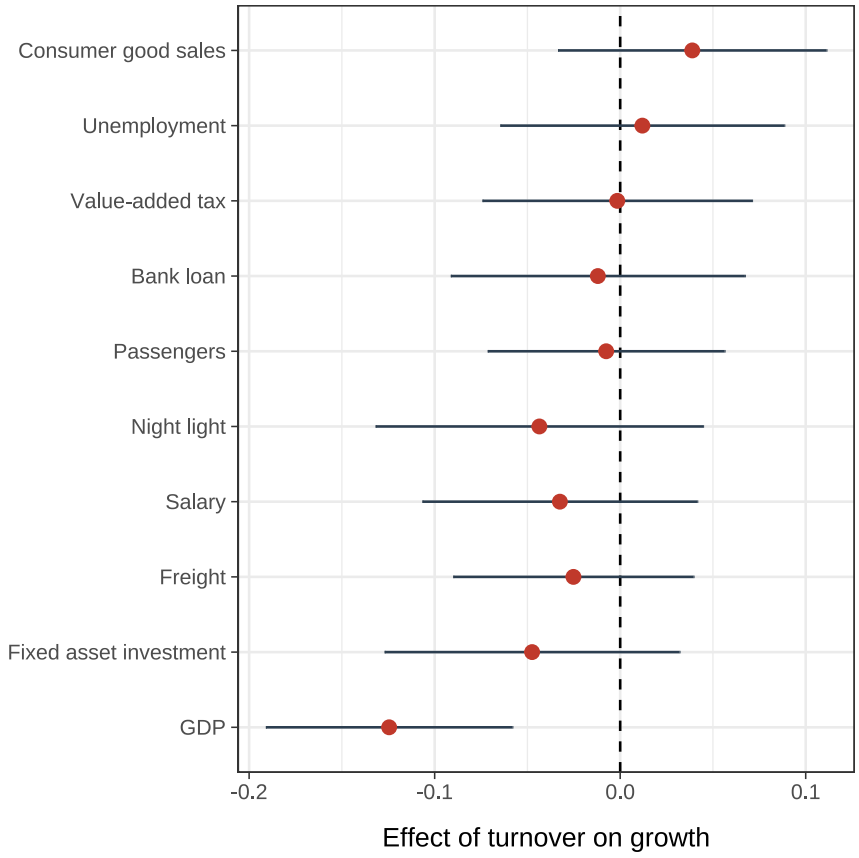

Figure 3: Estimated effects of city party secretary turnover on various economic indicators measured in growth rate. The dependent variables are standardized, so the coefficients reflect a one-standard-deviation change. Each regression conditions on city and year fixed effects. 95% confidence intervals based on standard errors clustered by province-year.

2008). If the turnover gap reflects reduced manipulation, bunching should be weaker in turnover years than in incumbent years. The incumbent-year density, for instance, falls just short of 12% growth and then spikes at it – the signature of figures rounded up to a focal number – while the turnover-year density shows a far less conspicuous break. Applying the discontinuity tests of Cattaneo, Jansson and Ma (2020), we confirm this pattern across integer values: pronounced bunching in incumbent years, markedly weaker bunching in turnover years (Appendix Table A3).

A third test compares reported growth against nighttime light growth directly. Holding light growth fixed, reported GDP growth runs systematically higher in incumbent years than in turnover years (Figure 4). Two cities with the same underlying growth – as captured by light growth – thus report differently depending on whether their leader is being replaced, a gap difficult to attribute to

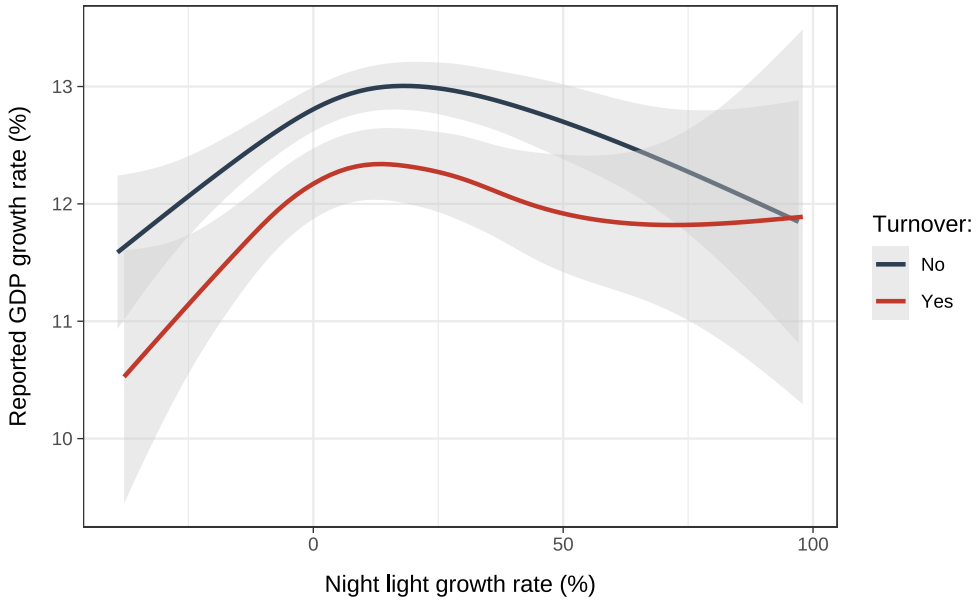

Figure 4: The relationship between nighttime light growth and reported GDP growth. Regression curves based on a generalized additive model with 95% confidence intervals.

anything but manipulation. Together, these tests converge on a most plausible interpretation: the turnover-year decline in reported growth reflects weaker manipulation, not a real slowdown.

4.2 Recalibration Method

Our recalibration strategy is to estimate a model of GDP growth on turnover-year observations, then use it to predict growth for every city-year in the sample. We model reported growth Y_t as a function of economic indicators X_t on turnover-year data:

$$Y_t = f(X_t) + \varepsilon_t \tag{3}$$

Our recalibration rests on two assumptions. The first, established above, is that GDP figures reported in turnover years carry less manipulation than those reported in incumbent years. We do not claim that turnover-year figures are entirely free of manipulation; we assume only that any residual manipulation is small and, crucially, independent of actual growth – so that it does not distort the estimated relationship $\hat{f}(X_t)$ but is instead absorbed into the error term ε_t.

The second assumption is that the relationship between actual growth and the predictors does not systematically differ between turnover and incumbent years. Evidence on the relationship between nighttime lights and reported GDP – both in growth-rate terms – supports this: the fitted curves for turnover and incumbent years run parallel, differing by a roughly constant vertical gap across a wide range of the data (Figure 4). The shared slope indicates that light growth maps onto GDP growth in the same way in turnover and incumbent years. Incumbent-year GDP growth thus follows the same relationship as in equation (3), augmented by a manipulation term:

$$Y_i = f(X_i) + M_i + \varepsilon_i \tag{4}$$

We estimate f from turnover-year data and apply it to all city-years, yielding predicted growth $\hat{Y} = \hat{f}(X)$. The discrepancy between reported and predicted growth, $Y - \hat{Y}$, recovers manipulation M, plus a random error. In turnover years, manipulation is negligible by assumption, so the discrepancy is essentially noise; in incumbent years, it captures, on top of the noise, the manipulation we seek to measure.

The dependent variable is official GDP growth rates from the CEIC database. The predictors are selected with manipulation in mind: because the method recovers manipulation as the gap between reported and predicted growth, the predictors themselves must be free of manipulation. We thus exclude indicators most exposed to manipulation, such as fixed-asset investment and employment – the scrutiny they draw makes them targets. Instead, we retain less scrutinized ones, like passenger or freight volumes. We do include value-added tax (VAT), despite its prominence, because it is costly to inflate: VAT revenue is shared with the central government, so overstating it would mean remitting more revenue upward (Chen et al. 2019, 101-102).

Alongside official statistics, we draw on four sources that lie outside the local statistical apparatus and are therefore beyond the reach of local manipulation: satellite nighttime lights, land transactions, customs records of foreign trade, and firm registration. For nighttime lights we use two measures. The first is the DMSP-OLS stable-lights series; the second is the global NPP-VIIRS-like product of Chen et al. (2020), a cross-sensor calibration that reconstructs VIIRS-scale radiances back

through the DMSP era and so provides consistent coverage of our full 2000–2013 window (rather than the raw VIIRS record, which begins only in 2012). For each we build a city-year indicator by aggregating light intensity across all pixels within the city's administrative boundary. Land transactions come from the Land Transaction Monitoring System, an official registry of several million records, from which we compute the total area and value of land sold in each city-year. Foreign trade comes from item-level Chinese Customs data, from which we compute total export value per city-year. Firm registration is constructed from systematic records of newly registered firms by counting the number in each city-year.

Beyond the predictors described above, which enter as annual growth rates,[6] the model also includes province and year fixed effects, so that $\hat{f}$ is identified from variation within provinces and years rather than from regional or temporal trends. The predictors, at this point, are only candidates. A practical constraint shapes which finally enter the model: many of the official statistical series are missing for a non-trivial share of observations, and adding predictors to improve fit shrinks both the training sample and the recalibrated sample, raising the risk of sample-selection bias. This bias is twofold once the recalibrated data are used to predict promotion. First, estimating $\hat{f}$ on a selected subsample of the training data biases the recalibration itself. Second, the recalibrated data have considerable missing values, so the promotion regressions run on a non-random set of city-years. We therefore seek a model that balances fit against sample coverage, tilting the balance slightly toward the latter.

We fit ten combinations of the candidate predictors using a random forest algorithm and compare them on fit and coverage (Appendix Table A4). Model fit does not improve as predictors are added: nighttime light alone, with province and year fixed effects (Column 1), returns an R^2 of 0.39, and adding the remaining proxies leaves fit flat. Several richer specifications fit slightly worse (Columns 3–8), and only the fullest models edge up to 0.40. Coverage, by contrast, falls sharply: the recalibrated sample shrinks from 4,508 city-years under the parsimonious specification to 2,553 under the model that stacks every proxy (Column 10). That nighttime light alone captures nearly all the explainable variation is unsurprising, as it is among the strongest individual predictors of GDP

growth. And because nighttime light is never missing, this specification recalibrates the full sample: $\hat{f}$ is estimated on all turnover-year observations, and the recalibrated data span every city-year. It thus preserves coverage at little cost in fit. We therefore adopt it as our baseline and use the richer specifications for robustness.

Applying this model generates two measures: recalibrated growth, our estimate of a city's actual economic performance, and the gap between reported and recalibrated growth, our measure of manipulation. Reported GDP growth exceeds recalibrated growth by an average of 0.44 percentage points – an over-statement of roughly four percent relative to recalibrated growth. The gap also varies widely across city-years (SD = 2.9 percentage points), and it is this variation that our promotion analysis exploits.

4.3 Validating the Recalibration

We validate the resulting measures in four steps. First, we show that recalibration strips out manipulation that the reported figures carry but existing approaches do not fully isolate. Second, the recalibrated series tracks real economic activity closely. Third, the gap responds to the very incentives that drive manipulation. Finally, we ask whether our conclusions survive downward manipulation of the turnover-year data.

4.3.1 Does the Recalibration Remove Manipulation?

To assess whether the recalibrated figures are free of manipulation, we compare them against cities' announced growth targets. At the start of each year, a city's government work report sets a GDP growth target for the year ahead – a threshold local officials have strong reason to clear, since missing a publicly announced target is a visible failure, while meeting it, by however slim a margin, averts one. Manipulation aimed at that threshold should leave a distinctive trace: an excess of city-years landing just above the target and a shortage just below. If the reported figures show this bunching and the recalibrated figures do not, the recalibration has removed exactly the distortion the target creates.

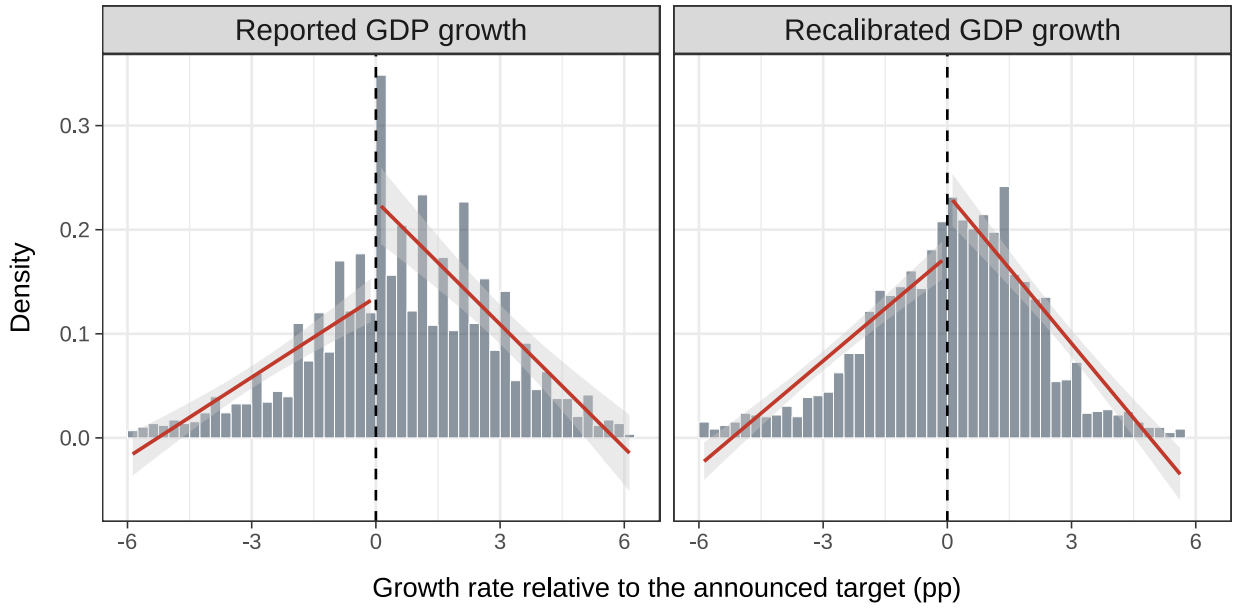

Figure 5: Bunching at the announced growth target.

The evidence supports this hypothesis: the bunching is present in the reported figures and absent from the recalibrated ones. Figure 5 plots the density of city-year growth rates relative to the announced target, which is normalized to zero. The reported series breaks at the threshold: observations pile up just above it and thin out just below, indicating that figures falling short of the target are pushed across it. The recalibrated series passes through the same threshold smoothly, with no discernible break. Density tests confirm the contrast: the discontinuity in the reported series is 0.21 and significant at the one percent level, while the recalibrated series yields 0.04, indistinguishable from zero ($p = 0.57$) (Appendix Table A5). The recalibration thus removes the manipulation that an externally defined incentive threshold induces.

This test also bears on how our method compares with existing approaches (e.g., Chen, Qiao and Zhu 2021; Wallace 2016). Like ours, the standard approach uses proxies for economic activity to estimate manipulation in reported growth; the difference lies in whether the GDP data used to fit the model are themselves manipulated. A model trained on manipulated data learns to reproduce part of that manipulation when the manipulation tracks actual growth, so it enters the predicted values and survives in the recalibrated figures (Appendix A2). The standard approach thus rests on an

assumption – independence between manipulation and actual growth – that our turnover-year design makes more plausible. By fitting the model on less contaminated data, we keep manipulation largely out of the estimated relationship and confine it to the residual, so that the discrepancy between reported and recalibrated growth provides a valid estimate of manipulation.

The test against growth target provides evidence for this difference. We construct the series the standard approach would produce – the same model with the same predictors, trained on all city-years rather than on turnover years – and apply the same density test (Appendix Table A5). If that approach leaves manipulation in its fitted values, those values should still cluster above the target. They do. The standard-method series carries a discontinuity of 0.11, roughly half the break in the reported figures and well above the 0.04 our recalibration leaves behind, though the estimate is significant only at the 10 percent level ($p = 0.09$). A model fitted to manipulated figures thus reproduces part of the target-clearing manipulation and passes it on to the growth estimate. Our recalibration, in contrast, removes the target-clearing manipulation more effectively by training on the cleaner turnover years.

4.3.2 Does Recalibrated Growth Capture Real Performance?

To test whether recalibrated growth reflects real economic activity, we assess it against indicators that are less susceptible to manipulation and play no part in the recalibration. From these – customs records of exports, land sold by area and by value, freight and passenger volumes, and new firm registrations – we build a composite index of real activity, and compare how closely reported and recalibrated growth follow it. Recalibrated growth tracks the index about as closely as the reported figures do: the two correlate with it almost identically, and the pattern holds across alternative fixed effects, alternative constructions of the index, and a rank-based correlation (Appendix Table A6). Removing manipulation from reported growth thus does not discard the real economic signal those figures carry. Instead, the two measures correspond to real activity about equally well.

This bears directly on the analysis that follows. Although reported and recalibrated growth track real activity equally well, as we show, only reported growth predicts promotion. This contrast

therefore cannot be attributed to recalibrated growth being the poorer measure of real performance. What separates the two is not their grip on real activity but the manipulation that reported growth contains.

4.3.3 Does the Gap Reflect Manipulation?

Our measure of manipulation – the gap between recalibrated and reported growth – tracks promotion incentives while remaining unrelated to traits that carry no such incentive. Regressing the gap on individual characteristics of city party secretaries (Appendix Table A7), we find that over-reporting is higher later in a secretary's term and for older secretaries – consistent with the intensifying push for promotion as a career advances – while characteristics with no clear bearing on promotion incentives, such as gender and ethnicity, show no relationship. This term-specific rise holds even with secretary fixed effects, which compare each secretary against their own turnover year. Over-reporting climbs over the following years and remains elevated through most of a term (Appendix Table A8).

Over-reporting also rises with the ambition of a city's growth target. Since a higher target is harder to meet through real growth, it is more likely to be met through manipulation. A target one percentage point above the city's recalibrated growth in the previous year is followed by 0.33 percentage points more over-reporting, against a sample mean of 0.66. When ambition is measured instead by the year-on-year change in the announced target, which involves no recalibrated figure at all, the estimate is 0.16 (Appendix Table A9). Both are precisely estimated while conditioning on city and year fixed effects, showing again that our measure of manipulation closely tracks performance incentives. Together, these two tests confirm that the gap behaves as a valid measure of manipulation should.

4.3.4 Sensitivity to Under-Reporting

Our recalibration assumes that reported GDP growth was not systematically inflated in turnover years. We have argued that city leaders had little incentive to inflate; we cannot, however, rule out

the opposite – that they deflated the figures, understating turnover-year growth to leave more room to claim credit later in their tenure. Such downward manipulation is costly as it would likely invite opposition from the incumbent mayor, whose own career rests on the same figures, thus making systematic deflation unlikely. Even so, we consider how it would affect our recalibration were it systematic.

Whether such deflation matters depends on its structure. If under-reporting were close to uniform across cities and years, it would lower the recalibrated figures on average but leave their relative differences across city-years unchanged. Because our promotion analysis exploits these relative differences rather than the absolute level of growth, such uniform under-reporting would not affect our conclusions. The under-reporting might add noise and a shift in the average to recalibrated growth, but it does not bias the comparisons we draw.

Under-reporting introduces distortion to the comparison only if it is correlated with actual growth. This possibility is hard to test because we cannot directly measure under-reporting, but we can probe it with the same nighttime-light evidence presented above (Figure 4). Suppose under-reporting in turnover years were systematically correlated with actual growth. It would then bend the relationship between nighttime light and reported GDP – steepening or flattening it, depending on the direction of the correlation – so that the turnover-year curve would no longer run parallel to the incumbent-year curve. What we observe is the opposite: the two curves share the same shape, separated only by a constant gap – not the divergence that growth-correlated under-reporting would produce.

5 Performance, Manipulation, and Promotion

We now use the recalibrated figures to test our two hypotheses. If promotion was tied to actual performance, recalibrated growth should predict which city party secretaries were promoted (*Hypothesis 1*). If instead the ostensibly meritocratic pattern rested on manipulation, over-reporting, not actual growth, should predict promotion (*Hypothesis 2*). We find the latter: recalibrated growth

bears no relationship to promotion, while over-reporting predicts it strongly.

5.1 Empirical Strategy

We test the hypotheses using data on the party secretaries of all prefectural- and sub-provincial-level cities between 2000 and 2013.[7] Our outcome is a binary indicator for whether a secretary was promoted during their term – that is, moved to a position of higher administrative rank.[8] Of the 1,392 leader terms in the data, 611 ended in promotion, a rate of 44%. Because a term spans several years, the chance of promotion in any given year is much lower: promotion occurs in about 12% of secretary-years, which is the dependent-variable mean in our regressions. Promotion probability also showed temporal variation over a secretary's tenure: lowest in the first year, rising steadily thereafter.

To estimate how a city's measured economic performance relates to its party secretary's promotion, we specify the following model, which we estimate with both reported and recalibrated growth. We use leader-year observations, which capture variation in performance over a secretary's term in office. The specification is:

$$promotion_{it} = \beta \times \frac{1}{t-1} \sum_{s=1}^{t-1} \phi_{is} + X_i'\gamma + \theta_{py} + \varepsilon_{it} \tag{5}$$

where the outcome is a binary indicator equal to one if city party secretary i was promoted in year t of their tenure ($t \geq 2$, since performance is undefined in the first year). It depends on the secretary's average economic performance over their prior years in office – from year 1 through year $t-1$, the signal available when the promotion decision was made – along with a vector of individual characteristics X_i – the secretary's age, gender, ethnicity, education, and administrative rank – with coefficients γ, and province-by-year fixed effects θ_{py}, where p and y denote the province and calendar year of observation (i, t). We use province-by-year fixed effects because the promotion tournament is a competition across cities within a province; conditioning instead on city fixed effects would assume secretaries compete against those who held the same post – their predecessors

and successors – which does not match how the tournament operates.

We measure a secretary's economic performance as their city's growth relative to the provincial average, averaged over their prior years in office. We construct it in two steps. First, for secretary i in the t-th year of their tenure, we compute ϕ_{it}, the deviation of their city's reported or recalibrated GDP growth from the provincial average that year – a measure of how the secretary performed relative to others competing for promotion. Second, we average ϕ over the secretary's observed prior years in office, $\frac{1}{t-1}\sum_{s=1}^{t-1}\phi_{is}$.[9] For a promotion decision in the third year, for instance, performance is the average of ϕ over the first two years.

5.2 GDP Growth and Promotion

Before turning to the regressions, we describe the relationship between measured performance and promotion. Within each province-year, we rank secretaries by their reported GDP growth, and separately by their recalibrated growth, then examine how each ranking relates to the probability of promotion the following year. Figure 6 plots the average probability of promotion against performance ranking for both measures. Secretaries ranked higher on reported GDP growth are more likely to be promoted. The pattern is much less visible, however, for recalibrated growth: promotion probability is roughly flat across the recalibrated ranking.

Table 1 reports the estimated relationship between measured performance and promotion. Columns 1 and 2 express performance in its natural units – percentage points of growth relative to the provincial average – while Columns 3 and 4 rescale each measure to unit standard deviation. We report both because reported and recalibrated growth differ in dispersion. Coefficients in natural units are therefore not directly comparable across the two measures. We thus rely on standardized coefficients – the change in promotion probability per one-standard-deviation increase in performance – when contrasting the two. Reported GDP growth predicts promotion: a one-standard-deviation increase in reported growth relative to the provincial average raises the probability of promotion by 0.011 (Column 3), roughly 9% of the sample mean. This is consistent with previous research documenting a positive correlation between reported economic performance and

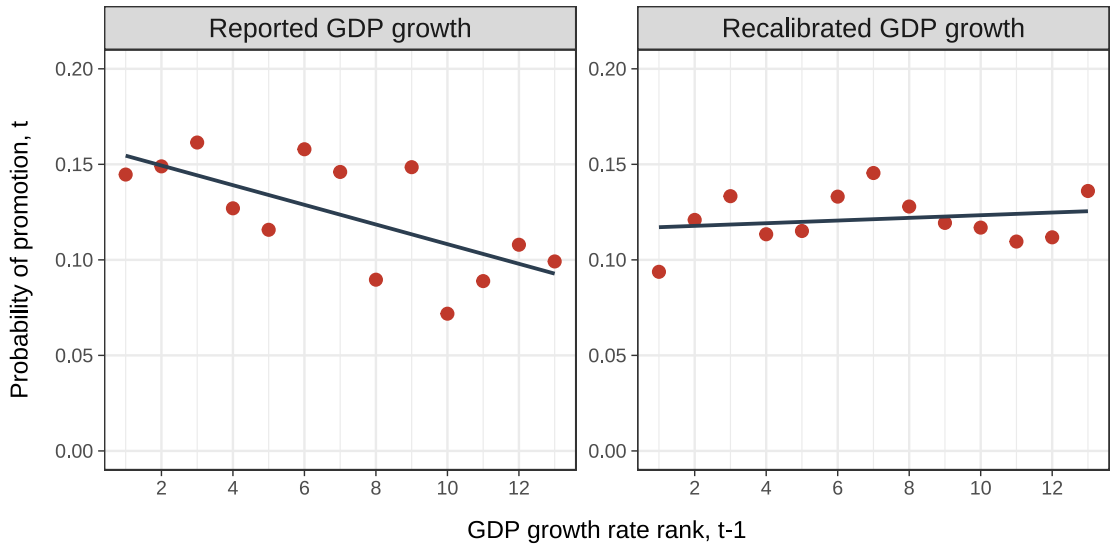

Figure 6: GDP growth and promotion probability.

the promotion of city party secretaries. Recalibrated growth, in contrast, bears no relationship to promotion (Columns 2 and 4): the estimate is small, statistically insignificant, and if anything slightly negative (-0.004 in natural units, -0.002 standardized). Actual economic performance, in other words, did not predict promotion – promoted secretaries were not systematically more capable of generating real growth.

The comparison between the standardized estimates bears directly on Hypothesis 1. In natural units the recalibrated coefficient carries a much larger standard error than the reported one, but only because the recalibrated measure varies less; once both measures are placed on a common scale, they are estimated with comparable precision, and the recalibrated effect is a tightly bounded near-zero whose confidence interval excludes the reported effect. This contrast is thus explained by the difference in effect size, rather than the precision of estimate linked to the recalibration itself.

Nor is the null a byproduct of how recalibrated growth is constructed. Because it is a model-fitted value, recalibrated growth is smoother and less variable than the reported series, raising the concern that smoothing itself – not the removal of manipulation – tends to produce smaller coefficients. We test this by applying the identical smoothing to reported growth, as the standard recalibration approaches do. This smoothed series predicts promotion nearly as strongly as the raw growth

Table 1: GDP Growth and the Promotion of City Party Secretaries

Performance measure:	*Promotion*			
	Levels		Standardized	
	(1)	(2)	(3)	(4)
Reported GDP growth	0.004**		0.011**	
	(0.002)		(0.005)	
Recalibrated GDP growth		-0.004		-0.002
		(0.007)		(0.004)
Individual characteristics	Yes	Yes	Yes	Yes
Dependent variable mean	0.123	0.120	0.123	0.120
Observations	3,655	3,750	3,655	3,750
Adjusted R^2	0.098	0.093	0.098	0.093
Province-Year FEs	Yes	Yes	Yes	Yes

***p < 0.01; **p < 0.05; *p < 0.1

figures, whereas recalibrated growth, smoothed in the same way but purged of manipulation, does not (Appendix Table A10). What separates the two is thus the manipulation removed, not the smoothing applied.

Nor does the pattern depend on the unit of analysis. Adopting the leader-term specification of Landry, Lü and Duan (2018) – one observation per secretary, and promotion at term completion regressed on average relative growth over the tenure – reported growth again predicts promotion, while recalibrated growth does not (Appendix Table A11, Columns 1–2). The contrast is thus not an artifact of the leader-year design.

5.3 GDP Fraud and Promotion

We now test Hypothesis 2 directly, using the gap between reported and recalibrated growth – our measure of over-reporting, which we showed earlier responds to manipulation incentives (Section 4.3) – to predict promotion. Table 2 shows that secretaries who over-reported more were more likely to be promoted. The estimate is positive and significant whether or not we control for recalibrated growth (Columns 1 and 2): holding constant the ability to produce real growth, party

Table 2: GDP Fraud and the Promotion of City Party Secretaries

Performance measure:	*Promotion*			
	Levels		Standardized	
	(1)	(2)	(3)	(4)
Manipulation	0.006**	0.008***	0.014**	0.017***
	(0.003)	(0.003)	(0.005)	(0.006)
Recalibrated GDP growth		-0.013		-0.008
		(0.008)		(0.005)
Individual characteristics	Yes	Yes	Yes	Yes
Dependent variable mean	0.122	0.122	0.122	0.122
Observations	3,553	3,553	3,553	3,553
Adjusted R^2	0.097	0.097	0.097	0.097
Province-Year FEs	Yes	Yes	Yes	Yes

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

secretaries more prone to manipulate were more likely to be promoted. In standardized terms, a one-standard-deviation increase in over-reporting raises the probability of promotion by 0.017 (Column 4), about 14% of the sample mean – larger than the standardized effect of reported growth in Table 1. Consistent with Hypothesis 2, under weak enforcement an ostensibly meritocratic promotion system in fact rewards those most willing to cheat.

5.4 Robustness and Extension

The two findings – that recalibrated growth does not predict promotion while over-reporting does – are robust to a range of alternative measurement, sample, and modeling choices (Appendix Table A12). Panel A shows that the recalibrated null holds throughout: the coefficient stays small and insignificant when we restrict the sample to a secretary's third and later years, when we recalibrate GDP with richer sets of economic proxies rather than nighttime light alone, when we measure performance by a secretary's most recent year rather than a running average, and when we drop sub-provincial cities. Panel B shows that the over-reporting effect remains positive and significant across the same perturbations, and under an alternative ratio-based measure of manipulation.

One might also worry that the association between over-reporting and promotion reflects patronage rather than manipulation: a secretary with a patron being the top leader of the province could both inflate figures with impunity and be promoted for reasons unrelated to those figures. We probe this with a standard proxy – whether a city party secretary was first appointed to the post by the province's current party secretary, the official with the most direct say over their advancement (Appendix Table A13). Adding this indicator to our main specification leaves the over-reporting coefficient essentially unchanged, so the relationship is not an artifact of connections.

A final question is whether this selection pattern was confined to our 2000–2013 window. Re-estimating both main specifications separately before and after 2013 – when central authorities signaled intention to scale back the weight of GDP growth in cadre evaluation – leaves the pattern intact (Appendix Table A14). Over-reporting continued to predict promotion at least as strongly after 2013 as before (the coefficient rises from 0.008 to 0.013), while recalibrated growth predicts promotion in neither period, and interaction tests detect no significant change in either relationship across the cutoff. The effort-prioritizing case we characterize thus holds across the sample: a signal of change in evaluation criteria on paper did not, at least not immediately, change how principals traded output against selection – which is why our setting offers no leverage on the trade-off itself, only on the one branch it realizes.

6 Conclusion

Performance-based promotion is widely taken to be meritocratic: by rewarding officials who deliver, it automatically selects the ablest for higher-level positions. We have shown that in the Chinese bureaucracy it did the opposite. Reported GDP growth predicted the promotion of city party secretaries, but the real growth buried inside those figures had little bearing on who rose; what predicted advancement was instead the gap between the two – the degree to which an official over-reported. The apparent meritocracy was an artifact of manipulation. Strip the inflation away, and promotion tracked not the capacity to generate growth but the willingness to fake it.

Our findings bear directly on a long-running debate over what drives elite promotion in China (Manion 2023), in which one camp stresses patronage connections (e.g., Shih, Adolph and Liu 2012) and another portrays the system as meritocratic (e.g., Bell 2016). The meritocratic interpretation rests largely on the documented correlation between reported growth and advancement. We find no evidence that the party-state systematically identifies and elevates officials who deliver superior economic performance; that correlation appears to be an artifact of statistical manipulation, dissolving once manipulation is held constant. Together with other recent evidence (Liu, Peng and Wang 2026), this cuts against the central premise of authoritarian meritocracy – that the able are identified and placed in higher office – and aligns instead with the view that autocrats have limited incentive to promote the most competent (Egorov and Sonin 2011; Zakharov 2016).

Notes

[1]The risk of detection also increases in actual output because inflating a larger output requires fabricating a larger absolute discrepancy, which is easier to detect.

[2]Proposition 2 concerns total effort; requiring each bureaucrat separately to exceed her own benchmark is a stronger condition. Both do so over some range of k if $c_2/c_1 > (3+\sqrt{5})/2 \approx 2.618$, against $1+\sqrt{2} \approx 2.414$ for the total (Appendix A1.3).

[3]Strictly, $D(k)$ approaches D^{NM} at both extremes of k. A politician who maximizes selection alone is therefore indifferent between the two. We treat suppression as the relevant case, since complete tolerance requires that every figure be inflated to the maximum.

[4]Secretaries typically hold the post for less than five years, so turnover is frequent: it occurs in every calendar year of our data, at an average annual rate of 27%.

[5]A similar gap appears for city mayor turnover, but it becomes insignificant once we condition on party secretary turnover – unsurprising given that the two often coincide; the mayoral gap thus appears to be driven by the secretary's turnover (Appendix Table A2).

[6]The two land-sales measures enter as log-differences rather than percentage growth rates, since land sales are zero in many city-years, which would leave a conventional growth rate undefined.

[7]We limit the sample to secretaries who served at least three years, excluding brief or interim appointments too short to establish a performance record.

[8]In a small number of cases we also code as promotions lateral moves to clearly more important positions at the same rank, such as appointment to the standing committee of a provincial party committee.

[9]When a city's growth is missing for a prior year, we average over the observed years, so the divisor is the number of non-missing prior years rather than $t - 1$.

References

Acemoglu, Daron, Leopoldo Fergusson, James Robinson, Dario Romero and Juan F Vargas. 2020. "The Perils of High-Powered Incentives: Evidence from Colombia's False Positives." *American Economic Journal: Economic Policy* 12(3):1–43.

Bell, Daniel A. 2016. *The China Model: Political Meritocracy and the Limits of Democracy*. Princeton University Press.

Bo, Zhiyue. 1996. "Economic Performance and Political Mobility: Chinese Provincial Leaders." *Journal of Contemporary China* 5(12):135–154.

Breton, Albert and Ronald Wintrobe. 1986. "The Bureaucracy of Murder Revisited." *Journal of Political Economy* 94(5):905–926.

Brown, Jennifer. 2011. "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars." *Journal of Political Economy* 119(5):982–1013.

Caixin Global. 2017. "Liaoning Government Admits False Growth Data from 2011–14." *Caixin Global* .

Catalinac, Amy, Bruce Bueno de Mesquita and Alastair Smith. 2020. "A Tournament Theory of Pork Barrel Politics: The Case of Japan." *Comparative Political Studies* 53(10-11):1619–1655.

Cattaneo, Matias D, Michael Jansson and Xinwei Ma. 2020. "Simple Local Polynomial Density Estimators." *Journal of the American Statistical Association* 115(531):1449–1455.

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

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

Chen, Zuoqi, Bailang Yu, Chengshu Yang, Yuyu Zhou, Shenjun Yao, Xingjian Qian, Congxiao Wang, Bin Wu, Jianping Wu, Lingxing Liao and Kaifang Shi. 2020. "The Global NPP-VIIRS-like Nighttime Light Data (Version 2) for 1992–2024.".

China Daily. 2018. "GDP Admissions by Local Govts Due to Necessity Not Sincerity." *China Daily* .

Eckhouse, Laurel. 2022. "Metrics Management and Bureaucratic Accountability: Evidence from Policing." *American Journal of Political Science* 66(2):385–401.

Egorov, Georgy and Konstantin Sonin. 2011. "Dictators and Their Viziers: Endogenizing the Loyalty–Competence Trade-Off." *Journal of the European Economic Association* 9(5):903–930.

Gehlbach, Scott and Alberto Simpser. 2015. "Electoral Manipulation as Bureaucratic Control." *American Journal of Political Science* 59(1):212–224.

Gregory, Paul R. 2009. *Terror by Quota: State Security from Lenin to Stalin (An Archival Study)*. New Haven: Yale University Press.

Harrison, Mark. 2011. "Forging Success: Soviet Managers and Accounting Fraud, 1943–1962." *Journal of Comparative Economics* 39(1):43–64.

Hassan, Mai. 2020. *Regime Threats and State Solutions: Bureaucratic Loyalty and Embeddedness in Kenya*. New York: Cambridge University Press.

He, Ning. 2026. "The Social Consequence of Bureaucratic Oversight: Evidence From the Great Chinese Famine." *Governance* 39(1):e70079.

He, Ning and Wenbing Wu. 2026. "The Shadow Cost of State Violence: Evidence from Bureaucratic Purges in China." *Comparative Political Studies* p. 00104140251400341.

Heinen, Sebastian. 2022. "Rwanda's Agricultural Transformation Revisited: Stagnating Food Production, Systematic Overestimation, and a Flawed Performance Contract System." *The*

Journal of Development Studies 58(10):2044–2064.

Höchtl, Wolfgang, Rudolf Kerschbamer, Rudi Stracke and Uwe Sunde. 2015. "Incentives and Selection in Promotion Contests: Is It Possible to Kill Two Birds with One Stone?" *Managerial and Decision Economics* 36(5):275–285.

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

Jia, Ruixue, Masayuki Kudamatsu and David Seim. 2015. "Political Selection in China: The Complementary Roles of Connections and Performance." *Journal of the European Economic Association* 13(4):631–668.

Kalgin, Alexander. 2016. "Implementation of Performance Management in Regional Government in Russia: Evidence of Data Manipulation." *Public Management Review* 18(1):110–138.

Landry, Pierre F. 2008. *Decentralized Authoritarianism in China: The Communist Party's Control of Local Elites in the Post-Mao Era*. Cambridge University Press.

Landry, Pierre F, Xiaobo Lü and Haiyan Duan. 2018. "Does Performance Matter? Evaluating Political Selection Along the Chinese Administrative Ladder." *Comparative Political Studies* 51(8):1074–1105.

Lazear, Edward P and Sherwin Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89(5):841–864.

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

Liu, Zhuang, Wenwei Peng and Shaoda Wang. 2026. "A Few Bad Apples? Academic Dishonesty, Political Selection, and Institutional Performance in China.". NBER working paper.

Manion, Melanie. 2023. *Political Selection in China: Rethinking Foundations and Findings*. Cambridge University Press.

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

Mattozzi, Andrea and Antonio Merlo. 2015. "Mediocracy." *Journal of Public Economics* 130:32–44.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142(2):698–714.

Montagnes, Pablo and Stephane Wolton. 2019. "Mass Purges: Top-Down Accountability in Autocracy." *American Political Science Review* 113(4):1045–1059.

Nove, Alec. 1958. "The Problem of Success Indicators in Soviet Industry." *Economica* 25(97):1–13.

Reuter, Ora John and Graeme B Robertson. 2012. "Subnational Appointments in Authoritarian Regimes: Evidence from Russian Gubernatorial Appointments." *The Journal of Politics* 74(4):1023–1037.

Rosen, Sherwin. 1986. "Prizes and Incentives in Elimination Tournaments." *American Economic Review* 76(4):701–715.

Sandefur, Justin and Amanda Glassman. 2015. "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics." *The Journal of Development Studies* 51(2):116–132.

Serrato, Juan Carlos Suárez, Xiao Yu Wang and Shuang Zhang. 2019. "The Limits of Meritocracy: Screening Bureaucrats under Imperfect Verifiability." *Journal of Development Economics* 140:223–241.

Shih, Victor, Christopher Adolph and Mingxing Liu. 2012. "Getting Ahead in the Communist Party: Explaining the Advancement of Central Committee Members in China." *American Political Science Review* 106(1):166–187.

The Economist. 2010. "Keqiang Ker-Ching: How China's Next Prime Minister Keeps Tabs on Its Economy." *The Economist* .

Vojnović, Milan. 2015. *Contest Theory: Incentive Mechanisms and Ranking Methods*. Cambridge University Press.

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

Weber, Max. 2019. *Economy and Society: A New Translation*. Harvard University Press.

Wintrobe, Ronald. 1998. *The Political Economy of Dictatorship*. Cambridge, UK; New York, NY:

Cambridge University Press.

Xinhua. 2017. "China to Apply Unified GDP Calculation Method." *Xinhua Net* .

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

Yao, Yang and Muyang Zhang. 2015. "Subnational Leaders and Economic Growth: Evidence from Chinese Cities." *Journal of Economic Growth* 20:405–436.

Zakharov, Alexei V. 2016. "The Loyalty-Competence Trade-Off in Dictatorships and Outside Options for Subordinates." *The Journal of Politics* 78(2):457–466.

Manipulation by Design: Performance Incentives and Adverse Selection in Bureaucracy

Online Appendix

Table of Contents

List of Tables

A1 Model Proof

A1.1 Equilibrium Manipulation

Proof of Proposition 1. Rewriting the utility in terms of reported output, $u_i = \Pr(r_i > r_j) - \widetilde{c}_i r_i$, turns the contest into a two-player all-pay auction with valuations $v_i = 1/\widetilde{c}_i(k)$; such a contest has a unique mixed-strategy equilibrium (Vojnović 2015, Ch. 1), which we use throughout.

To derive the effective cost, note that given any $r_i = (1+\rho_i)e_i$, a bureaucrat's cost has two parts: the cost of real effort $c_i e_i$ and the expected cost of being caught $k\rho_i^2 e_i$. Because the benefit depends only on r_i, we substitute $e_i = \frac{r_i}{1+\rho_i}$ and obtain the effective cost $\widetilde{c}_i = \frac{c_i + k\rho_i^2}{1+\rho_i}$.

The benefit depends only on r_i, while the effective cost $\widetilde{c}_i$ depends only on ρ_i; players therefore effectively choose r_i. Given r_i, a rational player chooses the ρ_i^* that minimizes $\widetilde{c}_i$ (adjusting e_i to hold r_i fixed). Differentiating, $\partial\widetilde{c}_i/\partial\rho_i = \frac{k\rho_i^2 + 2k\rho_i - c_i}{(1+\rho_i)^2}$, so the first-order condition is $k\rho_i^2 + 2k\rho_i - c_i = 0$; since this derivative changes sign from negative to positive at its root, the root is a minimum. Hence the optimum is $\rho_i^*(k) = \min\{\sqrt{1+\frac{c_i}{k}} - 1, 1\}$, and substituting back gives $\widetilde{c}_i(k) = \frac{c_i + k(\rho_i^*(k))^2}{1+\rho_i^*}$. Because $\rho_i^*(k)$ is increasing in c_i, the lower-ability bureaucrat manipulates more, $\rho_2^* \geq \rho_1^*$.

For the comparative statics in k, note that $\sqrt{1+c_i/k} - 1$ is strictly decreasing in k and exceeds 1 precisely when $c_i/k > 3$, that is, when $k < c_i/3$. Hence $\rho_i^*(k) = 1$ for $0 < k \leq c_i/3$, where it is constant, and $\rho_i^*(k) = \sqrt{1+c_i/k} - 1$ for $k > c_i/3$, where it is strictly decreasing. Each ρ_i^* is therefore non-increasing in k and strictly decreasing above $c_i/3$. Since $c_1 < c_2$, the higher-ability bureaucrat's manipulation begins to fall at a lower level of intolerance than her rival's. Substituting these values into $\widetilde{c}_i$ gives $\widetilde{c}_i(k) = (c_i + k)/2$ for $k \leq c_i/3$ and $\widetilde{c}_i(k) = 2k[\sqrt{1+c_i/k} - 1]$ for $k > c_i/3$, both strictly increasing in k, so the effective marginal cost of reported output rises with cheating intolerance. □

A1.2 Cheating Intolerance and Total Effort

Proposition 4 (Equilibrium total real effort)**.** *Suppose* $0 < c_1 < c_2$ *and* $k > 0$*. In the unique mixed-strategy Nash equilibrium, the expected efforts are*

$$\mathbb{E}[e_1^*] = \frac{1}{2\widetilde{c}_2(k)(1+\rho_1^*(k))} \qquad \mathbb{E}[e_2^*] = \frac{\widetilde{c}_1(k)}{2[\widetilde{c}_2(k)]^2(1+\rho_2^*(k))}.$$

Let $a_i(k) \equiv \sqrt{1+\frac{c_i}{k}}$*. The equilibrium total real effort is given by the following piecewise function:*

$$T(k) = \begin{cases} \dfrac{c_1+c_2+2k}{2(c_2+k)^2}, & 0 < k \le \dfrac{c_1}{3}, \\ \dfrac{1}{a_1(k)(c_2+k)} + \dfrac{2k[a_1(k)-1]}{(c_2+k)^2}, & \dfrac{c_1}{3} < k \le \dfrac{c_2}{3}, \\ \dfrac{a_2(k)+1}{4c_2a_1(k)} + \dfrac{c_1[a_2(k)+1]^2}{4c_2^2a_2(k)[a_1(k)+1]}, & k > \dfrac{c_2}{3}. \end{cases}$$

Proof of Proposition 4. As established in the proof of Proposition 1, the contest is a two-player all-pay auction with valuations $v_i = 1/\widetilde{c}_i(k)$. Its unique mixed-strategy equilibrium yields the expected efforts $\mathbb{E}[e_1^*] = \frac{1}{2\widetilde{c}_2(k)(1+\rho_1^*(k))}$ and $\mathbb{E}[e_2^*] = \frac{\widetilde{c}_1(k)}{2[\widetilde{c}_2(k)]^2(1+\rho_2^*(k))}$.

Let $a_i(k) \equiv \sqrt{1+\frac{c_i}{k}}$. The proof of Proposition 1 established the equilibrium manipulation levels and the effective costs they imply:

$$\rho_i^*(k) = \begin{cases} 1, & 0 < k \le \dfrac{c_i}{3}, \\ a_i(k)-1, & k > \dfrac{c_i}{3}, \end{cases} \qquad \widetilde{c}_i(k) = \begin{cases} \dfrac{c_i+k}{2}, & 0 < k \le \dfrac{c_i}{3}, \\ 2k[a_i(k)-1], & k > \dfrac{c_i}{3}. \end{cases}$$

Using the equilibrium expected-effort expressions, expected total real effort is

$$T(k) = \frac{1}{2\widetilde{c}_2(k)(1+\rho_1^*(k))} + \frac{\widetilde{c}_1(k)}{2[\widetilde{c}_2(k)]^2(1+\rho_2^*(k))}.$$

Because $c_1 < c_2$, the equilibrium manipulation levels define three regions. First, if $0 < k \le c_1/3$,

then $\rho_1^*(k) = \rho_2^*(k) = 1$, and $\widetilde{c}_1(k) = \frac{c_1+k}{2}$, and $\widetilde{c}_2(k) = \frac{c_2+k}{2}$. Therefore,

$$T(k) = \frac{1}{2(c_2+k)} + \frac{c_1+k}{2(c_2+k)^2} = \frac{c_1+c_2+2k}{2(c_2+k)^2}.$$

Second, if $c_1/3 < k \leq c_2/3$, then $\rho_1^*(k) = a_1(k) - 1$, and $\rho_2^*(k) = 1$. Thus, $\widetilde{c}_1(k) = 2k[a_1(k)-1]$, and $\widetilde{c}_2(k) = \frac{c_2+k}{2}$. Substituting into $T(k)$ gives

$$T(k) = \frac{1}{a_1(k)(c_2+k)} + \frac{2k[a_1(k)-1]}{(c_2+k)^2}.$$

Third, if $k > c_2/3$, then both bureaucrats choose interior manipulation: $\rho_i^*(k) = a_i(k) - 1$, $i \in \{1,2\}$. Hence, $\widetilde{c}_i(k) = 2k[a_i(k)-1]$. Substitution gives

$$T(k) = \frac{1}{4ka_1(k)[a_2(k)-1]} + \frac{a_1(k)-1}{4k[a_2(k)-1]^2 a_2(k)}.$$

Using $c_i = k[a_i(k)-1][a_i(k)+1]$, this can be rewritten as

$$T(k) = \frac{a_2(k)+1}{4c_2a_1(k)} + \frac{c_1[a_2(k)+1]^2}{4c_2^2a_2(k)[a_1(k)+1]}.$$

This proves the stated piecewise expression. □

A1.3 Manipulation and Total Effort

Proof of Proposition 2. Let $t \equiv \frac{c_2}{c_1} > 1$. The no-manipulation benchmark is $T^{\mathrm{NM}} = \frac{1}{2c_2}\left(1 + \frac{c_1}{c_2}\right) = \frac{c_1+c_2}{2c_2^2}$.

Consider the region $k > c_2/3$, where both bureaucrats choose interior manipulation. Using the expression from Proposition 4 and the expansion $\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + O(x^3)$, we obtain, as $k \to \infty$,

$$T(k) - T^{\mathrm{NM}} = \frac{t^2-2t-1}{8t^2} \cdot \frac{1}{k} + O(k^{-2}).$$

If $t > 1+\sqrt{2}$, then $t^2 - 2t - 1 > 0$. Therefore, the leading term in the expansion is positive.

It follows that there exists a finite $\bar{k}$ such that for every $k > \bar{k}$, $T(k) > T^{\mathrm{NM}}$. For any finite k, equilibrium manipulation is positive:

$$\rho_i^*(k) = \min\left\{\sqrt{1+\tfrac{c_i}{k}} - 1,\ 1\right\} > 0.$$

Thus, allowing a positive amount of manipulation can generate higher expected total real effort than the no-manipulation benchmark. □

The proposition establishes that some tolerance of manipulation raises total effort. The next result sharpens this into a statement about the effort-maximizing politician's choice: his optimal cheating intolerance is finite and strictly positive, so he deliberately leaves manipulation in place.

Corollary 1 (The effort-maximizing politician tolerates manipulation)**.** *Suppose $c_2/c_1 > 1+\sqrt{2}$. Then equilibrium total real effort $T(k)$ attains its maximum over $k \in (0,\infty)$ at a finite, strictly positive k^*, at which both bureaucrats manipulate, $\rho_2^*(k^*) \geq \rho_1^*(k^*) > 0$, and total effort strictly exceeds the no-manipulation benchmark, $T(k^*) > T^{\mathrm{NM}}$.*

Proof of Corollary 1. $T(k)$ is continuous on $(0,\infty)$, and its two boundary limits coincide with the benchmark. As $k \to \infty$, the expansion above gives $T(k) \to T^{\mathrm{NM}}$. As $k \to 0^+$, the region-1 expression from Proposition 4 gives

$$T(k) = \frac{c_1 + c_2 + 2k}{2(c_2+k)^2} \longrightarrow \frac{c_1+c_2}{2c_2^2} = T^{\mathrm{NM}}.$$

Extend T continuously to the compact interval $[0,\infty]$ by these limits; a continuous function on a compact set attains its maximum. By Proposition 2, $T(k) > T^{\mathrm{NM}}$ at some interior k when $c_2/c_1 > 1+\sqrt{2}$. The maximum therefore strictly exceeds the common boundary value T^{NM} and is attained at a finite $k^* \in (0,\infty)$, not at either endpoint.

To place k^* outside the maximal-manipulation region, note that on region 1 ($0 < k \leq c_1/3$),

$$T'(k) = -\frac{c_1+k}{(c_2+k)^3} < 0,$$

so T is strictly decreasing there and $k^* > c_1/3$. Because k^* is finite, $\rho_i^*(k^*) = \min\{\sqrt{1 + c_i/k^*} - 1,\ 1\} > 0$ for both bureaucrats, with $\rho_2^*(k^*) \geq \rho_1^*(k^*)$ by Proposition 1. The effort-maximizing politician thus sets a finite cheating intolerance that keeps manipulation active. □

Proposition 5 (Both bureaucrats exert more effort)**.** *Let e_i^{NM} denote bureaucrat i's expected real effort in the no-manipulation benchmark. If*

$$\frac{c_2}{c_1} > \varphi^2 = \frac{3 + \sqrt{5}}{2} \approx 2.618, \qquad \varphi \equiv \tfrac{1}{2}(1 + \sqrt{5}),$$

then there is a nonempty interval of cheating intolerance on which $\mathbb{E}[e_i^*] > e_i^{\mathrm{NM}}$ for both i, and hence $T(k) > T^{\mathrm{NM}}$.

Proof of Proposition 5. The decomposition $T^{\mathrm{NM}} = \frac{1}{2c_2}\left(1 + \frac{c_1}{c_2}\right)$ in the proof of Proposition 2 names each bureaucrat's benchmark effort, $e_1^{\mathrm{NM}} = \frac{1}{2c_2}$ and $e_2^{\mathrm{NM}} = \frac{c_1}{2c_2^2}$.

Consider the middle region $c_1/3 < k \leq c_2/3$, where $\rho_1^*(k) = a_1(k) - 1$ and $\rho_2^*(k) = 1$, so that $\widetilde{c}_1(k) = 2k[a_1(k) - 1]$ and $\widetilde{c}_2(k) = \frac{c_2 + k}{2}$. Substituting into the expected efforts of Proposition 4,

$$\mathbb{E}[e_1^*] = \frac{1}{a_1(k)(c_2 + k)}, \qquad \mathbb{E}[e_2^*] = \frac{2k[a_1(k) - 1]}{(c_2 + k)^2}.$$

Comparing each with its benchmark, and using $c_1 = k[a_1(k) - 1][a_1(k) + 1]$ to simplify the second,

$$\mathbb{E}[e_1^*] > e_1^{\mathrm{NM}} \iff \frac{2c_2}{c_2 + k} > a_1(k), \qquad \mathbb{E}[e_2^*] > e_2^{\mathrm{NM}} \iff \frac{2c_2}{c_2 + k} > \sqrt{1 + a_1(k)}.$$

Both hold if and only if $\frac{2c_2}{c_2 + k} > \max\{a_1(k), \sqrt{1 + a_1(k)}\}$. Since $a_1 \gtrless \sqrt{1 + a_1}$ according as $a_1 \gtrless \varphi$, the two requirements coincide at $a_1(k) = \varphi$, which by $c_1 = k(a_1^2 - 1)$ and $\varphi^2 - 1 = \varphi$ corresponds to $k = c_1/\varphi$. This value lies in the region considered: $c_1/3 < c_1/\varphi$ because $\varphi < 3$, and $c_1/\varphi \leq c_2/3$ because $c_2/c_1 > \varphi^2 > 3/\varphi$. Evaluating the requirement there,

$$\frac{2c_2}{c_2 + c_1/\varphi} > \varphi \iff c_2(2 - \varphi) > c_1 \iff \frac{c_2}{c_1} > \frac{1}{2 - \varphi} = \varphi^2,$$

the last equality because $2-\varphi=\varphi^{-2}$. Under the stated condition the inequality is strict, and both sides are continuous in k, so it holds on a neighborhood of c_1/φ. On that interval each bureaucrat exerts more real effort than in the benchmark; summing the two inequalities gives $T(k)>T^{\mathrm{NM}}$. $\square$

A1.4 Cheating Intolerance and the Promotion Gap

Proof of Proposition 3. In the unique mixed-strategy equilibrium of the two-player all-pay auction, the expected promotion probabilities are

$$\Pr(1 \text{ wins} \mid k) = 1 - \frac{\widetilde{c}_1(k)}{2\widetilde{c}_2(k)}$$

and

$$\Pr(2 \text{ wins} \mid k) = \frac{\widetilde{c}_1(k)}{2\widetilde{c}_2(k)}.$$

Therefore, $D(k) = 1 - \frac{\widetilde{c}_1(k)}{\widetilde{c}_2(k)}$.

Let $a_i(k) \equiv \sqrt{1+\frac{c_i}{k}}$. Recall from the proof of Proposition 4,

$$\widetilde{c}_i(k) = \begin{cases} \dfrac{c_i+k}{2}, & 0 < k \le \dfrac{c_i}{3}, \\ 2k[a_i(k)-1], & k > \dfrac{c_i}{3}. \end{cases}$$

Since $c_1 < c_2$, there are three regions.

First, if $0 < k \le c_1/3$, then both bureaucrats manipulate maximally. Thus,

$$D(k) = 1 - \frac{c_1+k}{c_2+k} = \frac{c_2-c_1}{c_2+k}.$$

Hence, $D'(k) = -\frac{c_2-c_1}{(c_2+k)^2} < 0$.

Second, if $c_1/3 < k \le c_2/3$, then $\rho_1^*(k) = a_1(k) - 1$, $\rho_2^*(k) = 1$. Therefore, $D(k) =$

$1 - \frac{4k[a_1(k)-1]}{c_2+k}$. Let $R(k) \equiv \frac{4k[a_1(k)-1]}{c_2+k}$. Then $D'(k) = -R'(k)$. Differentiating gives

$$\frac{R'(k)}{R(k)} = \frac{a_1(k)-1}{2ka_1(k)} - \frac{1}{c_2+k}.$$

Thus $R'(k) = 0$ if and only if $[a_1(k)-1](c_2+k) = 2ka_1(k)$. Let $t = c_2/c_1$. Using $k = \frac{c_1}{[a_1(k)-1][a_1(k)+1]}$, the preceding equality becomes $t[a_1(k)-1]^2 = 1$. Hence, $a_1(k) - 1 = \frac{1}{\sqrt{t}}$. The corresponding value of k is $k^\dagger = \frac{c_1 t}{1+2\sqrt{t}} = \frac{c_2}{1+2\sqrt{c_2/c_1}}$. Since $t > 1$, $\frac{c_1}{3} < k^\dagger < \frac{c_2}{3}$. Moreover, $R'(k) > 0$ for $k < k^\dagger$ and $R'(k) < 0$ for $k > k^\dagger$. Therefore,

$$D'(k) < 0 \quad \text{for} \quad \frac{c_1}{3} < k < k^\dagger,$$

and

$$D'(k) > 0 \quad \text{for} \quad k^\dagger < k \leq \frac{c_2}{3}.$$

Third, if $k > c_2/3$, then both bureaucrats choose interior manipulation. In this region,

$$D(k) = 1 - \frac{c_1[a_2(k)+1]}{c_2[a_1(k)+1]}.$$

Let $R(k) \equiv \frac{c_1[a_2(k)+1]}{c_2[a_1(k)+1]}$. Then

$$\frac{d}{dk}\log R(k) = -\frac{c_2}{2k^2 a_2(k)[a_2(k)+1]} + \frac{c_1}{2k^2 a_1(k)[a_1(k)+1]}.$$

Because $c_2 > c_1$, we have $a_2(k) > a_1(k)$. Moreover, $\frac{c_i}{a_i(k)[a_i(k)+1]} = k\frac{a_i(k)-1}{a_i(k)}$ is increasing in $a_i(k)$. Therefore, $\frac{c_2}{a_2(k)[a_2(k)+1]} > \frac{c_1}{a_1(k)[a_1(k)+1]}$. It follows that $\frac{d}{dk}\log R(k) < 0$, so $R'(k) < 0$, and hence $D'(k) > 0$ for all $k > c_2/3$.

Combining the three regions, $D(k)$ decreases on $(0, k^\dagger)$, reaches its minimum at $k^\dagger$, and increases on $(k^\dagger, \infty)$.

It remains to compare D with the benchmark. Both boundary limits equal D^{NM}. From the region-1 expression, $D(k) = \frac{c_2-c_1}{c_2+k} \to 1 - \frac{c_1}{c_2} = D^{\mathrm{NM}}$ as $k \to 0^+$, where $\rho_1^* = \rho_2^* = 1$; from the

region-3 expression, $D(k) = 1 - \frac{c_1[a_2(k)+1]}{c_2[a_1(k)+1]} \to 1 - \frac{c_1}{c_2} = D^{\mathrm{NM}}$ as $k \to \infty$, where $a_i(k) \to 1$ and $\rho_1^* = \rho_2^* = 0$. Because D falls strictly below this common value just to the right of 0 and returns to it only in the limit, $D(k) < D^{\mathrm{NM}}$ at every finite $k > 0$. The two limits arise for different reasons. As $k \to \infty$, manipulation itself vanishes, so nothing distorts the comparison. As $k \to 0^+$, manipulation is maximal but common to the two bureaucrats and costless: the factor $1 + \rho_i^*$ is then shared and cancels from $\widetilde{c}_1/\widetilde{c}_2$, while the expected penalty $k(\rho_i^*)^2 = k$ vanishes with it. Symmetry alone does not suffice – at any fixed $k > 0$ with $\rho_1^* = \rho_2^*$, the common penalty is added to both c_1 and c_2 and leaves $\widetilde{c}_1/\widetilde{c}_2 > c_1/c_2$. □

A2 Recalibration Methods Comparison

The standard approach estimates manipulation by fitting a model of reported GDP growth on proxies for economic activity and treating the gap between reported and predicted growth as manipulation. We show here that this procedure is biased whenever manipulation is correlated with actual growth, because the model is fit on the manipulated figures themselves.

Let Y denote reported GDP growth and X the predictors. Reported growth decomposes into three components: actual GDP growth explained by the predictors, $g(X)$; manipulation M; and a random error ε satisfying $E(\varepsilon \mid X) = 0$.

$$Y = g(X) + M + \varepsilon \tag{6}$$

For the predicted outcome not to capture manipulation, the procedure relies on the assumption that M is uncorrelated with the predictors X. This assumption can be violated even when the selected predictors are not themselves susceptible to manipulation. For instance, when manipulation is correlated with actual GDP growth, and actual growth is in turn captured by X, manipulation and the predictors cannot be independent.

Taking conditional expectations in (6) makes the consequence immediate:

$$E(Y \mid X) = g(X) + \hat{M}(X), \qquad \hat{M}(X) \equiv E(M \mid X). \tag{7}$$

A model fit on manipulated data estimates $E(Y \mid X)$, so its predicted values recover not $g(X)$ but $g(X) + \hat{M}(X)$. The recalibrated series is contaminated by exactly $\hat{M}(X)$, the component of manipulation the predictors can explain, which is nonzero precisely when M and X are correlated. Because $\hat{M}(X)$ moves with X whenever manipulation tracks actual growth, the bias runs toward retaining manipulation in the recalibrated figures rather than removing it. This is what our placebo exercise finds: a series constructed in this way retains enough manipulation to predict promotion nearly as strongly as the raw reported figures (Appendix Table A10).

The same quantity displaces the residual. Writing $u = Y - E(Y \mid X)$ for the model residual

and substituting (6) and (7),

$$u = \left[M - \hat{M}(X)\right] + \varepsilon, \tag{8}$$

so the residual recovers only the portion of manipulation the predictors fail to explain. It understates true manipulation by exactly $\hat{M}(X)$, and carries the random error besides, so it may reflect factors other than overreporting. Both halves of the standard procedure are therefore displaced by the same term: $\hat{M}(X)$ is added to the estimate of actual growth and subtracted from the estimate of manipulation. The displacement vanishes only under the independence assumption the procedure itself cannot verify, which is precisely what our turnover-year design is designed to avoid relying upon.

A3 Tables

Table A1: GDP Growth in the Turnover Year and Later Promotion

	Promotion			
	(1)	(2)	(3)	(4)
Constant	0.087***	0.117***	0.153***	0.299***
	(0.010)	(0.012)	(0.019)	(0.030)
GDP growth in turnover year	0.003	0.004	0.010	0.014
	(0.003)	(0.004)	(0.006)	(0.010)
Year of tenure	3rd	4th	5th	6th
Dependent variable mean	0.086	0.116	0.150	0.294
Observations	904	630	428	272

Notes: The regressor is the incoming secretary's city GDP growth in the turnover (first) year of the spell, expressed as a deviation from the province-year average. The outcome is an indicator for promotion, and each column estimates the relationship separately in the third through sixth year of tenure. Estimated by OLS with standard errors clustered by city in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A2: Leadership Turnover and Reported GDP Growth

	Reported GDP growth		
	(1)	(2)	(3)
Party secretary turnover	-0.477***		-0.417***
	(0.129)		(0.140)
Mayor turnover		-0.304**	-0.148
		(0.124)	(0.134)
Dependent variable mean	12.5	12.5	12.5
Observations	4,341	4,341	4,341
Adjusted R^2	0.386	0.384	0.386
City FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes

Notes: The outcome is a city's reported GDP growth rate in percentage points. *Party secretary turnover* and *Mayor turnover* are indicators equal to one in a calendar year during which the incumbent leaves office. Column 1 enters secretary turnover alone, Column 2 mayor turnover alone, and Column 3 both together. The mayoral gap is about two-thirds the size of the secretary's on its own and loses significance once secretary turnover is conditioned on, while the secretary's estimate remains large and significant. Because the two offices frequently turn over in the same year, this indicates the mayoral gap reflects the secretary's departure rather than the mayor's. All columns include city and year fixed effects, with standard errors clustered by province-year. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A3: Density Discontinuity Around Integers Values of Reported GDP Growth

	Turnover-year data		Incumbent-year data	
Integer value	Estimate	p-value	Estimate	p-value
10	0.050	0.085	0.052	0.013
11	-0.030	0.346	0.010	0.681
12	0.019	0.573	0.069	0.009
13	0.052	0.100	0.077	0.002
14	-0.003	0.927	0.009	0.726
15	0.040	0.225	0.068	0.005
16	0.006	0.808	0.047	0.013

Notes: Statistical tests for density discontinuity around integer values of reported GDP growth rates. Each estimate represents a density discontinuity (right minus left) around an integer value. Lower p-values (e.g., $p < 0.05$) indicate a statistically significant discontinuity.

Table A4: Recalibration Models and Results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Predictor importance										
Province dummies	57.62	44.34	43.26	38.88	36.57	26.60	25.44	27.57	28.20	27.23
Year dummies	83.49	79.22	57.66	49.89	43.84	39.87	40.44	28.53	23.53	22.68
Night light (VIIRS-like)	29.49		19.67	16.51	13.81	10.75	10.16	9.02	12.50	11.34
Night light (DMSP)		24.65								9.73
Export			11.96	8.66	4.18	4.43	2.77	2.59	1.51	2.59
Land sold (area)				9.34	6.40	5.65	5.48	1.71	7.70	6.44
Land sold (value)				12.66	7.73	6.52	4.52	1.69	5.53	5.42
New firms					5.59	9.41	7.36	4.75	2.17	2.07
Value-added tax						13.65	11.96	12.22	13.41	13.21
Passengers							3.75	0.66	2.70	3.45
Freight							5.13	4.84	4.57	3.09
Consumer good sales								20.21	17.13	15.27
Loan									3.45	3.66
Model results										
Mean (predicted GDP growth)	12.03	11.94	12.28	12.25	12.27	12.25	12.27	12.49	12.73	12.74
Mean (reported GDP growth)	12.47	12.43	12.75	12.73	12.75	12.75	12.75	12.96	13.21	13.20
Observations predicted	4,508	3,932	4,077	4,015	3,957	3,418	3,354	3,098	2,563	2,553
R-squared	0.39	0.39	0.34	0.33	0.31	0.35	0.34	0.37	0.40	0.40

Notes: Each column is a random-forest model that predicts reported GDP growth from economic indicators. Every model is trained on turnover-year city observations – the years a city party secretary takes office, when reporting is least distorted – and then applied to all city-years to recalibrate growth. Models 1 and 2 are alternatives, each using one nighttime light series alone; Models 3–9 add predictors cumulatively to Model 1; and Model 10 adds the second light series to Model 9. Coverage narrows as predictors with more missing values enter, though Model 2 stands outside this sequence and is narrower than Models 3–5. Model 1, our baseline, uses the NPP-VIIRS-like nighttime light series alone (Chen et al. 2020) with province and year fixed effects; Model 2 substitutes the DMSP-OLS light series; Models 3–9 add the external, manipulation-resistant proxies (exports, land sales, and new firm registrations) and then the official statistics (value-added tax, passenger and freight volumes, consumer sales, and bank loans); Model 10 includes every predictor, both light series among them. All predictors enter as annual growth rates, with the two land-sales measures entering as log-differences to accommodate the many city-years with zero recorded sales. The upper panel reports predictor importance, measured by the average reduction in accuracy across out-of-bag predictions when a feature is permuted (Mean Decrease Accuracy); a blank cell means the predictor is absent from that model. The lower panel reports the mean predicted and reported GDP growth over the recalibrated sample, the number of city-years that can be recalibrated (coverage, which shrinks as predictors with more missing values are added), and the out-of-bag R^2.

Table A5: Density Discontinuity at the Announced Growth Target

Growth measure	Observations	Estimate	p-value
Reported	2435	0.208	0.002
Recalibrated	2439	0.040	0.565
Standard method	2435	0.111	0.087

Notes: Statistical tests for density discontinuity at the growth target announced in the city government work report at the beginning of the year. The running variable is a city-year's growth rate minus its announced target, scaled to unit standard deviation so that the rows are comparable, and the cutoff is zero. Each estimate represents a density discontinuity (right minus left) at the cutoff, estimated by local linear regression with a common bandwidth of half a standard deviation. "Reported" uses the official GDP growth figures, "Recalibrated" our recalibrated figures, and "Standard method" the figures produced by the same random forest with the same predictors trained on all city-years rather than on turnover years, taking out-of-bag predictions so that the series is a genuine prediction rather than an in-sample fitted value. Lower p-values (e.g., $p < 0.05$) indicate a statistically significant discontinuity.

Table A6: External Validation: Recalibrated Growth and Independent Real Activity

Specification	N	Reported	Recalibrated
Baseline			
Equal-weight index, no FE	3,980	0.14 [0.11, 0.17]	0.15 [0.12, 0.18]
Fixed effects			
Year	3,980	0.06 [0.03, 0.09]	0.05 [0.02, 0.08]
Province + year	3,980	0.07 [0.03, 0.10]	0.06 [0.03, 0.09]
Province $\times$ year	3,980	0.03 [0.00, 0.07]	0.02 [-0.02, 0.05]
City + year	3,980	0.07 [0.03, 0.10]	0.07 [0.03, 0.10]
Index construction			
Require $\geq$2 proxies	4,210	0.14 [0.11, 0.17]	0.14 [0.11, 0.17]
Require $\geq$4 proxies	3,929	0.14 [0.11, 0.17]	0.15 [0.12, 0.18]
First principal component	3,980	0.08 [0.05, 0.11]	0.07 [0.04, 0.10]
Including VAT	4,205	0.21 [0.18, 0.23]	0.21 [0.18, 0.24]
Robustness			
Spearman rank correlation	3,980	0.18 [0.15, 0.21]	0.16 [0.13, 0.19]

Notes: Each cell is the partial correlation between a composite real-activity index and a GDP growth measure – reported or recalibrated – with the 95% confidence interval (Fisher z) in brackets. The index is the equal-weighted average of the standardized annual growth rates of six real-activity indicators that are external to both measures: exports (foreign trade), land sold by area, land sold by value, freight volume, passenger volume, and new firm registrations. Nighttime light is excluded because it constructs the recalibration, and value-added tax is excluded because it is mechanically part of GDP accounting; the "Including VAT" row adds VAT to the index. The baseline, index-construction, and Spearman rows use no fixed effects, while each fixed-effects row residualizes the index and the measure on the fixed effects named before computing the correlation. Index-construction rows vary the number of non-missing proxies required to form the index or replace the equal-weighted average with its first principal component. N is the number of city-years.

Table A7: What Explained the GDP Figure Manipulation?

	Gap		Ratio	
Variable	Estimate	p-value	Estimate	p-value
Tenure	0.046	**0.025**	0.006	0.136
Age	0.035	**0.013**	0.003	0.223
Higher rank	0.309	0.114	0.046	**0.034**
Male	0.198	0.585	0.026	0.545
Ethnic Han	-0.406	0.177	0.035	0.625
Education level	-0.174	0.053	-0.031	**0.046**
Provincial patron	-0.075	0.420	0.000	0.982

Notes: Each row reports a separate bivariate regression of a manipulation measure on the listed characteristic of the city party secretary. The outcome is either the gap (reported minus recalibrated GDP growth) or the ratio (reported over recalibrated). "Higher rank" indicates a vice-ministerial (sub-provincial) secretary, and "Provincial patron" indicates a secretary first appointed to the post by the current provincial party secretary. Estimates significant at the 5% level are in bold. Standard errors clustered by city.

Table A8: Over-Reporting Over a Secretary's Term

	(1)	(2)	(3)	(4)
Year of tenure = 2	0.197	0.427***	0.050	0.042**
	(0.121)	(0.102)	(0.035)	(0.018)
Year of tenure = 3	0.441***	0.550***	0.070**	0.043
	(0.085)	(0.092)	(0.028)	(0.034)
Year of tenure = 4	0.354***	0.504***	0.052	0.037
	(0.088)	(0.092)	(0.032)	(0.023)
Year of tenure = 5	0.325***	0.586***	0.053*	0.052**
	(0.124)	(0.125)	(0.032)	(0.021)
Year of tenure = 6	0.291***	0.376***	0.046	0.039*
	(0.109)	(0.133)	(0.032)	(0.022)
Year of tenure = 7	0.198	0.218	0.039	0.021
	(0.161)	(0.179)	(0.027)	(0.024)
Outcome	Gap	Gap	Ratio	Ratio
Dependent variable mean	0.348	0.359	1.02	1.02
Observations	5,035	4,882	5,035	4,882
Adjusted R^2	0.126	0.306	-0.028	0.146
Province-Year FEs	Yes	No	Yes	No
Secretary FEs	No	Yes	No	Yes

Notes: The outcome is a manipulation measure – the gap (reported minus recalibrated GDP growth) in Columns 1–2 and the ratio (reported over recalibrated) in Columns 3–4. Each coefficient is average over-reporting in that year of a secretary's tenure, relative to the turnover (first) year, which is the omitted reference; the seventh year of tenure and beyond are combined. Columns 1 and 3 include province-by-year fixed effects; Columns 2 and 4 include secretary fixed effects, comparing each secretary against their own turnover year. Standard errors clustered by city in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A9: Target Ambition and Over-Reporting

	Manipulation (gap)			
	(1)	(2)	(3)	(4)
Target − recalibrated growth	0.444***			
	(0.047)			
Target − lagged recalibrated growth		0.329***		
		(0.042)		
Change in announced target			0.155***	
			(0.036)	
Announced growth target				0.493***
				(0.048)
Recalibrated GDP growth				-0.275***
				(0.094)
Dependent variable mean	0.664	0.664	0.670	0.664
Observations	2,430	2,430	2,066	2,430
Adjusted R^2	0.346	0.288	0.254	0.356
City FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes

Notes: The outcome is the gap between reported and recalibrated GDP growth. The announced growth target is taken from the city government work report at the beginning of the year and is therefore predetermined with respect to realized growth. Column 1 measures ambition as the target minus the same year's recalibrated growth; because this regressor shares the recalibrated figure with the outcome, measurement error in the recalibration biases it upward. Columns 2 and 3 avoid this, measuring ambition as the target minus the previous year's recalibrated growth and as the year-on-year change in the announced target; Column 4 enters the target directly while conditioning on recalibrated growth. All columns include city and year fixed effects, with standard errors clustered by province-year in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A10: Placebo: The Recalibrated Null Is Not an Artifact of Smoothing

	Promotion		
	(1)	(2)	(3)
Reported GDP growth	0.011**		
	(0.005)		
Smoothed reported (standard method)		0.009**	
		(0.004)	
Recalibrated GDP growth			-0.002
			(0.004)
Individual characteristics	Yes	Yes	Yes
Dependent variable mean	0.123	0.120	0.120
Observations	3,655	3,750	3,750
Adjusted R^2	0.098	0.093	0.093
Province-Year FEs	Yes	Yes	Yes

Notes: The outcome is an indicator for promotion, and all performance measures are standardized to unit standard deviation, so coefficients are comparable across columns. "Reported GDP growth" is the raw figure. "Smoothed reported (standard method)" applies the same random-forest smoothing as our recalibration – the same predictors (province and year effects and nighttime light) and procedure – but trains on all city-years, the standard proxy approach, rather than on turnover years; it therefore shares the recalibration's smoothing while retaining manipulation. "Recalibrated GDP growth" trains on turnover years. That the smoothed-reported series predicts promotion nearly as strongly as reported growth, while recalibrated growth does not, shows the recalibrated null reflects the removal of manipulation, not the loss of variance from smoothing. Each column is a separate regression with individual characteristics and province-by-year fixed effects, and standard errors clustered by city in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A11: Leader-Term Robustness: Promotion at Term Completion

	Promotion at term completion			
Sample period	2000–2013		≤2007	
	(1)	(2)	(3)	(4)
Reported GDP growth	0.041***		0.040	
	(0.015)		(0.027)	
Recalibrated GDP growth		0.025		0.021
		(0.016)		(0.028)
Personal characteristics	Yes	Yes	Yes	Yes
Dependent variable mean	0.461	0.461	0.412	0.412
Observations	775	775	255	255
Adjusted R^2	0.140	0.136	0.178	0.170
Province FEs	Yes	Yes	Yes	Yes
Start year FEs	Yes	Yes	Yes	Yes

Notes: The unit of analysis is the leader term, following Landry, Lü and Duan (2018): the outcome is an indicator equal to one if the secretary was promoted at the completion of their term. Performance is the secretary's GDP growth relative to the mean of competing cities in the same province-year, averaged over the secretary's entire tenure; coefficients are scaled to a one-standard-deviation change in this measure. Columns 1 and 3 use reported growth and Columns 2 and 4 recalibrated growth. Columns 1–2 span the full 2000–2013 window; Columns 3–4 restrict to the pre-2008 window covered by earlier studies. Reported growth predicts promotion over the full window (Column 1) but not over the shorter pre-2008 window (Column 3), which locates the difference from earlier leader-term findings in the sample period rather than in the specification. All columns include political connection (a secretary first appointed to the post by the current provincial party secretary), age, age squared, tenure, and tenure squared, with province and start-year (cohort) fixed effects; standard errors are clustered by city in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A12: Robustness to Alternative Measurement, Sample, and Model Specification

	Promotion						
	Baseline	Tenure ≥ 3	Recal. (ext.)	Recal. (full)	Recent year	Excl. sub-prov.	Ratio
Panel A: Actual performance and promotion							
Recalibrated GDP growth	-0.004	0.008	0.004	0.005	-0.008	0.001	—
	(0.007)	(0.014)	(0.004)	(0.005)	(0.006)	(0.007)	
Observations	3,750	2,791	3,419	2,302	3,750	3,286	
Panel B: Manipulation and promotion							
Manipulation	0.008***	0.008**	0.010***	0.009**	0.006***	0.007**	0.056*
	(0.003)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)	(0.033)
Observations	3,553	2,645	3,299	2,302	3,546	3,097	3,553
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column re-estimates the two main specifications under a single change to measurement, sample, or model. Panel A reports the coefficient on recalibrated growth (as in Table 1); Panel B the coefficient on over-reporting, controlling for recalibrated growth (as in Table 2). Columns follow the order of the table: *Baseline* reproduces the main specifications; *Tenure* $\geq$ *3* drops second-year observations; *Recal. (ext.)* and *Recal. (full)* recalibrate GDP with progressively richer predictor sets; *Recent year* replaces the running average of performance with its most recent year; *Excl. sub-prov.* drops sub-provincial cities; and *Ratio* (Panel B only) measures manipulation as the ratio of reported to recalibrated growth. All specifications include individual characteristics and province-by-year fixed effects, with standard errors clustered by city. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A13: Over-Reporting and Promotion, Controlling for Patronage

	Promotion		
	(1)	(2)	(3)
Manipulation	0.008***	0.007***	0.007**
	(0.003)	(0.003)	(0.003)
Recalibrated GDP growth	-0.014	-0.014*	-0.012
	(0.008)	(0.008)	(0.008)
Provincial patron		-0.132***	-0.044**
		(0.017)	(0.018)
Individual characteristics	Yes	Yes	Yes
Dependent variable mean	0.122	0.122	0.122
Observations	3,541	3,541	3,540
Adjusted R^2	0.097	0.113	0.158
Province-Year FEs	Yes	Yes	Yes
Tenure FEs	No	No	Yes

Notes: The outcome is an indicator for promotion. Column 1 reproduces the specification of Table 2, Column 2, on the subsample with non-missing patronage; Column 2 adds a patronage indicator equal to one if the city party secretary was first appointed to that post by the current provincial party secretary; Column 3 further absorbs year-of-tenure fixed effects. The over-reporting coefficient is essentially unchanged by the patronage control. The negative patronage coefficient reflects the timing of the measure – a secretary is recorded as connected early in the spell, while the appointing provincial secretary is still in office, but is promoted later – and it shrinks by about two-thirds once tenure is absorbed (Column 3); it should not be read as an effect of patronage on promotion. All columns include individual characteristics and province-by-year fixed effects, with standard errors clustered by city in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A14: Performance and Promotion, Before and After 2013

	Promotion					
	(1)	(2)	(3)	(4)	(5)	(6)
Recalibrated GDP growth	-0.004	0.004	-0.001	-0.013	-0.013	-0.012**
	(0.007)	(0.009)	(0.006)	(0.008)	(0.010)	(0.006)
Recalibrated GDP growth × Post-2013			0.003			
			(0.010)			
Manipulation				0.008***	0.013***	0.009***
				(0.003)	(0.004)	(0.003)
Manipulation × Post-2013						0.004
						(0.004)
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Performance measure	Real growth	Real growth	Real growth	Manipulation	Manipulation	Manipulation
Sample	≤2013	>2013	All	≤2013	>2013	All
Dependent variable mean	0.120	0.108	0.114	0.122	0.108	0.115
Observations	3,750	3,503	7,253	3,553	3,492	7,045
Adjusted R^2	0.093	0.072	0.080	0.097	0.074	0.084
Province-Year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The two main promotion specifications, re-estimated separately before and after 2013. Columns 1–3 use recalibrated GDP growth alone (as in Table 1, Column 2); Columns 4–6 add over-reporting while controlling for recalibrated growth (as in Table 2, Column 2). Within each block, the first two columns are estimated on the ≤2013 and >2013 samples and the third pools all years and interacts the performance measure with a post-2013 indicator (the "× Post-2013" rows); the indicator's main effect is absorbed by the province-by-year fixed effects. All specifications include individual characteristics and province-by-year fixed effects, with standard errors clustered by city. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$