Curbing Bureaucratic Information Manipulation*

Yongwei Nian

May 8, 2023

Abstract

This paper examines the effect of a 2009 statistics reform in China aimed at curbing economic data manipulation by local governments. The detection of manipulation relied on survey teams pre-deployed in 40% of the counties, enabling a difference-in-differences design. Using newly collected county-level data from 2005 to 2018, I find a drop in GDP growth manipulation amounting to 5% of reported GDP growth in the treatment counties after the reform, as measured by the discrepancy between reported GDP growth and nighttime light intensity growth. Instrumenting the treatment counties using the random assignment of earlier rural survey teams in the 1980s generates similar results. Such a drop is mostly consistent with the reform generating a disciplining effect on local governments. Meanwhile, government policies shifted in directions conducive to economic growth, bank loans expanded, firm entry increased, and citizens' attitudes towards local governments improved. Corruption did not respond to the reform. These findings highlight the cost of bureaucrat misbehaviors distinct from corruption.

*Department of Economics, Bocconi University. Email: yongwei.nian@phd.unibocconi.it. Website: www.yongweinian.com. I thank Stefano Fiorin, Guido Tabellini, and Paolo Pinotti for their invaluable advice. I also thank Jan Bakker, Jean-Marie Baland, Martina Björkman Nyqvist, Luca Braghieri, Alexia Delfino, Sarah Eichmeyer, Marco Ottaviani, Eleonora Patacchini, Jaime Marques Pereira, Carlo Schwarz, Liyang Sun, Leonard Wantchekon, and all participants at Bocconi DLPE seminar and Bocconi Political Economy Breakfast for their useful discussions and suggestions.

1 Introduction

Bureaucracies are pivotal to economic development, a view that dates back at least to Max Weber and is further stressed by a growing economics literature (Weber, 1922; Jones and Olken, 2005; Besley et al., 2021). However, lower-level bureaucrats may have incentives to manipulate information, especially when the information is instrumental to their career advancement but imperfectly observed, thereby undermining bureaucratic effectiveness. This issue is particularly severe in China due to its heavy reliance on GDP growth to evaluate local leaders and the lack of alternative sources of information gathering such as media exposure and protests (Wallace, 2016; Xiong, 2018; Martinez, 2021). Such information manipulation may further impact the implementation of economic policies (Serrato, Wang and Zhang, 2019). Despite the cost, there is relatively little causal evidence on how to effectively reduce information manipulation.

In this paper, I study a statistics reform in China initiated in 2009 seeking to discipline local governments in processing statistics data. To this end, the National Bureau of Statistics (NBS) relied on survey teams in 40% of counties to check and verify the local statistics data. Upon detection of misbehaviors, the survey teams could directly report to the NBS and other higher-level supervisory authorities to punish those involved (e.g., local statistics bureaus or local officials). Crucially, these survey teams were initially deployed by the NBS in 2005 with the purpose of collecting several economic data, such as grain output and CPI, instead of disciplining local officials. Hence, the distribution of the survey teams was largely orthogonal to bureaucrat misbehaviors. Furthermore, the team members were appointed and led directly by the NBS, shielding them from local political interference. To identify the effects of the reform, I assemble a county-level dataset from 2005 to 2018 and employ a difference-in-differences design comparing counties with these survey teams (treatment) to counties without these survey teams (control) before and after 2009.

I document several sets of results. First, I show that compared to control counties, treatment counties experienced a significant drop in GDP growth manipulation, which is measured as the discrepancy between reported GDP growth and nighttime light intensity growth in the framework of Henderson, Storeygard and Weil (2012). I confirm that these two groups of counties were on the same trajectory before the reform using an event study specification. In terms of economic magnitude, in the most stringent specification with a set of preexisting demographic, economic, and geographic controls interacted with the post-reform dummy, I find a 0.58 percentage points drop in GDP growth manipulation in treatment counties relative to control counties, which amounts to 5.3% of the mean of reported GDP growth. This finding is robust to allowing the mapping between economic activities and light to be nonlinear or depend on various temporal and spatial characteristics, and robust to accounting for potential diverging trends between treatment and control counties by including county-specific time trends or re-weighting the observations to reach covariate balance, among a battery of other conventional specification tests.

To further tighten identification, I conduct two additional tests. The first is a placebo event study around 2005 when the survey teams were launched but had not conducted any disciplining actions. This test helps to further rule out pre-trends and alleviate concerns that these teams per se may affect my outcomes, even in the absence of the reform in 2009. Reassuringly, I find no diverging trends between treatment and control counties until 2009. The second is an instrumental variable estimation. By checking local gazetteers, I find that most of these survey teams launched in 2005 were restructured from earlier rural survey teams set up in 1984,[1] whose assignment at that time was done through a systematic random sampling of counties within provinces. The random assignment of counties with these earlier rural survey teams hence forms a valid instrument. The validity is further supported by balance tests showing that counties with these earlier rural survey teams were no different from other counties on a battery of demographic, economic, and geographic characteristics. The difference-in-differences estimates using this instrument have no substantial changes, suggesting that the baseline findings are unlikely to be driven by unobserved heterogeneity between treatment and control counties.

To shed light on the mechanisms and rule out alternative explanations, I show that the reform did not affect political turnover and local leader characteristics, which rules out a *selection* effect; that is, the removal of bad local leaders or the entry of good local leaders (Avis, Ferraz and Finan, 2018). Instead, the results have no substantial change when I exploit only within-leader variation in exposure to the reform, suggesting that the reform worked through a *disciplining* effect; that is, by inducing local leaders to refrain from GDP manipulation in view of future legal or political costs. I further bolster such disciplining effect by drawing on two heterogeneity tests: first, I show a less pronounced effect in regions with poorer institutional quality. In such regions, the survey teams may collude with local leaders, and local leaders may also interfere with the survey teams, leading the survey teams to be dysfunctional; second, I also find that the effect is decreasing in local leaders' career concerns. For local leaders with greater career concerns, the gains from GDP manipulation would outweigh the

[1] As I will describe later, these earlier teams were mainly used to collect information on agricultural output.

cost from potential punishment to a larger extent, hence they would react less to the reform. Having verified such disciplining effect, I next show that the findings are not driven by better soft information acquired by the survey teams, nor by improvement in local statistical capacity, which is measured by the performance of a county in coordinating and conducting the centrally launched economic census. Finally, I show that the findings do not capture the effects of other contemporaneous reforms that may also strengthen the monitoring of local leaders, such as the anti-corruption inspections.

Lastly, I delve into the real impact of the reform. This is motivated by a theoretical framework showing that county leaders in the reformed areas tended to allocate relatively more efforts into stimulating the economy given the relatively higher cost of GDP manipulation after the reform. Along this line, I first show through a textual analysis of county government work reports that government policies shifted in directions conducive to economic growth after the reform. In particular, I find greater emphasis on business attraction and market reform by local governments. As a placebo, however, I find no changes in social welfare. I further link the reform to downstream economic outcomes and find that bank loans and firm entry increased, and more so for smaller firms and firms with higher productivity, consistent with an improvement in resource allocation. In addition, using household survey data, I show that citizens improved their attitudes towards local governments after the reform: they increased trust in local officials and thought higher of local government performance. In contrast, I find no changes in trust in the general population. I conclude the analysis by checking the effect on corruption, using both perceived corruption and anti-corruption convictions. I find no changes in corruption after the reform, suggesting that GDP manipulation is a distinct aspect of bureaucrat misbehaviors.

This paper relates mainly to three strands of literature. Most relatedly, it contributes to the nascent literature documenting GDP manipulation, which mainly happens in autocracies (Wallace, 2016; Xiong, 2018; Martinez, 2021). This strand of literature uses data aggregated at the national or subnational level and shows that authoritarian leaders or career-minded subnational leaders tend to exaggerate GDP. This paper advances the literature in three aspects. First, it leverages data at the county level, which is the lowest level at which GDP is calculated in China. Such disaggregated data allows me to better control for unobserved heterogeneity across provinces or countries. Second, it draws on a unique quasi-experimental design by comparing counties with and without the pre-deployed survey teams before and after the reform, which not only allows me to better establish causality but also provides a rare opportunity to explore how to reduce GDP manipulation. To my best, this study is the first to formally investigate the effectiveness of a national policy on curbing GDP manipulation. Third, the rich micro-level data allows me to delve into the downstream impacts of GDP manipulation, which deepens our understanding of the far-reaching effect of such misbehavior.

This paper also complements the growing literature on the role of top-down monitoring in disciplining local officials. Olken (2007) conducts field experiments in Indonesian to show that top-down monitoring by government auditors can reduce corruption, while grassroot monitoring fails to do so. Avis, Ferraz and Finan (2018) show in the context of Brazil that being randomly audited in the past could decrease future corruption by 8%. They interpret the reduction as mostly explained by disciplining effect stemming from legal actions. Bobonis, Cámara Fuertes and Schwabe (2016) show in Puerto Rico that foreseeable audits could discipline politicians in the short run but not in the long run. Vannutelli (2021) shows in Italy that independent auditors, rather than those appointed by mayors, could improve municipal fiscal performance. However, these existing findings are mostly explained by electoral and judicial accountability. Hence, my findings complement this strand of literature by highlighting the power of top-down monitoring in settings like China with no electoral accountability and weak legal institutions. It also speaks to the effect of top-down monitoring in hierarchical organizations, where lower-level officials or agents are only accountable to those at the higher level.

Finally, this paper adds to the literature estimating causally the real consequences of corruption or bureaucrat misbehaviors in general. In the China setting, Giannetti et al. (2021) show by studying the anti-corruption campaign that corruption stifles firm performance and affects disproportionately small and young firms. Colonnelli and Prem (2022) show in the Brazil setting that after anti-corruption audits, the number of firms operating in government-dependent sectors increased by about 1.4%, total sales increased by about 6%, and bank loans and deposits rose by about 3%. Ajzenman (2021) shows in the context of Mexico that political corruption could erode citizens' values related to honesty, rule observance and trust, pointing to the self-reinforcing effect of corruption. This paper studies a less examined aspect—GDP manipulation, which diverts local leaders' effort away from stimulating the economy, and resonates with the existing literature by showing its effects on government policies and resource allocation (e.g., more credit to small firms). While the focus is on GDP manipulation, the implications can apply to other performance indicators on which local leaders are evaluated but cannot be fully observed by the principal.

The remainder of the paper is organized as follows. Section 2 introduces the institutional background of GDP manipulation and the reform. Section 3 develops a simple theoretical framework to illuminate how the reform could affect GDP manipulation. Section 4 describes the data and derives the estimating equation. Section 5 presents the main results and shows robustness. Section 6 dismantles the mechanisms and rules out several alternative explanations. Section 7 investigates the downstream effects of the reform. Section 8 discusses the difference from the literature on corruption. Section 9 concludes.

2 Institutional Background

In China, GDP is calculated using the value-added method; that is, by summing the value added in all sectors in a region, with the county as the lowest level of regions for GDP calculation. The calculation is done by the local statistics bureau, which is controlled by local leaders in terms of personnel and funding. It is worth noting that a Chinese county is co-led by two leaders: the party secretary controlling personnel and other political affairs, and the magistrate running the economy. However, both leaders are evaluated heavily on GDP growth and hence have incentives to manipulate (Yao and Zhang, 2015). These leaders have a couple of ways to manipulate GDP: directly asking local statistics bureau to make up numbers, requiring firms to overstate income or pay additional "tax" and return later, or double counting firms' non-local subsidiaries, among others.

In terms of the statistics reform, it was initiated by the National Bureau of Statistics (NBS), joint with other central authorities, in May 2009 with the goal of disciplining misbehaviors of local governments in processing statistics data. It mainly targeted local leaders who falsified statistics data by themselves, forced or instructed other agents to manipulate, retaliated against those detecting manipulation, or failed to find severe distortion in local statistics data. The last clause means that local leaders were still punished even if there was no evidence of their direct manipulation, alleviating concerns about the local statistical bureau acting as scapegoats upon detection. In addition, other agents participating in manipulation, such as the staff in local statistics bureau, were also punished. The punishment involves warning, demerit, demotion, or dismiss, depending on severity.

In terms of the detection of misbehaviors, both the local statistics bureau in each county and centrally managed survey teams in some counties, which I will describe their deployment shortly, were responsible. However, the local statistics bureau shared aligned incentives with local leaders because they were appointed and funded by local leaders; in contrast, the survey teams had a higher probability of detecting misbehaviors, as they were appointed and funded centrally. As emphasized in the literature, this type of independence is the key to the effectiveness of monitoring (Olken, 2007; Vannutelli, 2021). Upon detection, the survey teams could directly report to the NBS, and the NBS would take actions together with other authorities. The main takeaway so far is that the survey teams, which only existed in part of the counties, create the key source of variation in exposure to the reform.

What is crucial to my difference-in-differences identification strategy is when and how the survey teams were deployed. They were deployed by the NBS in 2005 in 40% of counties. Their job was to conduct sampling surveys to collect information on CPI, household income, grain output, and micro-firm dynamics. As these variables were frequently published and updated by the NBS, the survey teams could lighten the workload of the generally understaffed local statistics bureau. Starting from 2009 when the aforementioned reform was launched, these teams started to detect misbehaviors in terms of manipulating statistics data. While the NBS did not officially reveal the criteria regarding the selection of counties with these survey teams, by checking local gazetteers I find that most of these survey teams launched in 2005 were restructured from earlier rural survey teams set up in 1984. At that time, these earlier rural survey teams were mainly used to collect information on agricultural output, and counties with them were selected randomly within a province. I leave the detailed discussion of these earlier rural survey teams in Section 5.2.2, where they are used as an instrument. Unless explicitly noted, the survey teams refer to those launched in 2005 in my subsequent analysis.

3 Theoretical framework

In this part, I will leverage a simple economic tournament model to illustrate the sources of GDP manipulation under China's unique promotion rule and generate some testable predictions on the effect of the reform. Similar to the game setting in Lazear and Rosen (1981), I consider a single-period tournament and ignore discounting. There are two county leaders indexed by $i = 1, 2$ competing for promotion, which is decided by the principal (the upper-level government). Leader i can stimulate the economy with effort e_i and cost $g(e_i)$ or simply manipulate GDP growth by a degree of m_i with cost $h(m_i)$.[2] Both $g(\cdot)$ and $h(\cdot)$ are increasing and convex. Manipulating GDP itself is costly because one needs to align potential dissenters or carefully craft various components of GDP. To generate sharp predictions, I assume that each leader would

[2]While the modification of GDP statistics is directly done by local statistics bureaus, I only model the behavior of local leaders as local statistics bureaus are controlled by local leaders and thus act in concert with local leaders.

bear a total cost (stimulating the economy and manipulating GDP) no larger than $\bar{C}$, namely, $g(e_i) + h(m_i) \leq \bar{C}$. Furthermore, I assume that to make the manipulation less detectable, a leader manipulates GDP simultaneously with stimulating the economy, instead of after observing true GDP growth at the end of the period. [3] Hence, the reported GDP growth is given by

$$G_i = e_i + m_i + \varepsilon_i, \quad \varepsilon_i - \varepsilon_{-i} \sim \mathrm{U}[-\frac{1}{2\phi}, \frac{1}{2\phi}]$$

where the reported GDP growth is composed of the effort invested in stimulating the economy e_i, the degree of manipulation m_i, and an idiosyncratic shock ε_i. I assume the difference of the shocks between the two counties is uniformly distributed with mean 0 and density ϕ. Such distribution is known to all, but the realized values of the shocks are only known at the end of the period.

In addition, manipulating GDP would be detected by the survey teams deployed by the principal with probability θp, where $p \in (0, 1)$ denotes the exogenous rate of identifying manipulation associated with well-functioning survey teams, and θ denotes institutional quality, which captures how well the survey teams are functioning. For instance, in counties with poor institutional quality, local officials may interfere with the survey teams, and the survey teams may also collude with the local officials, thereby impeding the detection. Once detected, I assume that the leader suffers from a punishment taking the linear form of λm_i, where λ is a positive number reflecting the marginal punishment of manipulation. I do not assume a convex punishment of manipulation because in China the punishment of local officials' misbehavior is usually not convex or even concave.[4] Then the overall cost for leader i is

$$C_i = g(e_i) + h(m_i) + \mathbb{1}_{\{i \text{ detected}\}} \lambda m_i$$

where $\mathbb{1}_{\{i \text{ detected}\}}$ is an indicator equal to 1 if leader i is detected for manipulation and 0 otherwise.

Leader i's payoff is given by

$$U_i = \delta[\mathbb{1}_{\{i \text{ promoted}\}} u(R) + (1 - \mathbb{1}_{\{i \text{ promoted}\}}) u(r)] + (1 - \delta) u(\Omega) - C_i$$

where $\mathbb{1}_{\{i \text{ promoted}\}}$ is an indicator equal to 1 if leader i is promoted and 0 otherwise.

[3] This is also supported by anecdote evidence showing that local leaders asked firms to overstate income in the middle of a year.

[4] See http://www.gov.cn/xinwen/2016-04/18/content_5065309.htm for the punishment of corruption.

The utility function $u(\cdot)$ is increasing and concave. Leader i receives reward R if promoted and r if not, where $R \gg r > 0$. [5] Ω denotes exogenous rents extracted from her current office. The weight δ attached to future reward thus captures the degree of career concerns, which I will leverage to generate heterogeneous predictions.

Promotion rules I outline two rules that would plausibly characterize the key features of promotion in China, although there are no explicit and written criteria. The first rule posits that the principal mainly promotes the county leader with the highest reported GDP growth. Such feature has a widely-acknowledged theoretical foundation, as the high comparability across subnational units in China makes economic tournaments particularly suitable for promoting regional leaders (Maskin, Qian and Xu, 2000). It is also verified at various levels of governments by a growing literature showing that GDP growth is arguably the overarching performance measure dictating local leaders' career advancement (Li and Zhou, 2005; Jia, Kudamatsu and Seim, 2015; Chen and Kung, 2016). To capture the reform's effect, I assume that, upon detection of manipulation, the principal subtracts the degree of manipulation from a leader's reported GDP growth. Namely, leader i is promoted if

$$G_i - \mathbb{1}_{\{i \text{ detected}\}} m_i > G_{-i} - \mathbb{1}_{\{-i \text{ detected}\}} m_{-i} \qquad \text{(Rule 1)}$$

The second promotion rule shares the key feature with the first rule, except that, instead of subtracting the degree of manipulation from the reported GDP growth, the principal disqualifies a candidate for promotion upon detection of manipulation.[6] Specifically, this rule has the following features: (1) leader i is automatically promoted if undetected for manipulation but her opponent $-i$ is detected; (2) leader i is disqualified for promotion if detected for manipulation, irrespective of her opponent $-i$'s behavior; (3) if both are undetected, then leader i is promoted if $G_i > G_{-i}$; (4) if both leaders are detected for manipulation, then neither would be promoted. In short, leader i is promoted if

$$\{i \text{ undetected}, -i \text{ detected}\} \cup \{i \text{ undetected}, -i \text{ undetected}, G_i > G_{-i}\} \qquad \text{(Rule 2)}$$

Timing The timing of events in this tournament is summarized as follows:

[5] The positive reward r captures the fact that in China most local leaders would still stay in the same or similar position even if not promoted.

[6] This is realistic as the reform explicitly says that local leaders can be demoted or even dismissed depending on the severity of manipulation.

1. Both leaders simultaneously choose effort in stimulating the economy and the degree of GDP manipulation, before knowing the realization of ε_i.
2. ε_i is realized and all uncertainty is resolved.
3. The principal detects manipulation, punishes the involved, and makes promotion decision based on one of the aforementioned rules.

Equilibrium The equilibrium concept is a pure strategy Nash equilibrium.[7] To solve it, note that leader i maximizes her expected payoff taking leader $-i$'s choice as given. Then through usual maximization, one can get the equilibrium GDP manipulation m^* and effort in stimulating the economy e^*:[8]

$$\begin{aligned} m^* &= \mathbb{1}_{\{\text{Rule 1}\}} K\left[\frac{\theta p(\lambda+\delta V\phi)}{\delta V\phi} - 1\right] + \mathbb{1}_{\{\text{Rule 2}\}} K\left[\frac{\lambda\theta p}{(1-p)^2\delta V\phi} - 1\right] \\ e^* &= k(m^*) \end{aligned}$$

where $\mathbb{1}_{\{\text{Rule 1}\}}$ and $\mathbb{1}_{\{\text{Rule 2}\}}$ are indicators for the first and second promotion rules, respectively. $K(\cdot)$ is the inverse function of $k'(\cdot)$ and $k(\cdot) = g^{-1}(\bar{C} - h(\cdot))$. $V = u(R) - u(r)$, which measures the utility gains from promotion.

It is trivial to see $K'(\cdot) < 0$, then under both promotion rules, equilibrium GDP manipulation m^* is decreasing in institutional quality θ, probability of detection p, and marginal punishment λ. In contrast, m^* is increasing in promotion gains V, career concerns δ, and shock density ϕ.[9] Furthermore, equilibrium GDP manipulation is smaller under the second promotion rule, where a leader is disqualified for promotion upon detection of manipulation. In contrast, equilibrium effort in stimulating the economy e^* changes in opposite directions with respect to the aforementioned parameters, as e^* is a decreasing function of m^*.

To shed light on the reform's effect, let T and C denote treatment (counties with the survey teams) and control counties (counties without the survey teams), respectively. Then one could generate some testable predictions on the effect of the reform on GDP manipulation, effort in stimulating the economy, true economic growth, and the heterogeneity of the treatment effect, which are summarized below.

[7] I focus on pure strategies as it is empirically obscure to interpret mixed strategies in stimulating the economy and manipulating GDP in a static game.

[8] The assumptions on the functional forms ensure that the maximum exists.

[9] A higher density means that the leader can rely less on luck, so she needs to manipulate more.

Prediction 1. *Treatment counties should see a larger decrease in GDP manipulation than control counties:*

$$\Delta m_T - \Delta m_c = \underbrace{(m'_T - m_T)}_{<0} - \underbrace{(m'_C - m_C)}_{\approx 0} < 0$$

Intuitively, in treatment counties, the survey teams could effectively detect GDP manipulation, so the cost of GDP manipulation increases after the reform. Consequently, GDP manipulation decreases, consistent with the reform generating a disciplining effect on local leaders in treatment counties. In control counties, without the survey teams, the detection of GDP manipulation relies solely on local statistics bureaus. As local statistics bureaus share aligned incentives with local leaders and are also controlled by local leaders in terms of personnel and funding, they are essentially dysfunctional in terms of detection. Hence, there would be little change in GDP manipulation after the reform.

Prediction 2. *Treatment counties should see a larger increase in effort in stimulating the economy than control counties:*

$$\Delta e_T - \Delta e_c = \underbrace{(e'_T - e_T)}_{>0} - \underbrace{(e'_C - e_C)}_{\approx 0} > 0$$

Intuitively, in treatment counties, it would be relatively less costly to invest effort in stimulating the economy because of the heightened cost of GDP manipulation. In contrast, in control counties, the cost of GDP manipulation has little change after the reform due to the lack of effective detection, as discussed in 1.

Prediction 3. *Let $y = e + \varepsilon$ denote true economic growth. The effect of the reform on true economic growth is indeterminate:*

$$\Delta y_T - \Delta y_c = \underbrace{(\Delta e_T - \Delta e_c)}_{>0} + \underbrace{(\Delta \varepsilon_T - \Delta \varepsilon_C)}_{\gtreqless 0} \gtreqless 0$$

This is because true economic growth is composed of both the effort invested in stimulating the economy and the realized values of the idiosyncratic shocks. Depending on the relative strength of the changes in these two forces, the effect of the reform on true economic growth could be positive, zero, or negative.

Prediction 4. *The magnitude of the treatment effect on GDP manipulation $|\Delta m_T - \Delta m_c|$ is decreasing in career concern δ but increasing in institutional quality θ:*

$$\frac{\partial |\Delta m_T - \Delta m_c|}{\partial \delta} < 0, \quad \frac{\partial |\Delta m_T - \Delta m_c|}{\partial \theta} > 0$$

This prediction concerns the heterogeneity of the treatment effect. The first heterogeneity is because local leaders with greater career concerns would attach a larger weight to promotion. As such, they manipulate more to boost their promotion chances. The second heterogeneity is intuitive as poorer institutional quality could undermine the functioning of the survey teams, leading to a smaller effect of the reform.[10]

4 Data and empirical strategy

4.1 Main data

Below I list the main data that will be used in this paper. Other data will be described where they first appear in the paper.

The list of counties with survey teams The list of counties with survey teams deployed in 2005 is collected from the annual reports published by various levels of statistical bureaus, and supplemented by local gazetteers.

County-level outcomes and covariates County-level data on GDP and other statistics are collected from county statistics yearbooks. County-level data on harmonized nighttime light intensity are collected from Li et al. (2020). These data will be used to construct GDP growth manipulation. Other county data on demographic, economic, and geographic characteristics, which are used to conduct balance tests, are collected from multiple sources including the 2010 population census, the National Oceanic and Atmospheric Administration (NOAA), and the United States Geological Survey (USGS).

Household surveys Household survey data, such as the China Family Panel Studies, is collected from the Institute of Social Science Survey (ISSS) of Peking University and used to construct citizens' trust in local officials and other variables.

Local leader résumés Local leader résumés are collected from various government websites, Baidu Baike (China's equivalent of Wikipedia), and occasionally complemented by online news reports. They are used to construct variables on leader characteristics.

My main analysis will focus on the period 2005-2018, for which I have detailed

[10]Note that one could also show similar heterogeneity prediction for efforts in stimulating the economy, but I leave that to future work as my current data on this part is incomplete and I could be underpowered to find heterogeneity empirically.

information on various variables. I define treatment counties as those with the surveys teams deployed in 2005 and control counties as those without, excluding the following types: (1) counties in the four centrally-managed cities: Beijing, Shanghai, Tianjin, and Chongqing. These counties are ranked one level higher than other counties and hence are not suitable to serve as controls; (2) counties in Tibet where data are unavailable; (3) counties outside mainland China;[11] (4) urban districts. In the end, I have 1779 counties in total, of which 40% are treated. The spatial distribution of treatment counties can be found in Figure 1, which is quite even across space.

4.2 Empirical strategy

Deriving estimation equation To derive the estimation equation, I incorporate GDP growth manipulation and the reform into the framework of Henderson, Storeygard and Weil (2012), who establish the relationship between nighttime light intensity and real economic activities. First, denote the reported GDP growth (with manipulation), true GDP growth (unobservable), and nighttime light intensity growth in county c and year t as z_{ct}, y_{ct}, and l_{ct}, respectively. Assume the degree of manipulation is m_{ct}. Then the GDP growth observed by the local statistics bureau is $z_{ct} - m_{ct}$ (without manipulation). According to Henderson, Storeygard and Weil (2012), the mapping from $z_{ct} - m_{ct}$ to y_{ct}, and the mapping from l_{ct} to y_{ct} can be written respectively as:

$$z_{ct} - m_{ct} = y_{ct} + \epsilon^{z}_{ct} \tag{1}$$

$$l_{ct} = \gamma y_{ct} + \epsilon^{l}_{ct} \tag{2}$$

Combining equations (1) and (2), the degree of manipulation can be written as:

$$m_{ct} = z_{ct} - \frac{1}{\gamma} l_{ct} + \epsilon^{m}_{ct} \tag{3}$$

where ϵ^{m}_{ct} is a combination of the error terms ϵ^{z}_{ct} and ϵ^{l}_{ct}. Then the difference-in-differences equation to test the effects of the reform on GDP growth manipulation can be written as:

$$m_{ct} = \beta Treat_c \times Post_t + \delta_c + \lambda_t + \varepsilon^{m}_{ct} \tag{4}$$

where $Treat_c$ and $Post_t$ are dummy variables for treatment counties (the 40% afore-

[11] Specifically, these include counties in Hongkong, Macau, and Taiwan. They are excluded due to institutional and administrative differences from mainland China.

mentioned counties with survey teams deployed in 2005) and post-reform years (years after 2009), respectively. As one cannot directly observe m_{ct}, substituting equation (3) into equation (4) and rearranging gets:

$$\underbrace{z_{ct}}_{\text{ReportedGDPGrowth}} = \frac{1}{\gamma} \underbrace{l_{ct}}_{\text{LightGrowth}} + \beta Treat_c \times Post_t + \delta_c + \lambda_t + \varepsilon^z_{ct} \tag{5}$$

where ε^z_{ct} is a combination of ϵ^m_{ct} and ε^m_{ct}. δ_c denotes county fixed effects, controlling for time-invariant factors at the county level that may correlate with the treatment or the outcome; λ_t denotes year fixed effects, controlling for time-varying shocks common to all counties. As the treatment varies at the county level, I cluster the standard errors by county (Abadie et al., 2017), and check robustness to alternative levels of clustering. I expect the coefficient of interest β to be negative, which implies that counties with the survey teams would engage in less manipulation relative to other counties after the reform. Note that both local statistics bureaus and the survey teams could engage in detecting manipulation after the reform, which may affect the interpretation of β. Under the assumption that local statistics bureaus are dysfunctional in detection, which is plausible as they are controlled by local leaders, the coefficient β could be well interpreted as the overall effect of the reform. If this assumption is not true, the coefficient β is a lower bound of the effect of the reform, but this is still meaningful.

Identification concerns The identification assumption is that, reported GDP growth, after adjusting for light growth, should evolve in parallel between treatment and control counties in the absence of the reform in 2009. This assumption is essentially unverifiable. Pre-reform parallel trends between treatment and control counties, which is commonly estimated using event study specifications, can lend support to such assumption but cannot fully verify it. One still needs to consider two types of concerns: first, the relationship between light and economic activities may differ across counties or years, which is specific to my setting; second, treatment counties may differ significantly from control counties ex ante. To address the first concern, I will allow the effect of light to vary by a host of spatial and temporal characteristics to check the sensitivity of the estimates. The second concern about unbalance between treatment and control counties ex ante is conventional in difference-in-differences designs. Although perfect balance in levels is not necessarily required in such designs, significant unbalance may raise concerns about the validity of using the control groups as counterfactual. As shown in Appendix Table B1, treatment counties were quite similar to control counties ex ante, except for population and GDP (but, importantly, not GDP growth or light growth). The pre-deployment nature of the survey teams implies that such unbalance should not be related to the reform strategically and therefore should be largely orthogonal to the outcomes. In the specification checks, I will demonstrate the robustness of the estimates to flexibly controlling for size and other baseline covariates, to allowing for county-specific trends, and to achieving covariate balance through entropy balancing (Hainmueller, 2012), among others. I will also leverage institutional knowledge to design two additional tests: the first is a placebo test around 2005 when the survey teams were launched but had not conducted any disciplining actions, and the second is an instrumental variable approach using the random assignment of earlier rural survey teams as an instrument, which I will elaborate later on.

5 Results

5.1 Main results

Event study Figure 2 shows the dynamic effect of the reform estimated using equation (6), which is an event study variant of the baseline equation (5) with the post-reform dummy replaced by a set of year dummies. The year before the reform, 2008, is omitted as the reference year. The estimated effects in the pre-treatment period, namely, all β_js for $j < 2009$, are essentially indistinguishable from zero and statistically insignificant. A F-test of joint significance of all the pre-treatment estimates generates a p-value of 0.96, implying that the parallel trends assumption is plausibly satisfied. After the reform, there is an immediate and persistent negative effect, suggesting that the reform decreased GDP growth manipulation, which is consistent with Prediction 1. In Appendix Figure A1, I further show a decomposition of the effect of the reform, by checking the dynamic effect on reported GDP growth and light growth separately. [12] The results further confirm that the reform decreased GDP growth manipulation: there is a sharp drop in reported GDP growth but little change in light growth after the reform. The latter is consistent with Prediction 2 that the effect of the reform on true economic growth is indeterminate.

[12] Specifically, I estimate the following equation, where Y_{ct} denotes either reported GDP growth or light growth:

$$Y_{ct} = \sum_{j=2005,\, j\neq 2008}^{j=2018} \beta_j Treat_c \times 1_{\{t=j\}} + \delta_c + \lambda_t + \varepsilon_{ct}$$

$$ReportedGDPGrowth_{ct} = \alpha LightGrowth_{ct} + \sum_{j=2005,\ j\neq 2008}^{j=2018} \beta_j Treat_c \times 1_{\{t=j\}} + \delta_c + \lambda_t + \varepsilon_{ct} \tag{6}$$

Average effect Table 1 summarizes the dynamic treatment effects above into an average treatment effect. Column (1) reports the results using equation (1), controlling for only county and year fixed effects, and light growth. The point estimate on Treat x Post is negative and statistically significant (coef.=−0.751, s.e.=0.023), implying that relative to the control counties, treatment counties experienced a 0.751 percentage points drop in GDP growth manipulation after the reform. The estimate on light growth is also consistent with that in the literature (Martinez, 2021).[13] Through columns (2)-(4), I gradually introduce a set of demographic, economic, and geographic controls (interacted with the post-reform dummy), which are used in the balance tests. The precision of the estimates improves, although the size drops slightly. In the most stringent specification in column (4) with all the county controls, the estimate shows a 0.576 percentage points drop in GDP growth manipulation in treatment counties relative to the control counties after the reform. This drop is also economically significant, which amounts to 5.3% of the mean or 6.9% of the standard deviation of the reported GDP growth. In sum, these findings suggest the effectiveness of the combination of monitoring and punishment in reducing bureaucratic misbehaviors in data processing, which resonates with the key insights of Becker and Stigler (1974).

Spillover Having established the negative effect of the reform on GDP growth manipulation, I turn to check if there exists any spillover effect, which could bias my baseline estimation even if the treatment is exogenous. To this end, I estimate equation (7), where $TreatedNeighbors_c$ denotes the strength of spillover and $\beta^{Spillover}$ captures the spillover effect of the reform. Following Huber (2023), I use the number of treatment counties among a county's neighbors to proxy for the strength of spillover to that county, where neighbors are defined as other counties sharing a common boundary segment with that county. The results are reported in Appendix Table B3. In column (2), the estimated spillover effect is small and statistically insignificant (coef.=−0.01, s.e.=0.071). Considering the average number of treated neighbors for a county is 2,

[13]Specifically, in a similar specification in Martinez (2021)'s cross-country analysis, the coefficient estimate on light growth is about 0.027-0.039 (s.e.=0.006-0.007), and in my setting, it is about 0.017-0.023 (s.e.=0.005).

such estimates imply that moving from a county with no treated neighbors to the average county would decrease GDP growth manipulation by 0.02 percentage points. Given the direct effect of about 0.58 percentage points, the spillover effect is also economically negligible. In contrast, the direct effect is barely changed compared to the baseline effect as reported in column (1). In the remaining two columns, I use dummies to indicate the strength of spillover, and the results have no substantial changes. In Appendix Table B4, I further show that the results are robust to using alternative definitions of neighbors or weighting the treated neighbors by their sizes.

$$\begin{aligned} ReportedGDPGrowth_{ct} = \alpha LightGrowth_{ct} + \ & \beta^{Direct} Treat_c \times Post_t \\ & + \beta^{Spillover} TreatedNeighbors_c \times Post_t + \delta_c + \lambda_t + \varepsilon_{ct} \end{aligned} \tag{7}$$

5.2 Robustness checks

5.2.1 Alternative specifications

In this section, I will show the robustness of the baseline results to a host of alternative specifications that alleviate the aforementioned identification concerns. The first concern is about heterogeneous mapping between nighttime light intensity and true economic activities. As previously discussed, the baseline equation (5) assumes a uniform and linear relationship between nighttime light intensity and true economic activities. However, such relationship may be nonlinear and change across counties or over years. To alleviate such concerns, I allow the effect of light to: (1) be nonlinear by including a 3rd-order polynomial of light; (2) vary by county longitude and latitude; (3) vary by county area; (4) vary by GDP (5) vary by population (6) vary by urbanization rate; (7) vary by economic structure (proxied by share of population in the primary and secondary sectors); (8) vary by year; (9) vary by province; (10) vary by both province and year; (11) vary by treatment status; (12) vary by treatment status and linearly by year. As shown in Panel A of Figure 3, the results are essentially unaffected by these alternative specifications, which suggests that my baseline findings are not an artifact of heterogeneous light effect.

Another concern is about covariate unbalance. As shown in Appendix Table B1, treatment counties were larger than control counties ex ante and may therefore differ significantly from control counties later on, leading to violations of the parallel trends assumption. To alleviate such concerns, I estimate the following alternative specifications: (1) I flexibly control for size effects by including county size decile bin fixed effects interacted with year fixed effects, where county size is proxied by preexisting GDP, population, and area; (2) I add county-specific time trends that allow treatment and control counties to be on differential linear trajectories (Angrist and Pischke, 2014). This could relax the identification assumption, although the precision of the estimates may decrease;[14] (3) I add province-by-year fixed effects. In this way, I am only comparing counties in the same province and year, and the covariates should be more balanced; (4) I include all the county covariates interacted with year fixed effects to allow treatment and control counties to trend differentially depending on the covariates; (5) I select the most relevant covariates using the Double LASSO method (Belloni, Chernozhukov and Hansen, 2014); (6) I re-weight observations to make treatment and control counties similar in terms of observables using the entropy balancing method (Hainmueller, 2012). These results are plotted in Panel B of Figure 3 and are quite similar to the baseline estimates from equation (5), implying that my baseline estimates are not driven by possible differential trends caused by covariate unbalance.

Finally, Appendix Figure A2 shows that the results are not driven by certain particular regions, by leaving out each province individually. Appendix Table B5 shows that the results are not driven by a few smaller counties by weighting the regression by county size (e.g., population or GDP). Appendix Table B6 shows that the results are robust to alternative levels of clustering, such as by prefecture, province, or longitude-latitude grid cells. Appendix Figure A3 further shows that the results are robust to randomization inference, which could be more reliable in difference-in-differences settings with few treatment groups (MacKinnon and Webb, 2020).

5.2.2 Placebo reform: the launch of the survey teams in 2005

I corroborate the baseline results by conducting a placebo event study around 2005 when the survey teams were launched but had not conducted any disciplining actions. This could further help to see if there existed any pre-trends. In addition, it could alleviate concerns that the survey teams per se may affect the outcomes, even without the reform in 2009. The specification is equation (8), which is similar to the baseline event study specification in equation (6) except that the sample period here is from 2001 to 2008, with the year 2004 omitted as the reference year. Figure 4 shows the event study estimates. In contrast to the sharp drop in reported GDP growth around 2009 in Figure 2, there was no discernible change in reported GDP growth around 2005. This pattern persisted until 2009. Furthermore, there were no diverging trends before 2005. Such results suggest that the reform effect is unlikely to be driven by preexisting differences between treatment and control counties or differential effects

[14] Using linear time trends in DiD specifications could absorb part of the effect and the treatment variation, which leads to less precise estimates (Goodman-Bacon, 2021).

(net of the reform effect) generated by the survey teams.

$$
\begin{aligned}
ReportedGDPGrowth_{ct} = \alpha LightGrowth_{ct} + \sum_{j=2001,\ j\neq 2004}^{j=2008} \beta_j Treat_c \times 1_{\{t=j\}} \\
+ \delta_c + \lambda_t + \varepsilon_{ct}
\end{aligned} \tag{8}
$$

5.2.3 Instrumental variable strategy

While the baseline difference-in-differences strategy is immune to any time-invariant confounders, one may still worry that treatment counties may differ from other counties along certain characteristics that would vary over time, which could bias the baseline estimates. To alleviate this concern, in this section I leverage a unique institutional feature to construct an instrumental variable for the treatment and conduct an instrumented difference-in-differences estimation.

In 1984, to gauge agricultural production, the National Bureau of Statistics set up a group of teams called *rural survey teams* in part of the counties. At that time, China was still essentially an agricultural country. The counties with these rural survey teams were chosen randomly within a province. In particular, the NBS adopted a commonly used probability sampling method called *systematic random sampling*: one first selects a random start and then selects counties with a fixed and periodic interval. This sampling rule and the list of counties selected are collected from provincial gazetteers.[15] In theory, each county should have the same probability of being selected, leading to perfect within-province randomness of assignment of counties with the rural survey teams. Of course, in practice, the randomness may be affected by particular patterns in the county sequence or the limited number of counties in some provinces. In terms of specific work, these rural survey teams were led by the NBS, but in terms of personnel and funding, they were led by local leaders. Given the dramatic change in economic structure caused by market reform in recent years, in 2005, these rural survey teams were abolished, and most of them were restructured into more comprehensive and independent survey teams (led solely by the NBS), which are the survey teams examined in the previous parts.

I define a dummy variable $Treat_c^{1984}$ equal to 1 if a county had a rural survey team in 1984. Given the previous discussion, I expect this instrument to be strongly correlated with $Treat_c$ and also satisfy the exclusion restriction due to the within-province randomness of assignment. Figure 5 shows the distribution of the rural survey teams

[15]Gazetteers are called *Difangzhi* in Chinese and are a series of encyclopedias covering a wide range of topics: history, geography, economics, politics, culture, social sciences, etc. They are compiled by local officials and noted literati in each area and updated every dozens of years.

in 1984. The significant overlap with the survey teams launched in 2005 (see Figure 1) suggests the high relevance of this instrument, which is also confirmed by the first-stage F-statistic that I will show shortly. The exclusion restriction requires that this instrument does not affect other dimensions of the counties except for having a survey team in 2005. To check whether this instrument is correlated with other county characteristics, Appendix Table B2 provides a balance test for this instrument. Counties with and without the rural survey teams in 1984 are well balanced across a number of demographic, economic, and geographic factors. Appendix Figure A4 further plots the distribution of the standardized differences between the two groups of counties for all these factors, which are well centered around zero, suggesting no systematic patterns between the two groups. The absolute values of the standardized differences never exceed 7%, which is far below the threshold of 25% as suggested by Imbens and Rubin (2015) for covariate balance. Finally, one may be worried about certain legacy effect generated by the rural survey teams. Specifically, these teams per se may affect the outcomes examined in this paper, namely, GDP manipulation later in the 2000s. However, this is unlikely as the rural survey teams did not conduct any disciplining actions and were essentially led by local governments. Furthermore, even if the rural survey teams generated some legacy effect, as long as it did not change after 2009, which seems very plausible, such legacy effect will be differenced out by my DID strategy.

Given the relevance and the exogeneity of the instrument, I perform 2SLS estimation where the first- and second-stage equations are:

$$Treat_c \times Post_t = \theta LightGrowth_{ct} + \alpha Treat_c^{1984} \times Post_t + \delta_c + \lambda_t + \epsilon_{ct} \tag{9}$$

$$ReportedGDPGrowth_{ct} = \vartheta LightGrowth_{ct} + \beta \widehat{Treat_c \times Post_t} + \delta_c + \lambda_t + \varepsilon_{ct} \tag{10}$$

where all variables are as defined previously, and the coefficient β captures the average causal effect of the reform, among those counties with a survey team in 2005 due to having a rural survey team in 1984. I first conduct an instrumented event study and plot the coefficient estimates in Figure 6. The patterns are quite similar to those shown in Figure 2, albeit with less precision. To summarize the effects, in Table 2 I report the 2SLS results based on equation (10). The F-statistic on the excluded instrument in the first stage ranges from 1945 to 2210, which is much larger than the rule-of-thumb cutoff value of 10 for weak instrument (Staiger and Stock, 1997). The large size of the F-statistic also alleviates concerns about distortions in the conventional inference procedure in 2SLS as raised by the recent econometric literature (e.g. Lee et al., 2022).[16] The estimated treatment effects are slightly smaller than OLS estimates, but are still economically sizable.[17] Considering the results in the last column with all county controls interacted post-reform dummy, the estimate shows a 0.487 percentage points drop in GDP growth manipulation in treatment counties relative to the control counties after the reform in 2009 (coef.=-0.487, s.e.=0.225), which accounts for 4.5% of the mean or 5.8% of the standard deviation of the reported GDP growth. Such results show that the baseline findings are unlikely to be driven by unobserved heterogeneity between treatment and control counties.

6 Interpretation

Having shown the robustness of the baseline findings, in this section I zoom in on the underlying mechanisms. Specifically, I show that the baseline findings are mostly consistent with the reform generating a disciplining effect on local leaders. I also rule out several alternative explanations that may also generate similar patterns, namely, better political selection, better soft information, improvement in local statistical capacity, and contemporaneous reforms that may also strengthen the monitoring of local leaders.

6.1 Disciplining effect as the key mechanism

As posited by the theoretical model, the reform could generate a *disciplining* effect on local leaders: namely, local leaders would refrain from GDP manipulation in view of future legal or political costs. To bolster such a mechanism, I focus on within-leader variations in exposure to the reform. This means that any reduction in GDP manipulation could only be attributed to behavioral changes within the same leader's term, rather than differential propensities to manipulate across leaders. To achieve this, I incorporate leader fixed effects in column (2) of Table 3. Considering that the term length of county leaders is typically 4 or 5 years, this is a very demanding specification as the source of variation now solely comes from county leaders whose terms straddled 2009. The within-leader estimates are similar to the baseline estimates in column (1), albeit with less precision. In column (3), I conduct a similar exercise by

[16] Lee et al. (2022) argue that inference relying on the first-stage F-statistic exceeding a certain cutoff may be biased and propose adjustment factors to inflate the conventional standard errors. However, given the large size of the F-statistic in my setting, the adjustment factor is close to 1 according to Lee et al. (2022), implying that the conventional standard errors are still valid in my setting.

[17] The change in the effect size could be driven by omitted variable bias in OLS or that the 2SLS estimates only capture the effect among compliers (i.e., those counties having a survey team in 2005 due to having a rural survey team in 1984).

focusing on a trimmed sample where only leaders whose terms straddled 2009 are included, thus ruling out any changes caused by leader turnover. The results are also similar to the baseline results in column (1).

Another way to underpin the mechanism is to check the heterogeneity of the disciplining effect as elaborated in Prediction 4. The first heterogeneity concerns how well the survey teams functioned, which is captured by institutional quality. To elucidate, consider the teams working in regions with poorer institutional quality. Then the teams may collude with local leaders or be intervened by local leaders, thereby becoming less effective. Hence, one should expect a smaller effect in areas with poorer institutional quality. To test such heterogeneity, I proxy for a region's institutional quality using the provincial marketization index, which is initially developed by Fan, Wang and Zhu (2003) and used widely as an indicator for institutional quality in China (Li et al., 2011; Fan, Wong and Zhang, 2013; Qian, Strahan and Yang, 2015).[18] Another heterogeneity pertains to the weight that local leaders attach to promotion gains, or career concerns. Specifically, when deciding the optimal amount of GDP manipulation, a leader had to weigh promotion gains against potential costs. Thus, greater career concerns could cause a leader to tilt more towards GDP manipulation as it could boost her promotion probability, thus resulting in a smaller reform effect. I proxy for career concerns using a leader's age, which is an inverse indicator of career concerns because the upper-level government tends to promote younger leaders (Jia, Kudamatsu and Seim, 2015; Persson and Zhuravskaya, 2016; Martinez-Bravo et al., 2022).[19] As shown in Figure 7, the reform effect was smaller in counties with poorer institutional quality and counties where local leaders had greater career concerns, which is consistent with Prediction 4. Hence, I find robust evidence supporting the disciplining effect as the underlying mechanism.

6.2 Political selection

The survey teams may lead to a better political selection and subsequently reduce GDP manipulation, which is akin to the so-called *selection* effect generated by random audits (Avis, Ferraz and Finan, 2018). This may occur if some leaders were replaced due to their misbehaviors being detected, or if the reform facilitated the entry of better leaders. In either case, one should expect to observe an improvement in leader quality in the wake of the reform. I test this explanation in Appendix Table B7. In column (1), I show no effect of the reform on leader turnover (coef.=-0.001, s.e.=0.006). In the

[18] The index is available annually and I take the average of the index in the pre-reform period.

[19] I focus on the ages of party secretaries as they are more powerful than magistrates, although the results using magistrates are similar.

remaining columns, I further show no effect on a battery of observable leader characteristics.[20] While such observable characteristics are just crude proxies for leader quality, many of the estimated effects are economically small and precisely estimated, which rules out even relatively small changes in leader characteristics. Hence, such findings are largely inconsistent with a better political selection channel. Alternatively, if the drop in GDP manipulation is driven by better political selection, then one could expect a larger drop in counties with higher political turnovers. However, this is also not the case as shown in Appendix Table B8.

6.3 Soft information

The survey teams may assist the upper-level government in achieving soft information about the performance of local leaders, thereby dampening the role of GDP growth in promotion (Hart, 1995; Aghion and Tirole, 1997; Stein, 2002). Consequently, local leaders may be less inclined to manipulate in the aftermath of the reform. To explore this possibility, I examine whether the effect is less pronounced in counties closer to the upper-level government or in counties where the leaders are socially connected to the upper-level government, as shorter distances or social ties could also facilitate the flow of soft information and consequently weaken the impact of the survey teams (Agarwal and Hauswald, 2010; Petersen and Rajan, 2002; Bandiera, Barankay and Rasul, 2009; Fisman, Paravisini and Vig, 2017). I follow the literature to proxy social ties using shared hometown background between county leaders and leaders in the upper-level government (Chu et al., 2021; Do, Nguyen and Tran, 2017). As shown in Appendix Table B9, I do not find a significant differential effect in such counties, suggesting that soft information is unlikely to be a driving force of the baseline results.

6.4 Local statistical capacity

Local statistics bureaus may improve their statistical capacity through interactions with the survey teams, and hence could more accurately measure economic activity (Martinez, 2021). This may also lead to a drop in the gap between reported GDP growth and light growth. Given the difficulty in directly measuring the statistical capacity of a county, I utilize an award from the National Bureau of Statistics (NBS) for counties' outstanding performance in coordinating and conducting economic censuses, which are initiated every 4 or 5 years by the NBS. The award is only reputational

[20] I focus on characteristics of party secretaries as they are more powerful than magistrates, although the results using magistrates are similar.

and involves no material rewards. Data on recipient counties of such award is collected from the NBS and is available for the years 2004, 2008, 2013, and 2018. I estimate the following event study specification:

$$Y_{ct} = \sum_{k=2004,2013,2018,k\neq 2008} \beta_k Treat_c \times 1_{\{t=k\}} + \delta_c + \lambda_g + \varepsilon_{ct} \tag{11}$$

where Y_{ct} is a dummy variable equal to 1 if county c wins such an award in year t. The year 2008 is omitted as the reference group. In Appendix Figure A5, I show that the estimates for β_ks are close to zero both before and after the reform, suggesting that the baseline results are unlikely driven by improvement in local statistical capacity.

6.5 Contemporaneous reforms

The baseline results may also be confounded by some contemporaneous reforms that may also strengthen the monitoring of local leaders. The inclusion of province-by-year fixed effects in the robustness checks could rule out all confounding reforms at the province level. In this section, I examine two noteworthy reforms at the county level. The first reform is the province-managing-county (PMC) reform since 2003 (Li, Lu and Wang, 2016). This PMC reform stipulated that the provincial government could bypass the prefecture government and directly administer the county government in fiscal matters (Fiscal PMC), or even in all aspects for a few counties (Full PMC). This data is collected from various government websites. The second reform pertains to the top-down inspections during the recent anti-corruption campaign launched by President Xi Jinping. These inspections mainly focused on curbing corruptions but may also create disciplining effects in other aspects. This data is complied by Wang (2021), who collects the detailed timings and sites of inspections from government websites and newspapers. Appendix Figure A6 presents the rollout of these reforms. Appendix Table B10 shows that the results have no substantial changes after accounting for these reforms.

7 Real effects

As shown in Prediction 2, the reform is expected to generate positive effect on local leaders' effort in stimulating the economy, which is disproportionately larger in treatment counties. In this section, I test this prediction by checking the effect on four aspects: local government policies, financial development, firm entry, and citizen attitudes towards local governments. These aspects individually may not confirm the effect on local leaders' effort, but collectively they would provide compelling evidence.

7.1 Government policies

The reform may shift local government policies in directions conducive to the economy. To test this hypothesis, I conduct a simple textual analysis on county governments' annual work reports.[21] The reports are delivered as a speech by local governments to the People's Congress at the same level, which is held at the beginning of each year. Each report contains two parts: a summary of the government's achievement in the last year and a work plan for the year ahead, which contains detailed and well-structured development policies. The emphasis on each policy area could vary significantly both cross-sectionally and temporarily, as the reports are essentially at the discretion of local leaders. Hence, the reports are well-suited to examine local government policy changes (Jiang, Meng and Zhang, 2019; Campante, Chor and Li, 2022).[22]

I create an original dataset on county-level government reports collected from the official websites of each county. As a first step, I randomly select three provinces:[23] *Guangdong*, *Shaanxi*, and *Zhejiang*, and focus on the years 2007, 2008, 2009, 2013, and 2017.[24] This corresponds to 97 counties. Given such a small sample, the current results should be interpreted with caution. I then define four policy topics: *business attraction*, *infrastructure*, *market reform*, and *policy experimentation*, which are major contributory factors to China's recent economic success and are also frequently mentioned in the reports (Xu, 2011; Jiang, Meng and Zhang, 2019). As a placebo, I create keywords for a fifth topic: *social welfare*, as it is hard to think of a plausible way through which the reform could affect social welfare, at least in the short run. For each topic, I define a list of keywords and count the total number of mentions of these keywords in each

[21] Another way to test this is to check the changes in various components of county fiscal expenditure (e.g., infrastructure versus administrative expenditure), which is from the China Prefecture, City, and County Public Finance Statistics published by the Ministry of Finance of China (Jia, Liang and Ma, 2021). However, the publication of such data stopped after 2007.

[22] Jiang, Meng and Zhang (2019) use prefecture-level government work reports and unsupervised Latent Dirichlet Allocation (LDA) topic models to examine social welfare policies at the prefecture level. Campante, Chor and Li (2022) also use prefecture-level government work reports and a dictionary approach to measure governments' emphasis on political stability.

[23] Unlike the prefecture-level government websites, the county government websites have no uniform layouts, and much of the collection has to be done manually, which is a laborious process. In the future, all provinces will be included.

[24] To avoid confusion, here the years are labeled in a manner that a work report in year t contains the achievements of the government in the same year and its plans for year $t+1$, although such reports are released at the beginning of year $t+1$.

report.[25] The list of keywords for each topic can be found in Appendix Table B11. To control for the length of each report, I normalize the keywords counts by the total number of sentences in each report.

I employ the baseline DiD specification augmented with county controls and present the results in Table 4. While there is no effect on infrastructure and policy experimentation, there is a significant increase in business attraction (coef.=0.008, s.e.=0.004) and market reform (coef.=0.019, s.e.=0.005).[26] Relative to the means, the estimates indicate that, in response to the reform, local governments increased their emphasis on business attraction and market reform by 35% and 30%, respectively. I also find that local governments did not change their emphasis on social welfare, which partially alleviates the concern that the previous findings are caused by a general shift in government policies. Figure 8 reports the corresponding event study graphs, which show little difference between treatment and control counties in these policy outcomes before 2009, and a persistent increase in business attraction and market reform after 2009. Overall, these results suggest a shift of government policies in directions conducive to economic growth. One remaining concern is that these policy shifts may only reflect local governments' visions instead of any real actions. I will address this concern in the next sections by showing the effect on banking, firm entry, and citizen attitudes towards local governments.

7.2 Financial development

To provide complementary evidence underpinning government policy changes, I check the effect on the banking sector to see if deposits and loans expanded after the reform. I collect disaggregated data on deposits and loans at the county level from the China Banking Regulatory Commission. One drawback of such data is that it only covers the period 2006 to 2011, which limits the study of long-term effect. Nevertheless, it would still be reassuring if one finds a short-term effect. Table 5 shows that the reform generated positive effects on deposits and loans, loans to small firms, and the number of firms that obtained loans, where these variables are transformed by inverse hyperbolic sine to deal with zeros, and the coefficients can thus be roughly interpreted as percent changes. While the estimates are less precise, the economic magnitudes

[25] Alternatively, one could measure the topics using unsupervised LDA topic models. Given the small number of counties and hence small size of corpus, here I use simple dictionary method to measure various policy topics and leave the LDA method to future analysis.

[26] Binswanger and Oechslin (2020) show that better economic statistics could discourage policy experimentation, as the government is less likely to receive the "benefit of the doubt" if the numbers reveal a failure of past attempts. In this regard, the null effect on policy experimentation could still be meaningful.

are generally larger than 10%. To benchmark such magnitudes, Colonnelli and Prem (2022) show that random audits on local governments in Brazil increased bank loans and deposits by about 3%. Notably, the effect on loans to small firms is the largest and statistically significant (coef.=0.263, s.e.=0.116). The role of small and medium enterprises (SMEs) in economic development is well documented in the literature and has been instrumental in China's recent economic progress. However, it is also true that such firms face severe credit constraints (Beck, Demirgüç-Kunt and Maksimovic, 2008; Ayyagari, Demirgüç-Kunt and Maksimovic, 2010). The significant increase in loans to small firms thus speaks to positive policy shifts by local governments. The event study graphs in Appendix 9 provide a more stark pattern by showing significant effects two years after the reform, with no pre-trends before the reform. Finally, one alternative interpretation is that the increase in loans could be driven by the increase in firm demand rather than government-led credit supply. To alleviate this concern, I provide additional evidence in Appendix Table B12 where I find more pronounced effects among banks controlled by local governments.

7.3 Firm entry

To further check whether the policy shifts by local governments generated any real impact, I examine whether firm entry increased after the reform. I use the universe of firm registration data from Dong et al. (2021), which is available for the years 2005, 2010, and 2015. The firms are classified by ownership into four types: private firms, state-owned enterprises (SOEs), foreign-owned firms, and collectively owned firms. I aggregate the firm-level registration data at the county-year level. To deal with zeros, the dependent variable is the inverse hyperbolic sine transformation of the number of registrations so that the coefficients can be roughly interpreted as percent changes. Table 6, column 1 shows a positive and significant effect on the total number of firm registrations (coef.=0.046, s.e.=0.026). To put this 4.6% increase in firm entry into perspective: Giannetti et al. (2021) show that the anti-corruption campaign launched in 2013 in China increased firm entry by 6.7% for a province-industry that was ex ante one standard deviation more corrupt than the average. Although the specification is different, it nonetheless provides some reassurance that my estimate is of considerable economic significance.

When examining the effect by ownership in the remaining columns, I also find an increase in the entry of private firms (coef.=0.048, s.e.=0.027), SOEs (coef.=0.169, s.e.=0.062), and foreign-owned firms (coef.=0.041, s.e.=0.051). The positive and significant effect on private firms is consistent with the fact that these firms have higher productivity (Song, Storesletten and Zilibotti, 2011) and thus may contribute greatly to economic growth. The larger effect on SOEs may seem puzzling as SOEs are generally viewed as less productive, but can be well reconciled with two facts: (1) the total number of SOEs at that time accounts for less than 5% of all firms, so the increase is marginal in an absolute sense; (2) the productivity of SOEs actually was converging to that of private firms after nearly a decade of productivity-enhancing reforms in the state sector since the late 1990s (Hsieh and Song, 2015). Appendix Figure A7 confirms this fact.[27] Finally, as shown in the last column, the effect on collectively owned firms is negative, consistent with the fact that these firms are inefficient (see Appendix Figure A7), although the estimate is imprecise (coef.=-0.035, s.e.=0.062). The event study graphs in Figure 10 show that the increase in firm entry happened immediately after the reform. To alleviate concerns about diverging pre-trends, which cannot be examined with only three periods of data, I instrument the treatment using the randomly assigned rural survey teams in 1984. The 2SLS estimates shown in Appendix Figure A8 are similar, albeit with less precision. Taken together, I find suggestive evidence showing that the reform boosted firm entry, especially for those with higher productivity, which corroborates the previous findings of a positive shift in government policies.

7.4 Citizen attitudes

So far, I have shown that the reform induced local leaders to choose better policies. To further bolster this argument, in this section I test whether citizens' attitudes towards local officials and governments improved as reflected in survey data. I use the China Family Panel Studies (CFPS) and pool two waves, 2014 and 2016, together.[28] The CFPS surveys citizens' trust in local officials by asking, "*To what extent do you trust local officials?*" The answer ranges from 0 to 10, with larger integers denoting higher trust. To alleviate concerns that people may not express their opinions faithfully, I drop the top decile of people who show the highest concerns about the survey, which are observed and recorded by the investigators. Due to the lack of pre-reform survey data, I follow the empirical strategy in Duflo (2001) to estimate a cohort DiD

[27] Appendix Figure A7 calculates the simple average total factor productivity (TFP) by ownership for manufacturing firms from 1998 to 2007 using the commonly used Annual Survery of Industrial Firms dataset and the Olley and Pakes method. The findings are: (1) foreign firms had the highest TFP; (2) private firms had the second-highest TFP; (3) SOEs' TFP used to be the lowest but converged quickly to that of private firms and actually exceeded private firms after 2006; (4) collective firms generally had the lowest TFP. The patterns are robust to weighting the TFP by value added or using alternative methods to calculate TFP (e.g., the Levinsohn and Petrin method)

[28] The earlier waves, 2010 and 2012, are not used as it may take time for both local governments to take actions and the citizens to change attitudes.

specification that utilizes two sources of variation: (1) treatment counties versus control counties; (2) affected cohorts versus unaffected cohorts within the same county. The latter source of variation is built on insights in the psychology literature that political attitudes are most permeable during teenage years and keep stable after one's 30s (Krosnick and Alwin, 1989). I create three cohort groups: those born in the 1970s, in the 1980s, and in the 1990s. The 1970s cohort would be older than 30 during the reform period and therefore are defined as unaffected by the reform.[29] The estimation equation is:

$$Y_{icg} = \sum_{k=1980s,1990s,k\neq 1970s} \beta_k Treat_c \times 1_{\{g=k\}} + \delta_c + \lambda_g + W_i\Omega + X_{cg}\Psi + \varepsilon_{icg} \tag{12}$$

where Y_{icg} denotes the trust in local officials for individual i living in county c and born in cohort $g \in \{1970s, 1980s, 1990s\}$. The 1970s cohort is the omitted group. W_i denotes a set of individual controls, including schooling and its square, age and its square, dummy for male, dummy for living in the urban area, and dummy for survey wave. δ_c and λ_g are county and cohort fixed effects, which help to partial out county- and cohort-specific time-invariant confounding factors, respectively. To the extent that the treatment $Treat_c$ may be correlated with county characteristics, which may further have differential impacts on trust by cohorts, I include in X_{cg} the same set of county controls as before, interacted with cohort fixed effects.

Table 7 presents the results, which are also visualized in Figure 11. Column 1 shows that the reform generated a positive shift in trust in local officials, and the effects are mainly concentrated among the young cohort, namely the 1990s cohort (coef.=0.472, s.e.=0.169). In terms of magnitude, the estimate indicates a 10% increase in trust (relative to the mean trust across all three cohorts) among the 1990s cohort. To tighten identification, I utilize another question in the CFPS to conduct a placebo check. The question asks, "*Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?*" Column 2 uses a dummy variable indicating trust in most people as dependent variable. The effect is statistically insignificant and economically small relative to the mean. I further utilize another two questions in the CFPS to explore additional impacts. The first question asks, "*What is your overall evaluation of the county government's achievements last year?*" The raw answer ranges from 1 to 5, with smaller integers denoting higher achievements. To ease interpretation, I reverse this answer with 5 denoting the highest achievements. The second question asks,

[29] Older cohorts, such as those born in the 1950s and the 1960s, are not used as controls as they grew up in turbulent times when China suffered from several catastrophic events (e.g., the Great Famine and the Cultural Revolution), making them less comparable to younger cohorts.

"*In general, how serious do you think the problem of government corruption is in our country?*" The answer ranges from 0 to 10, with larger integers denoting higher perceived corruption. The results in Column 3 indicate that the reform increased citizens' evaluation of local government performance both for the 1980s (coef.=0.105, s.e.=0.062) and the 1990s (coef.=0.103, s.e.=0.06) cohorts. These effects correspond to about 3% of the mean. Column 4 shows a negative effect on perceived corruption, although s-tatistically insignificant. Two factors may contribute to such insignificance. First, the question is not geared towards local governments. Second, corruption typically refers to bribery or misappropriation of public property, which may thus have little correlation with data manipulation. Finally, Appendix Figure A9 provides 2SLS results using the randomly assigned rural survey teams in 1984 as an instrument, which provide similar patterns. These results suggest that the reform improved citizens' attitudes towards local governments, thereby bolstering the previous findings that government policies improved after the reform.

8 Relation to the literature and discussion

The findings in the last section resonate with the literature on the economic effects of anti-corruption crackdowns or audits (Giannetti et al., 2021; Colonnelli and Prem, 2022). That said, as I find that the reform did not change citizens' perception about corruption, it is likely that the reform in my setting works through mechanisms slightly different from how anti-corruption campaigns work. Specifically, while corruption often involves the abuse of public office for private gains and thus directly stifles the economy, GDP growth manipulation does not directly affect the economy. Instead, it diverts away local leaders' efforts. To firmly support such an argument, one still needs to address the concern that the perceived corruption may not reflect actual corruption. I therefore examine the effect on corruption using an objective corruption measure—the number of officials convicted for corruption during the recent anti-corruption campaign since 2013, using the data complied by Wang and Dickson (2022).[30] The data is the most comprehensive database on China's corruption convictions and includes each official's name, position, locality, and reason for convictions. An inspection of the reason for convictions reveals that most cases involve bribery (60%) and appropriation of public property (22%), which is consistent with the conventional definition of corruption. The remaining convictions involve other misbehav-

[30] Wang and Dickson (2022) collect the data from Tencent—the largest Internet company in China. In 2011 Tencent launched a searchable online database of all corruption investigations across China, and the authors scraped the website in August 2016. Unfortunately, the website is closed currently.

iors such as sexual scandals.[31] As the data contains very few corruption convictions before 2012, I drop convictions happening before 2012 and then collapse the data by county. I run a cross-sectional regression using as instrument the randomly assigned rural survey teams in 1984. Namely, I estimate:

$$Treat_c = \alpha + \beta Treat_c^{1984} + \epsilon_c \tag{13}$$

$$Conviction_c = \gamma + \delta \widehat{Treat_c} + \varepsilon_c \tag{14}$$

where $Conviction_c$ is a measure of corruption convictions in county c and other variables are as defined previously. As a starting point, I focus on bribery and appropriation of public property in Table 8. The results are generally small and statistically insignificant. To alleviate the concern that such corruption convictions data reflects law enforcement (Glaeser and Saks, 2006; Zhu, 2017), I also control for the number of anti-corruption inspections using the data from Wang (2021) in the last two columns, and the results are unchanged. In the first two columns of Appendix Table B13, I further show no effect on all types of corruption and the remaining types of corruption (apart from bribery and appropriation of public property). To alleviate the concern that there may exist lags between the occurrence of corruption behaviors and subsequent convictions, which may mechanically drive the results to zero if all the corruption behaviors happened before 2009, I focus on more recent corruption convictions, e.g., those in 2015 and 2016, in the last two columns.[32] The results have no substantial change.

9 Conclusion

There have been long-standing debates and anecdote evidence on GDP manipulation, especially in autocracies, although causal evidence is still scarce. Further, how to curb such misbehavior remains obscure. This paper focuses on China and provides new and causal evidence regarding this issue. By constructing a dataset at the county level and exploiting various empirical strategies, including an instrumented difference-in-differences design, I provide causal evidence that top-down monitoring of local governments can reduce GDP manipulation, even in contexts like China where the legal system is dysfunctional. I further demonstrate that the monitoring generated far-reaching effects by eliciting government efforts in directions conducive to eco-

[31] However, the reason for investigation does not include GDP manipulation or other data manipulation.

[32] The corruption convictions in 2015 and 2016 account for 61% of all convictions.

nomic development. As a result, credit expanded, firm entry increased, and citizen attitudes towards local governments improved.

These striking downstream patterns provide a new perspective to how individual local officials could affect the macro-level economic outcomes. Unlike corruption, which could directly stifle economic development (Shleifer and Vishny, 1993), GDP manipulation has a more subtle impact—it hurts the economy by inducing an unfavorable shift of local officials' efforts. In this regard, the findings in this paper also have implications for a variety of settings such as hierarchical organizations with imperfect observation from the top.

This study has some limitations, and I outline some directions for future research. First, the question of how top-down monitoring affects the economy is essentially a general equilibrium question. The research design in this paper, by construction, may only estimate the lower bound and partial equilibrium effect of the monitoring. Future work may explore this impact using structural approaches. Second, future work could enrich our understanding of the monitoring effect by unpacking the impact on firms once more recent firm-level data are available, and the effect on political promotion by tracking the career paths of local officials.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. "When should you adjust standard errors for clustering?" National Bureau of Economic Research.

Agarwal, Sumit, and Robert Hauswald. 2010. "Distance and private information in lending." *The Review of Financial Studies*, 23(7): 2757–2788.

Aghion, Philippe, and Jean Tirole. 1997. "Formal and real authority in organizations." *Journal of political economy*, 105(1): 1–29.

Ajzenman, Nicolás. 2021. "The power of example: Corruption spurs corruption." *American Economic Journal: Applied Economics*, 13(2): 230–57.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2014. *Mastering'metrics: The path from cause to effect.* Princeton university press.

Avis, Eric, Claudio Ferraz, and Frederico Finan. 2018. "Do government audits reduce corruption? Estimating the impacts of exposing corrupt politicians." *Journal of Political Economy*, 126(5): 1912–1964.

Ayyagari, Meghana, Asli Demirgüç-Kunt, and Vojislav Maksimovic. 2010. "Formal versus informal finance: Evidence from China." *The Review of Financial Studies*, 23(8): 3048–3097.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2009. "Social connections and incentives in the workplace: Evidence from personnel data." *Econometrica*, 77(4): 1047–1094.

Becker, Gary S, and George J Stigler. 1974. "Law enforcement, malfeasance, and compensation of enforcers." *The Journal of Legal Studies*, 3(1): 1–18.

Beck, Thorsten, Asli Demirgüç-Kunt, and Vojislav Maksimovic. 2008. "Financing patterns around the world: Are small firms different?" *Journal of financial economics*, 89(3): 467–487.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on treatment effects after selection among high-dimensional controls." *The Review of Economic Studies*, 81(2): 608–650.

Besley, Timothy J, Robin Burgess, Adnan Khan, and Guo Xu. 2021. "Bureaucracy and development."

Binswanger, Johannes, and Manuel Oechslin. 2020. "Better statistics, better economic policies?" *European Economic Review*, 130.

Bobonis, Gustavo J, Luis R Cámara Fuertes, and Rainer Schwabe. 2016. "Monitoring corruptible politicians." *American Economic Review*, 106(8): 2371–2405.

Campante, Filipe R, Davin Chor, and Bingjing Li. 2022. "The Political Economy Consequences of China's Export Slowdown." Working Paper.

Chen, Ting, and James Kai-sing Kung. 2016. "Do land revenue windfalls create a political resource curse? Evidence from China." *Journal of Development Economics*, 123: 86–106.

Chu, Jian, Raymond Fisman, Songtao Tan, and Yongxiang Wang. 2021. "Hometown ties and the quality of government monitoring: evidence from rotation of Chinese auditors." *American Economic Journal: Applied Economics*, 13(3): 176–201.

Colonnelli, Emanuele, and Mounu Prem. 2022. "Corruption and Firms." *The Review of Economic Studies*, 89(2): 695–732.

Dong, Lei, Xiaohui Yuan, Meng Li, Carlo Ratti, and Yu Liu. 2021. "A gridded establishment dataset as a proxy for economic activity in China." *Scientific Data*, 8(1): 1–9.

Do, Quoc-Anh, Kieu-Trang Nguyen, and Anh N Tran. 2017. "One mandarin benefits the whole clan: hometown favoritism in an authoritarian regime." *American Economic Journal: Applied Economics*, 9(4): 1–29.

Duflo, Esther. 2001. "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." *American economic review*, 91(4): 795–813.

Fan, Gang, Xiaolu Wang, and Hengpeng Zhu. 2003. "NERI index of marketization of China's provinces." *National Economic Research Institute, Beijing*.

Fan, Joseph PH, TJ Wong, and Tianyu Zhang. 2013. "Institutions and organizational structure: The case of state-owned corporate pyramids." *The Journal of Law, Economics, and Organization*, 29(6): 1217–1252.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig. 2017. "Cultural proximity and loan outcomes." *American Economic Review*, 107(2): 457–492.

Giannetti, Mariassunta, Guanmin Liao, Jiaxing You, and Xiaoyun Yu. 2021. "The externalities of corruption: Evidence from entrepreneurial firms in China." *Review of Finance*, 25(3): 629–667.

Glaeser, Edward L, and Raven E Saks. 2006. "Corruption in america." *Journal of public Economics*, 90(6-7): 1053–1072.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277.

Hainmueller, Jens. 2012. "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies." *Political Analysis*, 20(1): 25–46.

Hart, Oliver. 1995. *Firms, contracts, and financial structure.* Clarendon press.

Henderson, J Vernon, Adam Storeygard, and David N Weil. 2012. "Measuring economic growth from outer space." *American Economic Review*, 102(2): 994–1028.

Hsieh, Chang-Tai, and Zheng Michael Song. 2015. "Grasp the large, let go of the small: The transformation of the state sector in China." National Bureau of Economic Research.

Huber, Kilian. 2023. "Estimating general equilibrium spillovers of large-scale shocks." *The Review of Financial Studies*, 36(4): 1548–1584.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Jia, Junxue, Xuan Liang, and Guangrong Ma. 2021. "Political hierarchy and regional economic development: Evidence from a spatial discontinuity in China." *Journal of Public Economics*, 194: 104352.

Jiang, Junyan, Tianguang Meng, and Qing Zhang. 2019. "From Internet to social safety net: The policy consequences of online participation in China." *Governance*, 32(3): 531–546.

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.

Jones, Benjamin F, and Benjamin A Olken. 2005. "Do leaders matter? National leadership and growth since World War II." *The Quarterly Journal of Economics*, 120(3): 835–864.

Krosnick, Jon A, and Duane F Alwin. 1989. "Aging and susceptibility to attitude change." *Journal of Personality and Social Psychology*, 57(3): 416.

Lazear, Edward P, and Sherwin Rosen. 1981. "Rank-order tournaments as optimum labor contracts." *Journal of Political Economy*, 89(5): 841–864.

Lee, David S, Justin McCrary, Marcelo J Moreira, and Jack R Porter. 2022. "Valid t-ratio Inference for IV." *American Economic Review*.

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.

Li, Kai, Tan Wang, Yan-Leung Cheung, and Ping Jiang. 2011. "Privatization and risk sharing: Evidence from the split share structure reform in China." *The Review of Financial Studies*, 24(7): 2499–2525.

Li, Pei, Yi Lu, and Jin Wang. 2016. "Does flattening government improve economic performance? Evidence from China." *Journal of Development Economics*, 123: 18–37.

Li, Xuecao, Yuyu Zhou, Min Zhao, and Xia Zhao. 2020. "A harmonized global night-time light dataset 1992–2018." *Scientific Data*, 7(1): 1–9.

MacKinnon, James G, and Matthew D Webb. 2020. "Randomization inference for difference-in-differences with few treated clusters." *Journal of Econometrics*, 218(2): 435–450.

Martinez-Bravo, Monica, Gerard Padró i Miquel, Nancy Qian, and Yang Yao. 2022. "The rise and fall of local elections in China." *American Economic Review*, 112(9): 2921–58.

Martinez, Luis R. 2021. "How Much Should We Trust the Dictator's GDP Growth Estimates?" *Forthcoming, Journal of Political Pconomy*.

Maskin, Eric, Yingyi Qian, and Chenggang Xu. 2000. "Incentives, information, and organizational form." *The Review of Economic Studies*, 67(2): 359–378.

Olken, Benjamin A. 2007. "Monitoring corruption: evidence from a field experiment in Indonesia." *Journal of Political Economy*, 115(2): 200–249.

Persson, Petra, and Ekaterina Zhuravskaya. 2016. "The limits of career concerns in federalism: Evidence from China." *Journal of the European Economic Association*, 14(2): 338–374.

Petersen, Mitchell A, and Raghuram G Rajan. 2002. "Does distance still matter? The information revolution in small business lending." *The journal of Finance*, 57(6): 2533–2570.

Qian, Jun, Philip E Strahan, and Zhishu Yang. 2015. "The impact of incentives and communication costs on information production and use: Evidence from bank lending." *The Journal of Finance*, 70(4): 1457–1493.

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.

Shleifer, Andrei, and Robert W Vishny. 1993. "Corruption." *The quarterly journal of economics*, 108(3): 599–617.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing like china." *American economic review*, 101(1): 196–233.

Staiger, Douglas O, and James H Stock. 1997. "Instrumental variables regression with weak instruments." *Econometrica*, 65(3): 557–586.

Stein, Jeremy C. 2002. "Information production and capital allocation: Decentralized versus hierarchical firms." *The journal of finance*, 57(5): 1891–1921.

Vannutelli, Silvia. 2021. "From lapdogs to watchdogs: Random auditor assignment and municipal fiscal performance in Italy." *working paper*.

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

Wang, Erik H. 2021. "Frightened mandarins: the adverse effects of fighting corruption on local bureaucracy." *Forthcoming, Comparative Political Studies*.

Wang, Yuhua, and Bruce J Dickson. 2022. "How corruption investigations undermine regime support: Evidence from China." *Political Science Research and Methods*, 10(1): 33–48.

Weber, Max. 1922. *Economy and society.* Tübingen.

Xiong, Wei. 2018. "The mandarin model of growth." National Bureau of Economic Research.

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(4): 405–436.

Zhu, Boliang. 2017. "MNCs, rents, and corruption: Evidence from China." *American Journal of Political Science*, 61(1): 84–99.

Figures and Tables

Figure 1: Treatment counties vs control counties

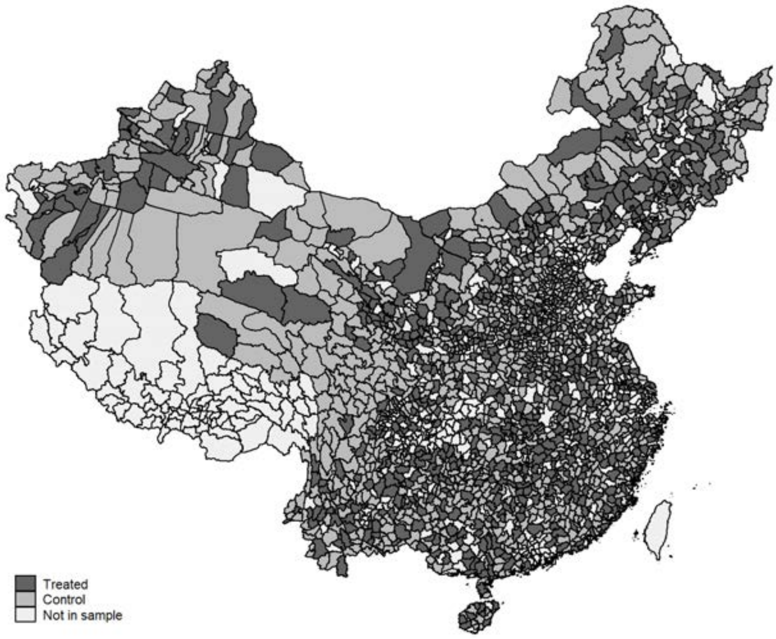
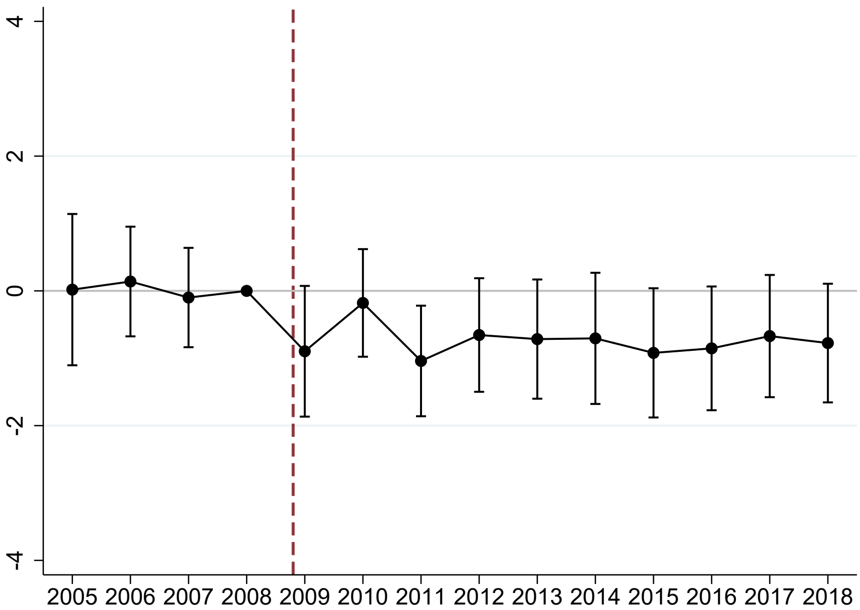


Figure 2: Dynamic effect on GDP growth manipulation



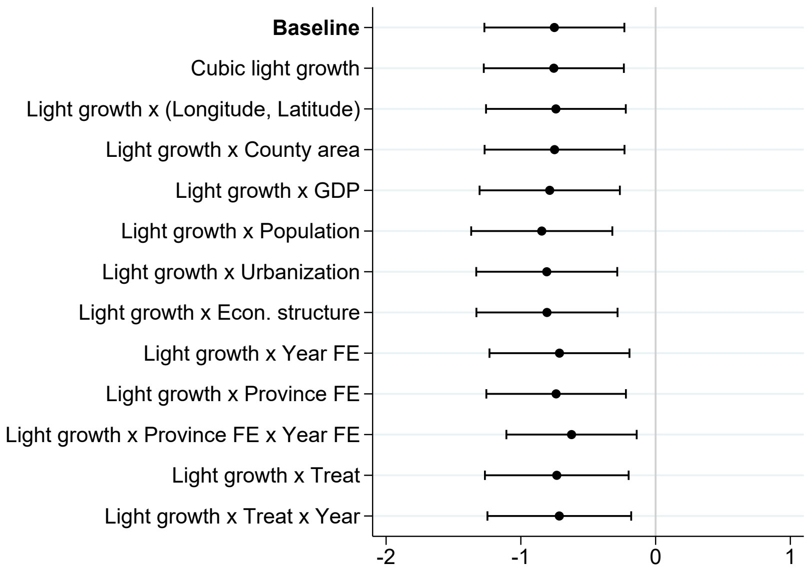
Notes: This figure shows the dynamic effect of the reform on GDP growth manipulation, which is estimated using the following equation:

$$ReportedGDPGrowth_{ct} = \alpha LightGrowth_{ct} + \sum_{j=2005,\ j\neq 2008}^{j=2018} \beta_j Treat_c \times 1_{\{t=j\}} + \delta_c + \lambda_t + \varepsilon_{ct}$$

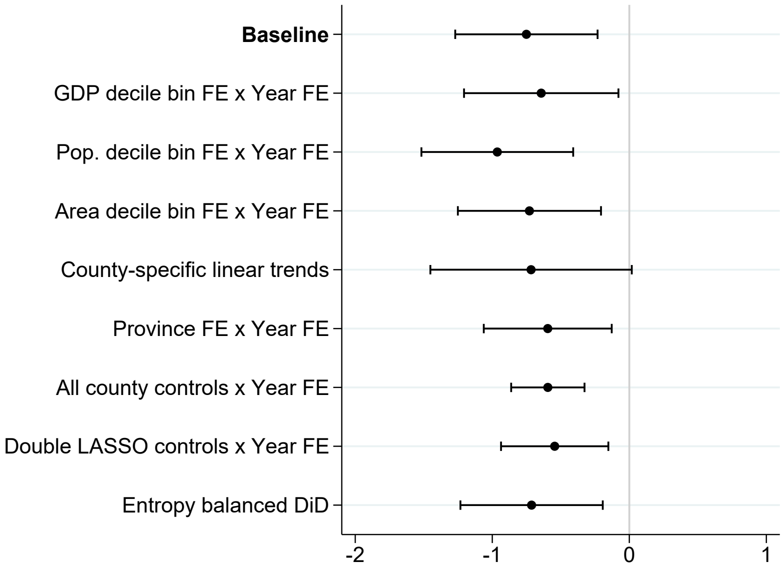
where the year 2008, one year before the reform, is omitted as the reference year. Standard errors used to construct the 90% confidence intervals, which are denoted by the spikes, are clustered at the county level.

Figure 3: Sensitivity to alternative specifications

A: Flexible mapping between light growth and economic growth

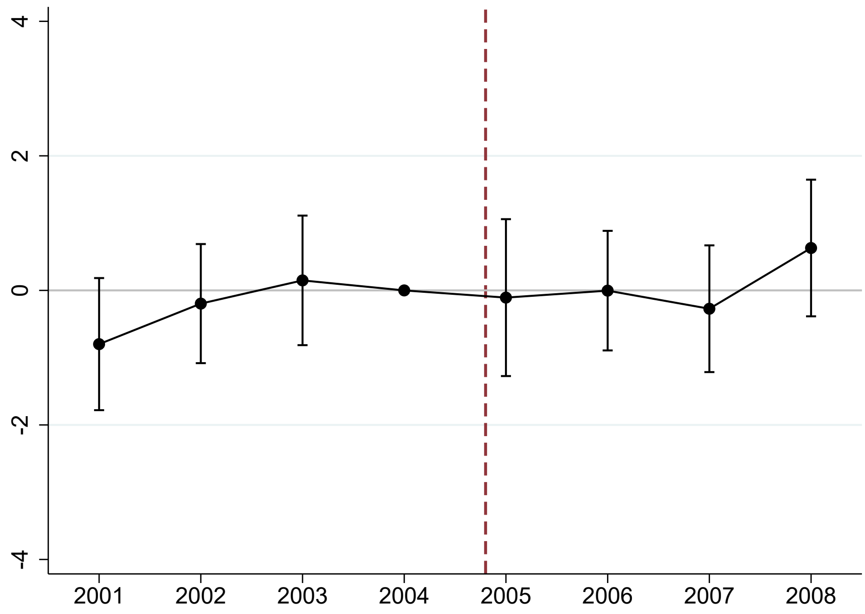


B: Addressing unbalance between treatment and control groups



Notes: This figure checks the sensitivity of the baseline results to alternative specifications, and is created by estimating variants of the baseline equation (5). Specifically, Panel A addresses the concern that the mapping between light growth and economics may not be uniform across counties or years, by allowing the mapping to vary flexibly; Panel B addresses the concern that the results may be confounded by preexisting unbalance between treatment counties and control counties, by directly controlling for the sources of unbalance or re-weighting the observations to achieve covariate balance, among others. Standard errors used to construct the 90% confidence intervals, which are denoted by the spikes, are clustered at the county level.

Figure 4: Using the launch of the survey teams in 2005 as a placebo



Notes: This figure conducts a placebo event study around 2005, when the survey teams were deployed but had not started disciplining local officials. The estimating equation is:

$$ReportedGDPGrowth_{ct} = \alpha LightGrowth_{ct} + \sum_{j=2001,\ j\neq 2004}^{j=2008} \beta_j Treat_c \times 1_{\{t=j\}} + \delta_c + \lambda_t + \varepsilon_{ct}$$

where the year 2004, one year before the deployment of the survey teams, is omitted as the reference year. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

Figure 5: Distribution of the rural survey teams in 1984

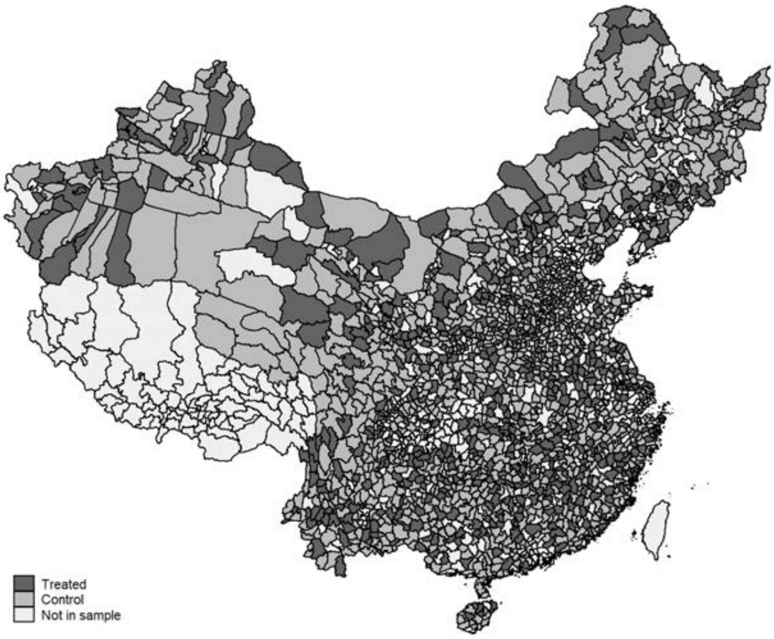
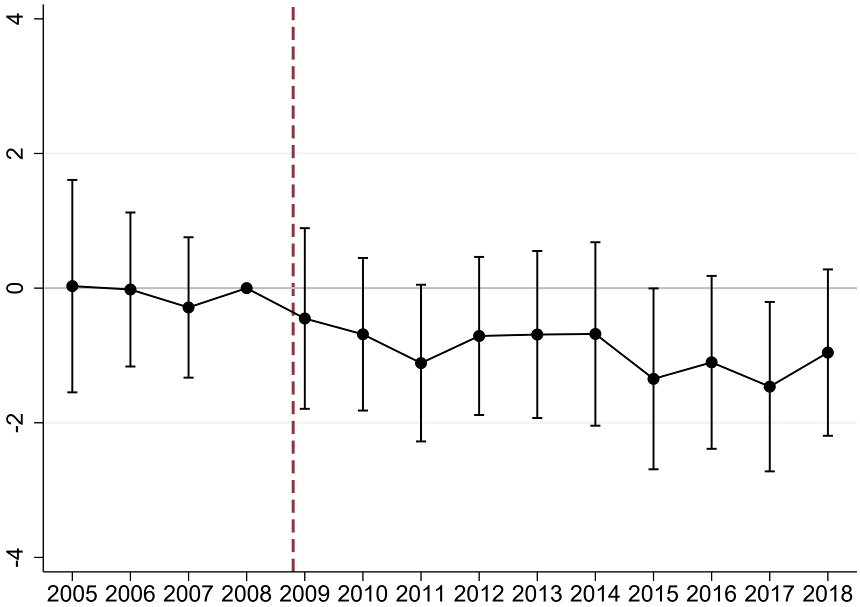


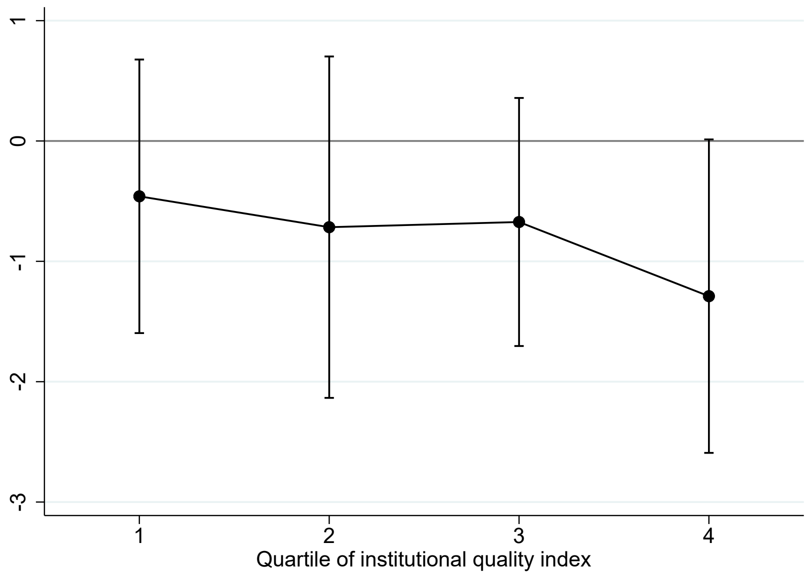
Figure 6: Dynamic effect on GDP growth manipulation - 2SLS estimates



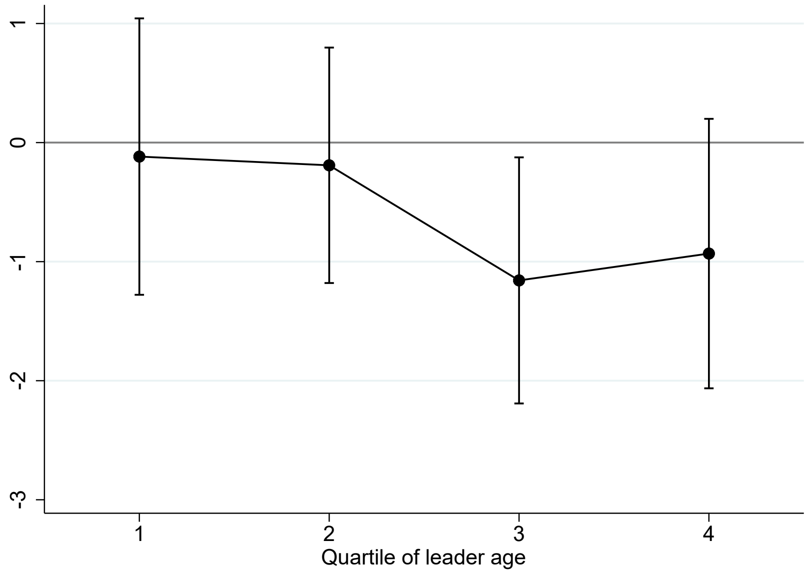
Notes: This figure shows the 2SLS estimates on the dynamic effect of the reform on GDP growth manipulation, and is created by estimating an event study variant of the 2SLS equation (10), with $Post_t$ replaced by a set of year dummies. The year 2008, which is one year before the reform in 2009, is omitted as the reference year. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

Figure 7: Heterogeneous treatment effect

A: By institutional quality

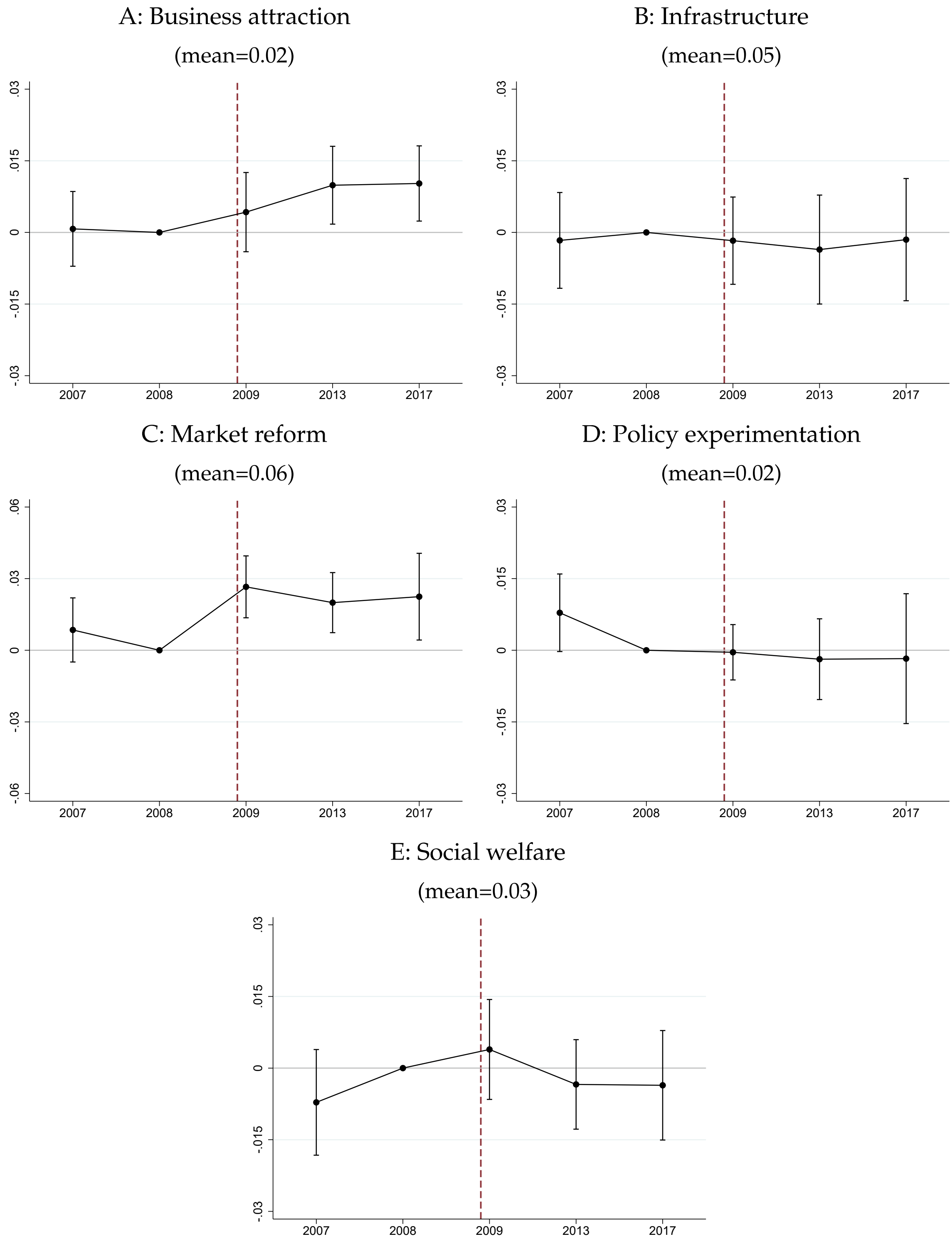


B: By career concerns



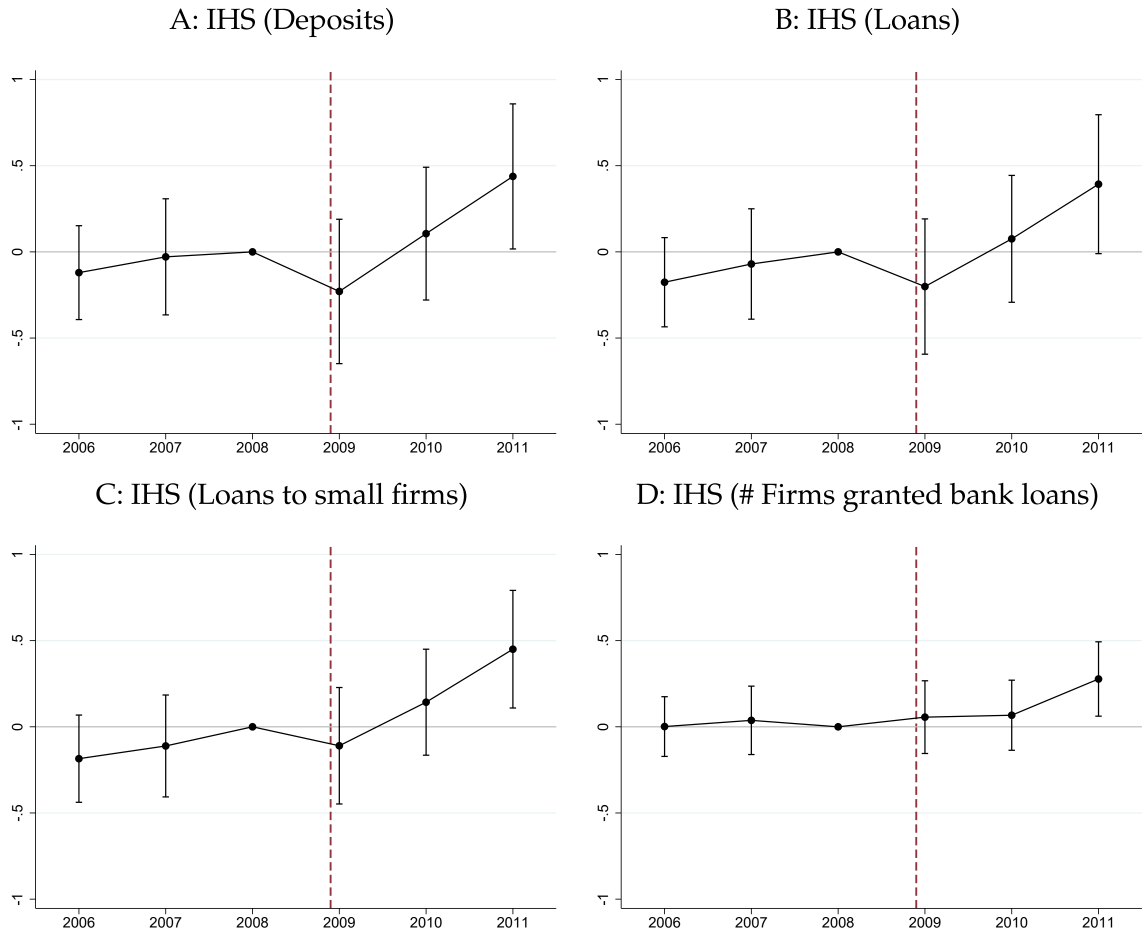
Notes: This figure shows the heterogeneous effect of the reform on GDP growth manipulation, and is created by estimating the baseline equation (5) for each quartile of institutional quality (Panel A) or career concerns (Panel B). Institutional quality is proxied by the marketization index developed by Fan, Wang and Zhu (2003). Career concerns are proxied by the ages of party secretaries, with older ages denoting minor concerns. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

Figure 8: Dynamic effect on government policies



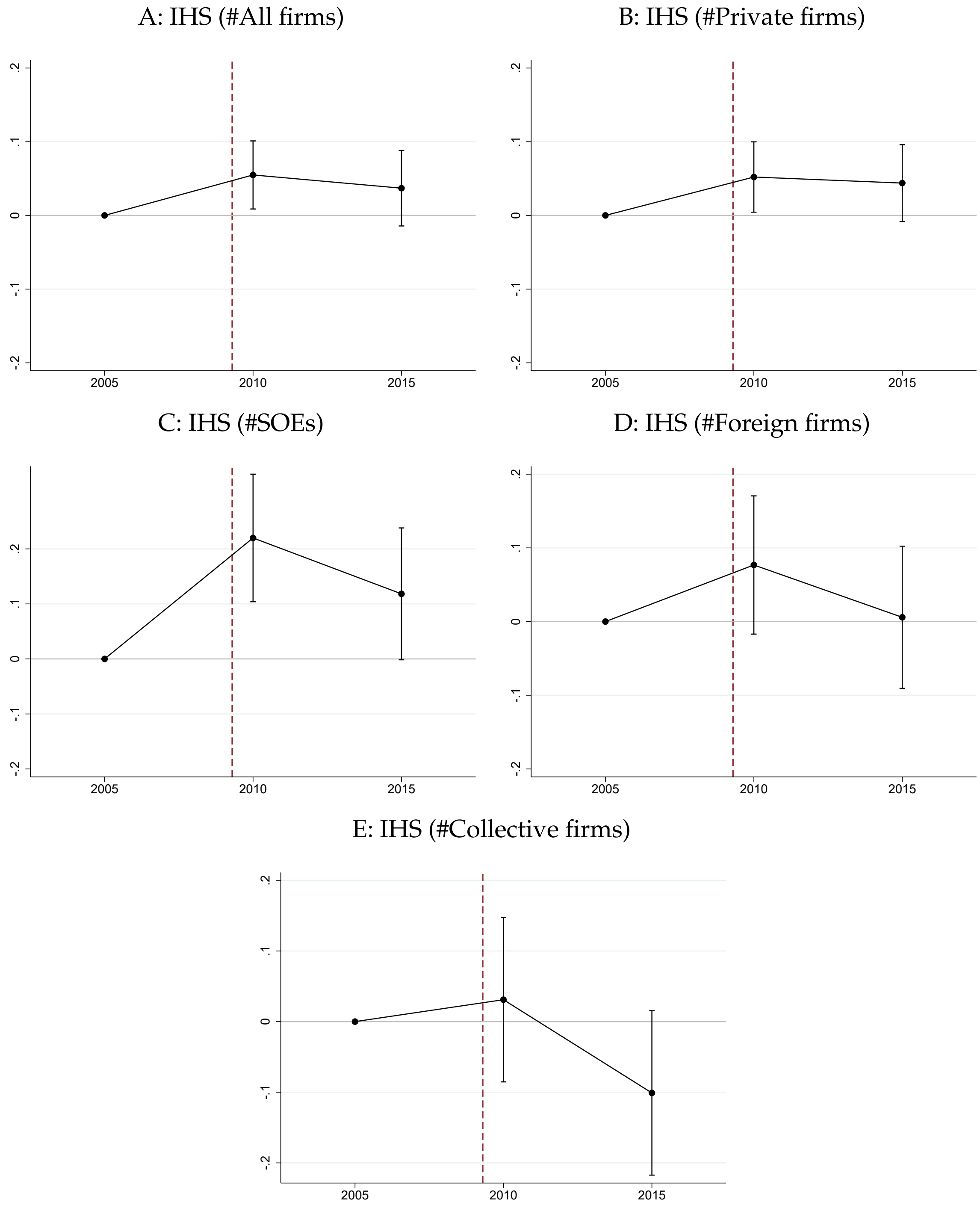
Notes: This figure shows the dynamic effect of the reform on government policies along five dimensions. The first four dimensions (Panel A-D) are dimensions conducive to economic growth, and the last dimension severs as a placebo. The estimating equation is an event study variant of the specification in Table 4. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

Figure 9: Dynamic effect on bank deposits and loans



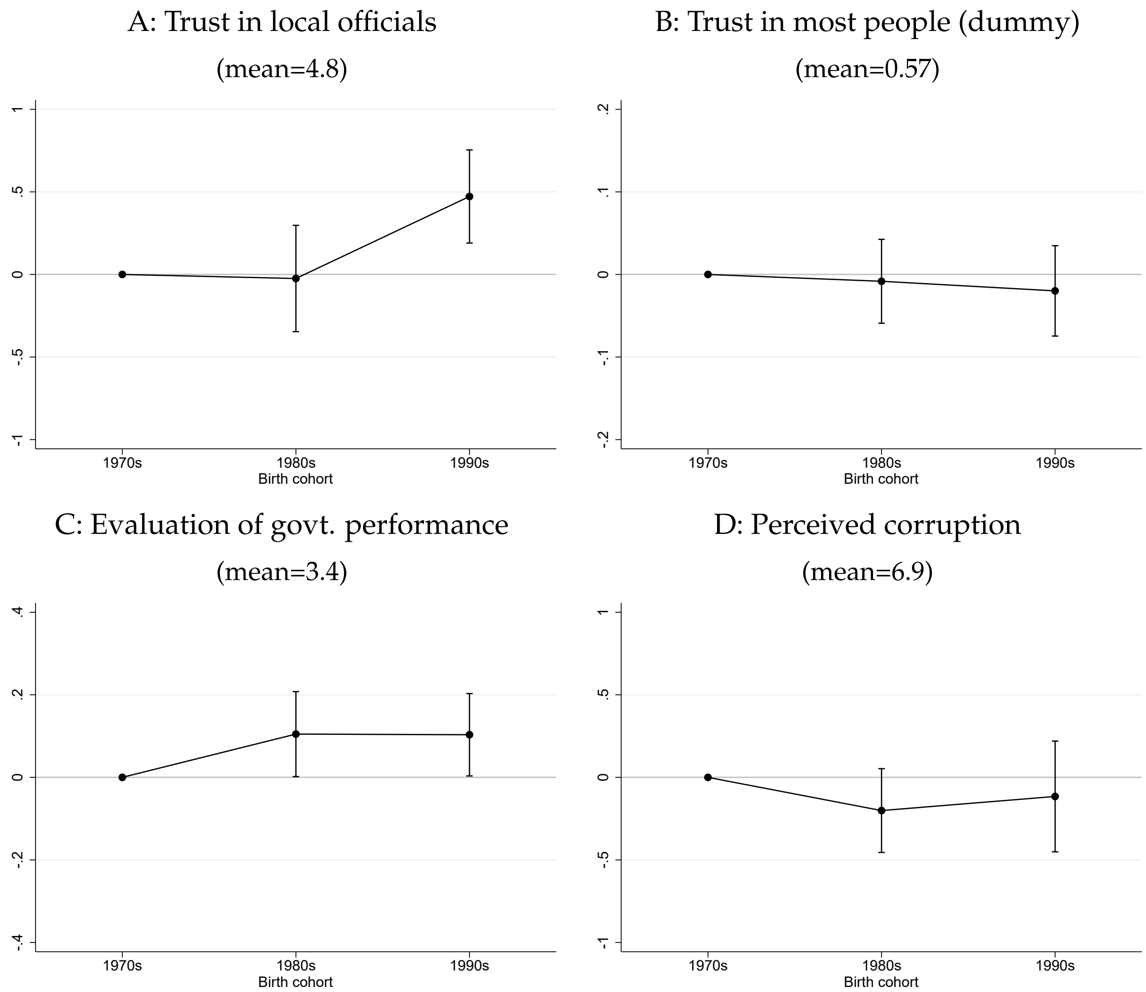
Notes: This figure shows the dynamic effect of the reform on bank deposits and loans. The estimating equation is an event study variant of the specification in Table 5. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level. IHS denotes inverse hyperbolic sine transformation.

Figure 10: Dynamic effect on firm entry



Notes: This figure shows the dynamic effect of the reform on firm entry. The estimating equation is an event study variant of the specification in Table 6. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level. IHS denotes inverse hyperbolic sine transformation.

Figure 11: Dynamic effect on citizen attitudes



Notes: This figure shows the dynamic effect of the reform on citizen attitudes, and is created by visualizing the results in Table 7. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

Table 1: Effect on GDP growth manipulation

	(1)	(2)	(3)	(4)
Dep. var.:	Reported GDP growth (%)			
Treat x Post	-0.751**	-0.869***	-0.552***	-0.576***
	(0.316)	(0.331)	(0.162)	(0.161)
Light growth (%)	0.023***	0.022***	0.017***	0.017***
	(0.005)	(0.005)	(0.005)	(0.005)
Demographic controls x Post	No	Yes	Yes	Yes
Economic controls x Post	No	No	Yes	Yes
Geographic controls x Post	No	No	No	Yes
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County
Observations	23,360	22,580	20,343	20,273
R-squared	0.269	0.269	0.362	0.362
Mean of dep. var.	10.97	10.97	10.84	10.84

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. Standard errors clustered at the county level are reported in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 2: Effect on GDP growth manipulation - 2SLS results

	(1)	(2)	(3)	(4)
Dep. var.:	Reported GDP growth (%)			
Treat x Post	-0.835*	-0.750*	-0.498**	-0.487**
	(0.444)	(0.448)	(0.225)	(0.225)
Light growth (%)	0.023***	0.022***	0.017***	0.017***
	(0.005)	(0.005)	(0.005)	(0.005)
Demographic controls x Post	No	Yes	Yes	Yes
Economic controls x Post	No	No	Yes	Yes
Geographic controls x Post	No	No	No	Yes
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County
Observations	22,998	22,580	20,343	20,273
F-stat of excl. inst.	1945	2210	2044	2080
Mean of dep. var.	10.98	10.97	10.84	10.84

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. This table provides 2SLS estimates of the baseline results with $Treat_c^{1984}$ (dummy for counties with the randomly assigned rural survey teams in 1984) as an instrument for Treat. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Disciplining effect as the key mechanism

	(1) Baseline	(2) With leader FE	(3) Trimmed sample with no turnover
Dep. var.		Reported GDP growth (%)	
Treat x Post	-0.751**	-0.797*	-0.789**
	(0.316)	(0.417)	(0.389)
Light growth (%)	0.023***	0.016***	0.017***
	(0.005)	(0.005)	(0.006)
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Leader FE	No	Yes	No
Cluster level	County	County	County
Observations	23,360	21,101	15,021
R-squared	0.269	0.343	0.297
Mean of dep. var.	10.97	10.87	11.50

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. Column (2) includes leader fixed effects. Column (3) focuses on a trimmed sample with only leaders whose terms straddle 2009 so that there is no turnover in the sample. Standard errors clustered at the county level are reported in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 4: Effect on government policies

	(1)	(2)	(3)	(4)	(5)
Topic:	Business attraction	Infrastructure	Market reform	Policy experimentation	Social welfare
Dep. var.:	Number of key words found / Number of sentences				
Treat x Post	0.008**	-0.002	0.019***	-0.003	0.001
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
County controls x Post	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County	County
Observations	314	314	314	314	314
R-squared	0.618	0.585	0.661	0.681	0.508
Mean of dep. var.	0.023	0.044	0.063	0.016	0.029

Notes: The unit of observation is county x year. The sample includes only 97 counties and the years 2007, 2008, 2009, 2013, and 2017. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Effect on bank loans and deposits

	(1)	(2)	(3)	(4)
	IHS transformation of			
Dep. var.:	Deposits	Loans	Loans to small firms	Number of firms granted loans
Treat x Post	0.158	0.174	0.263**	0.123
	(0.129)	(0.123)	(0.116)	(0.076)
County controls x Post	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County
Observations	8,922	8,922	8,922	8,922
R-squared	0.311	0.329	0.424	0.520

Notes: The unit of observation is county x year. The sample period is 2006-2011. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. The dependent variable is transformed by inverse hyperbolic sine (IHS) to deal with zeros. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Effect on firm entry

	(1)	(2)	(3)	(4)	(5)
Firm type:	All	Private	SOEs	Foreign	Collective
Dep. var.:	IHS (Number of firm registrations)				
Treat x Post	0.046*	0.048*	0.169***	0.041	-0.035
	(0.026)	(0.027)	(0.062)	(0.051)	(0.062)
County controls x Post	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County	County
Observations	4,494	4,494	4,494	4,494	4,494
R-squared	0.943	0.944	0.535	0.726	0.571

Notes: The unit of observation is county x year. The sample includes the years 2005, 2010, and 2015. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. The dependent variable is transformed by inverse hyperbolic sine (IHS) to deal with zeros. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Effect on citizens attitudes

Dep. var.:	(1) Trust in local officials	(2) Trust in most people	(3) Evaluation of local govt. performance	(4) Perceived corruption
Treat x 1980s cohort	-0.025	-0.008	0.105*	-0.201
	(0.193)	(0.031)	(0.062)	(0.152)
Treat x 1990s cohort	0.472***	-0.020	0.103*	-0.116
	(0.169)	(0.033)	(0.060)	(0.202)
County controls x Cohort FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County
Observations	10,825	10,825	10,649	10,747
R-squared	0.079	0.085	0.085	0.103
Mean of dep. var.	4.850	0.568	3.394	6.918

Notes: The unit of observation is individual (two waves from the CFPS, 2014 and 2016). Treat is a dummy variable indicating counties with the survey teams deployed in 2005. 1980s cohort and 1990s cohort are dummy variables indicating individuals born in the 1980s and the 1990s, respectively. The omitted group is those born in the 1970s. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Effect on corruption convictions - 2SLS estimates

	(1)	(2)	(3)	(4)	(5)
Dep. var.:	IHS (Number of convictions)				
	Panel A: Bribery				
Treat	0.055	0.022	-0.006	0.020	-0.008
	(0.071)	(0.066)	(0.061)	(0.066)	(0.061)
#Anti-corruption inspections				-0.023*	-0.015
				(0.013)	(0.013)
Mean of dep. var.	2.603	2.718	2.718	2.718	2.718
	Panel B: Appropriation				
Treat	0.041	0.039	0.020	0.039	0.020
	(0.052)	(0.054)	(0.054)	(0.054)	(0.054)
#Anti-corruption inspections				0.001	-0.006
				(0.011)	(0.012)
Mean of dep. var.	0.990	0.974	0.974	0.974	0.974
	Panel C: Bribery ∪ Appropriation				
Treat	0.075	0.048	0.020	0.047	0.018
	(0.070)	(0.066)	(0.062)	(0.066)	(0.062)
#Anti-corruption inspections				-0.018	-0.018
				(0.014)	(0.014)
Mean of dep. var.	3.191	3.303	3.303	3.303	3.303
County controls	No	Yes	Yes	Yes	Yes
Province FE	No	No	Yes	No	Yes
Observations	1,752	1,498	1,498	1,498	1,498
F-stat of excl. inst.	1942	2041	2214	2041	2204

Notes: The unit of observation is county. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. The dependent variable is transformed by inverse hyperbolic sine (IHS). This table conducts 2SLS estimation with $Treat_c^{1984}$ (dummy for counties with the randomly assigned rural survey teams in 1984) as an instrument for Treat. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix

Curbing Bureaucratic Information Manipulation

Yongwei Nian

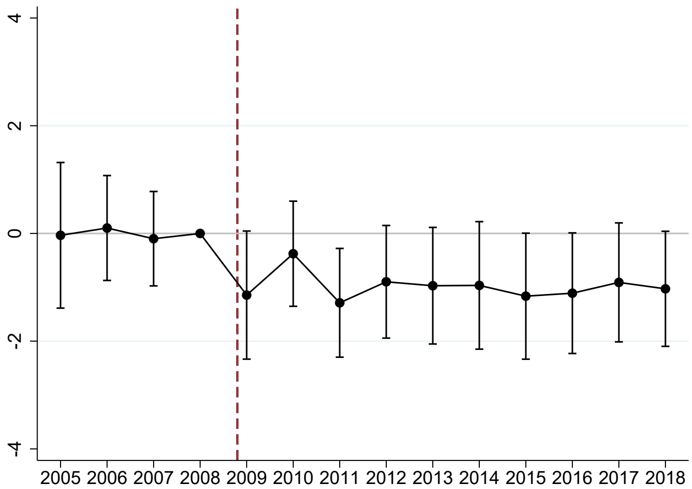
A: Additional Figures

B: Additional Tables

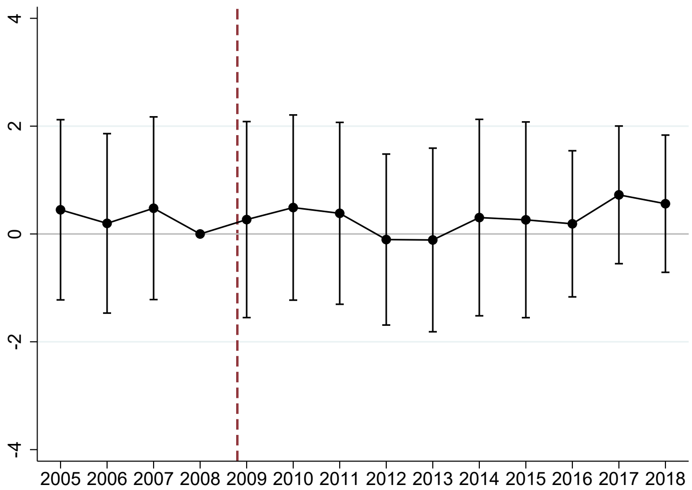
A Additional Figures

Figure A1: Decomposing the effect of the reform

A: Effect on reported GDP growth



B: Effect on light growth

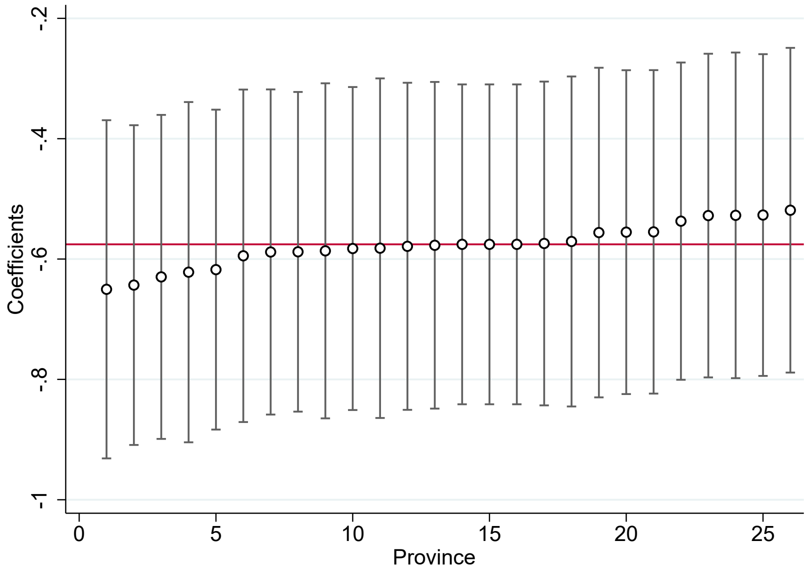


Notes: This figure shows the dynamic effect of the reform on reported GDP growth (Panel A) and light growth (Panel B) separately. Specifically, I estimate:

$$Y_{ct} = \sum_{j=2005,\ j\neq 2008}^{j=2018} \beta_j Treat_c \times 1_{\{t=j\}} + \delta_c + \lambda_t + \varepsilon_{ct}$$

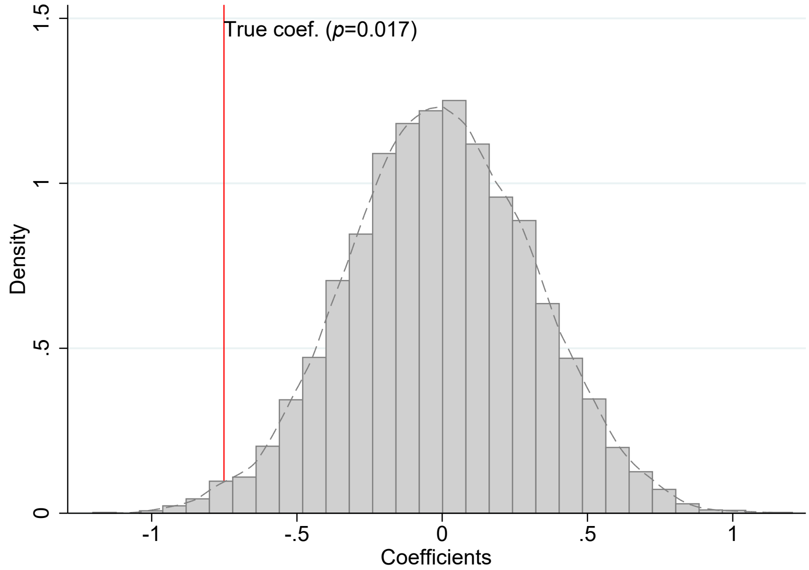
where Y_{ct} denotes either reported GDP growth (Panel A) or light growth (Panel B). Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

Figure A2: Leaving out each province individually



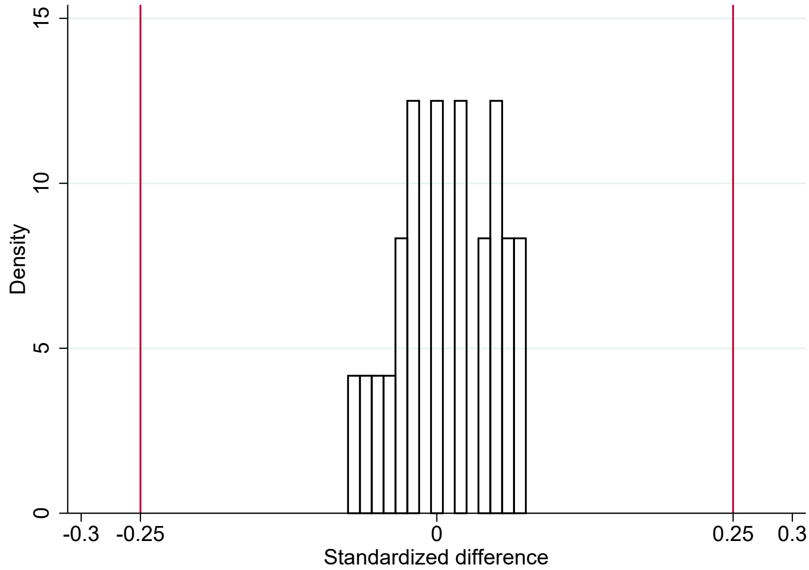
Notes: This figure is created by excluding one province each time and rerunning the baseline specification, with county controls included. The horizontal line denotes the baseline estimates. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

Figure A3: Randomization inference



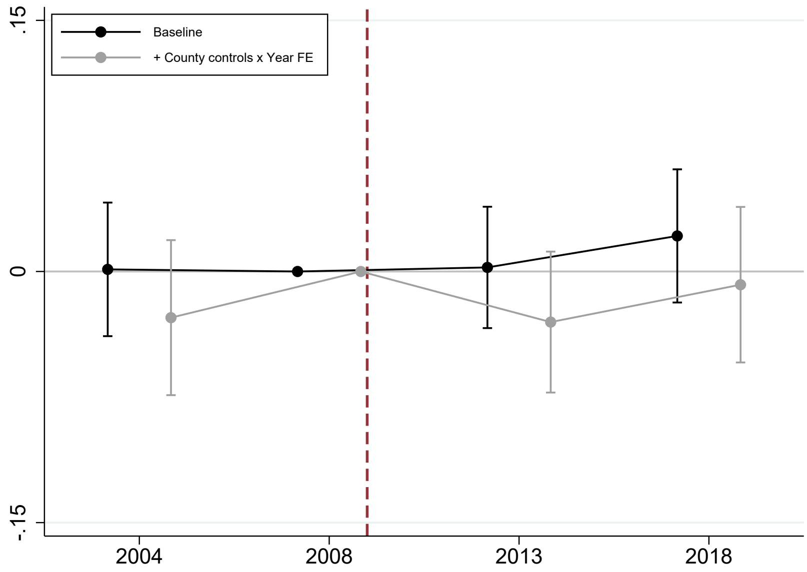
Notes: This figure plots the distribution of coefficient estimates from 10,000 randomized permutations of the treatment using the baseline equation (5). The vertical line denotes the coefficient estimate using the true treatment and the p-value is computed following MacKinnon and Webb (2020). Namely, the fraction of the absolute values of the permutation estimates smaller than the absolute value of the true estimate.

Figure A4: Standardized difference for balance tests of IV



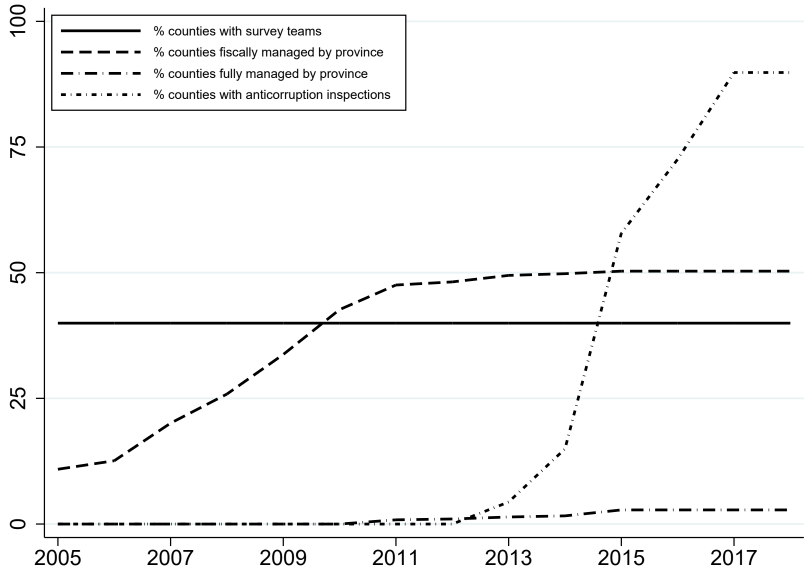
Notes: The standardized difference is calculated as the difference between sample means, normalized by the square root of the average of the sample variances, namely, $(\bar{x}_t - \bar{x}_c)/\sqrt{(s_t^2 + s_c^2)/2}$. The two vertical lines denote the 25% threshold recommended by Imbens and Rubin (2015).

Figure A5: Testing for statistical capacity
(mean=0.13)



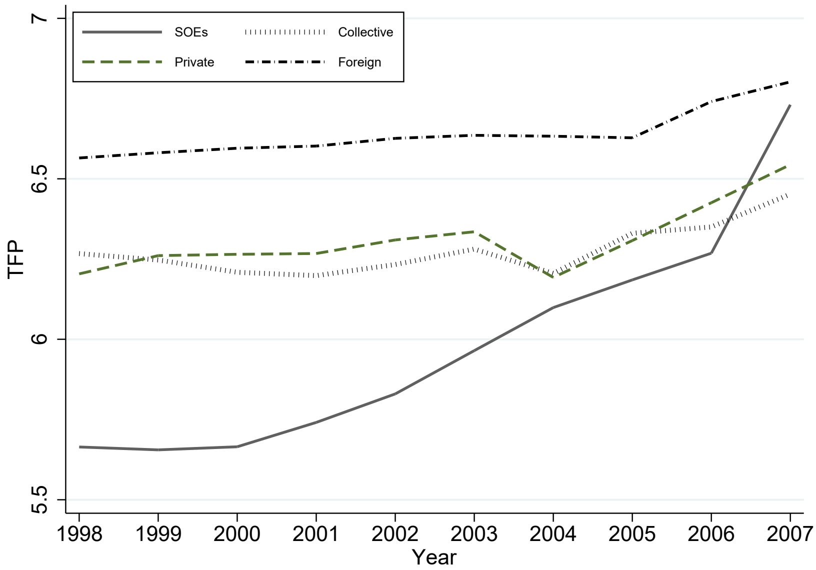
Notes: This figure shows the dynamic effect of the reform on local statistical capacity, and is created by estimating equation (11). The dependent variable is a dummy denoting whether a county won an award for outstanding performance in conducting economic census. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

Figure A6: Rollout of contemporaneous reforms



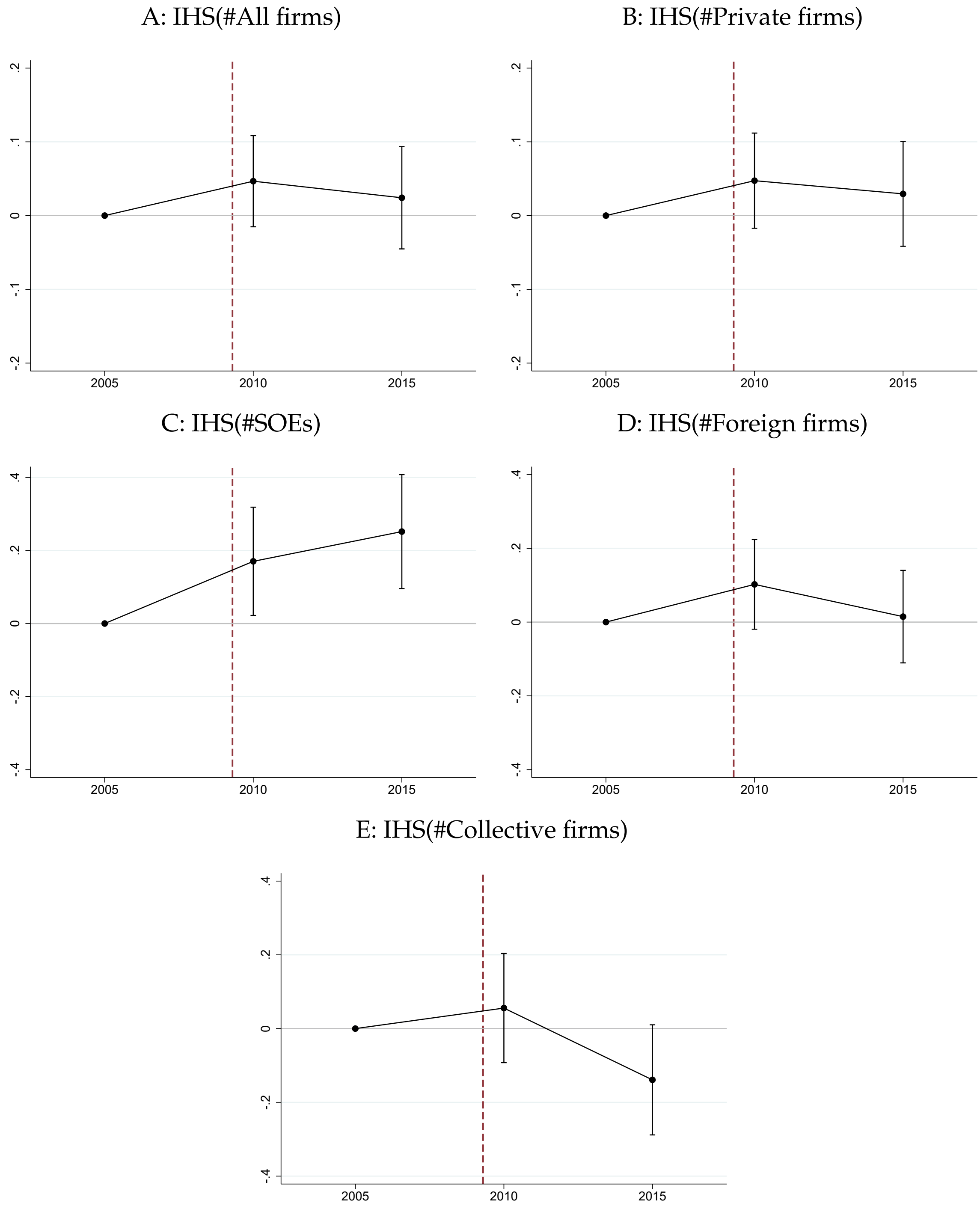
Notes: This figure shows the rollout of various contemporaneous reforms discussed in Section 6.5.

Figure A7: Manufacturing TFP by ownership



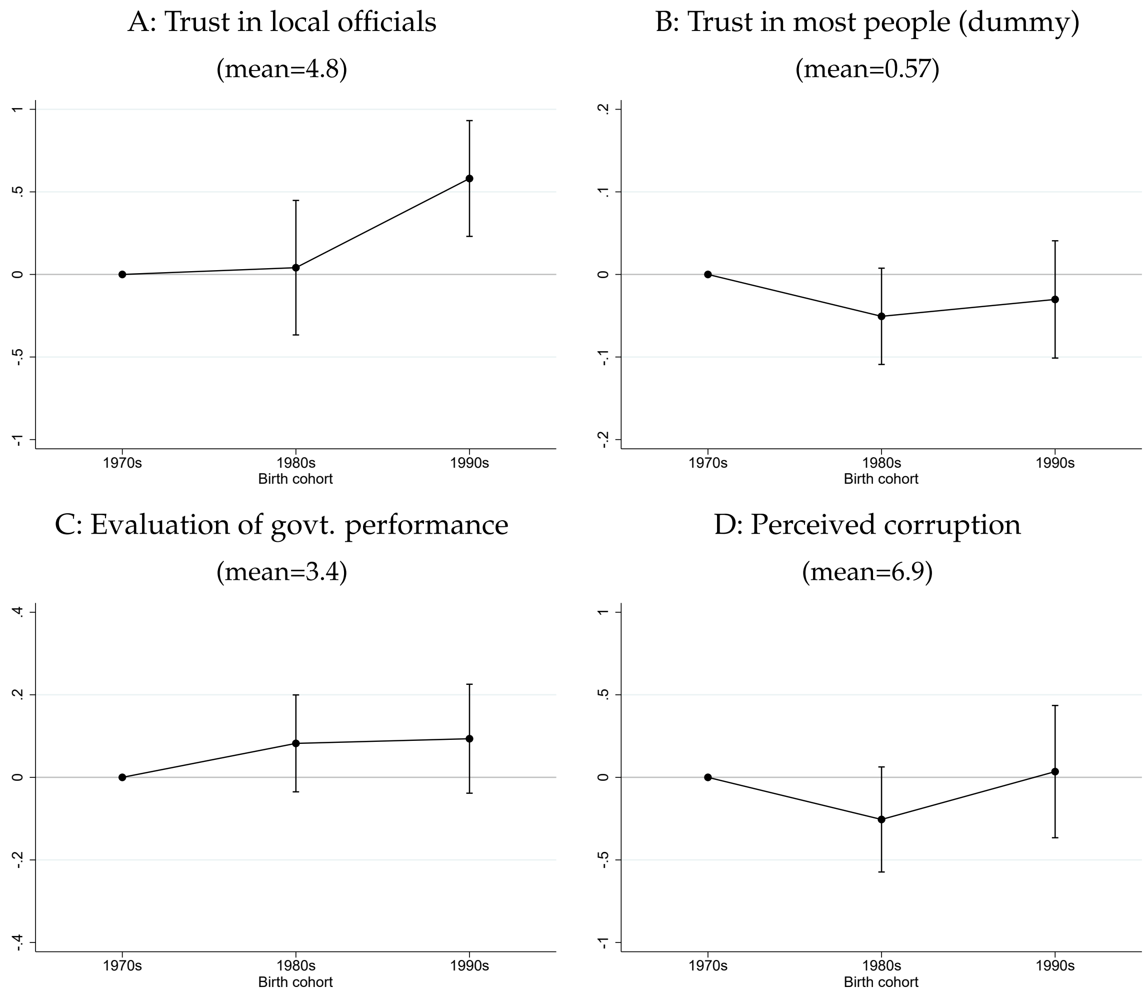
Notes: This figure is created by calculating the simple average total factor productivity (TFP) by ownership for manufacturing firms from 1998 to 2007 using the commonly used Annual Survery of Industrial Firms dataset and the Olley and Pakes method.

Figure A8: Dynamic effect on firm entry - 2SLS estimates



Notes: This figure is created by estimating event study variants of the specifications in Table 6, with $Treat_c^{1984}$ as instrument for $Treat_c$. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

Figure A9: Dynamic effect on citizen attitudes - 2SLS estimates



Notes: This figure is created by using $Treat_c^{1984}$ as an instrument for $Treat_c$ and reestimating the specifications in Table 7. Standard errors used to construct the 90% confidence intervals, denoted by the spikes, are clustered at the county level.

B Additional Tables

Table B1: Balance test

	Treat		Control		Difference		
	Mean	SD	Mean	SD	T - C	SE	*p*-value
			Panel A: Demography				
log Population (2010)	13.00	0.70	12.50	0.81	0.50	0.04	0.00
Share urban (%, 2010)	33.65	12.30	34.64	14.31	-0.99	0.66	0.13
Share 15-64 (%, 2010)	72.53	4.56	72.44	4.53	0.10	0.22	0.67
Years of schooling (2010)	8.23	0.73	8.19	1.05	0.04	0.05	0.34
			Panel B: Economic development				
Unemployment rate (%, 2010)	2.08	1.39	2.16	1.52	-0.09	0.07	0.23
Share primary sectors (%, 2010)	64.97	18.20	64.30	17.93	0.68	0.88	0.44
Share secondary sectors (%, 2010)	16.10	12.79	15.49	11.54	0.61	0.59	0.30
log GDP (2004)	12.47	0.94	11.97	1.03	0.50	0.05	0.00
log GDP (2008)	12.96	0.97	12.48	1.06	0.48	0.05	0.00
GDP growth (%, 2002-2004 average)	11.28	6.26	11.57	6.53	-0.29	0.33	0.38
GDP growth (%, 2006-2008 average)	12.86	6.38	12.67	6.63	0.19	0.32	0.54
Light growth (%, 2002-2004 average)	18.24	12.27	18.82	16.28	-0.58	0.72	0.42
Light growth (%, 2006-2008 average)	6.86	10.31	7.67	11.93	-0.81	0.55	0.14
Distance to major roads (km, 2010)	69.18	96.10	75.21	80.62	-6.03	4.24	0.16
Distance to major railways (km, 2010)	70.16	102.37	74.22	90.45	-4.06	4.63	0.38
			Panel C: Geography				
County area (km^2)	3900	7453	4128	10222	-228	446	0.61
Precipitation (inches, 2004)	0.04	0.09	0.03	0.10	0.00	0.00	0.55
Temperature (degrees, 2004)	13.91	5.23	13.57	5.41	0.34	0.26	0.19
Precipitation (inches, 2008)	0.05	0.11	0.05	0.11	0.01	0.01	0.20
Temperature (degrees, 2008)	13.73	5.08	13.43	5.25	0.30	0.25	0.23
Distance to major rivers (km)	59.17	61.15	57.76	59.03	1.41	2.90	0.63
Distance to country border (km)	346.52	251.00	345.25	251.49	1.26	12.16	0.92
Distance to coastline (km)	616.84	612.04	640.02	568.79	-23.18	28.46	0.42
Distance to prefecture center (km)	60.14	41.78	62.66	46.66	-2.52	2.17	0.25

Notes: This table provides balance tests by comparing counties with the survey teams deployed in 2005 to those without. The *p*-value reported in the last column is from a *t*-test of mean equality between groups.

Table B2: Balance test for IV

	Treat		Control		Difference		
	Mean	SD	Mean	SD	T - C	SE	*p*-value
			Panel A: Demography				
log Population (2010)	12.73	0.82	12.69	0.80	0.05	0.04	0.25
Share urban (%, 2010)	34.09	12.69	34.33	14.00	-0.24	0.68	0.72
Share 15-64 (%, 2010)	72.62	4.35	72.40	4.65	0.22	0.23	0.32
Years of schooling (2010)	8.24	0.82	8.19	0.99	0.05	0.05	0.30
			Panel B: Economic development				
Unemployment rate (%, 2010)	2.13	1.38	2.16	1.51	-0.03	0.07	0.64
Share primary sectors (%, 2010)	63.98	17.66	64.89	18.24	-0.91	0.90	0.31
Share secondary sectors (%, 2010)	15.92	11.90	15.64	12.14	0.28	0.60	0.64
log GDP (2004)	12.20	0.99	12.13	1.04	0.07	0.05	0.21
log GDP (2008)	12.71	1.01	12.63	1.08	0.07	0.05	0.16
GDP growth (%, 2002-2004 average)	11.26	6.32	11.54	6.50	-0.28	0.34	0.42
GDP growth (%, 2006-2008 average)	12.96	6.56	12.62	6.51	0.34	0.33	0.29
Light growth (%, 2002-2004 average)	18.67	14.60	18.67	15.04	0.00	0.74	1.00
Light growth (%, 2006-2008 average)	7.24	10.92	7.48	11.60	-0.24	0.57	0.67
Distance to major roads (km, 2010)	74.78	99.81	72.54	80.07	2.23	4.40	0.61
Distance to major railways (km, 2010)	77.02	105.61	71.13	90.01	5.89	4.80	0.22
			Panel C: Geography				
County area (km^2)	3883	7003	4160	10298	-277	463	0.55
Precipitation (inches, 2004)	0.030	0.083	0.035	0.095	-0.005	0.005	0.27
Temperature (degrees, 2004)	13.83	5.22	13.60	5.41	0.23	0.27	0.39
Precipitation (inches, 2008)	0.04	0.10	0.05	0.12	-0.01	0.01	0.14
Temperature (degrees, 2008)	13.66	5.07	13.45	5.26	0.21	0.26	0.43
Distance to major rivers (km)	58.51	59.50	58.40	60.42	0.11	3.00	0.97
Distance to country border (km)	340.46	252.42	348.81	250.34	-8.36	12.55	0.51
Distance to coastline (km)	632.09	618.89	631.84	571.17	0.25	29.48	0.99
Distance to prefecture center (km)	62.20	44.79	61.42	44.92	0.78	2.25	0.73

Notes: This table provides balance tests by comparing counties with the survey teams in 1984 (which serves as an IV) to those without. The *p*-value reported in the last column is from a *t*-test of mean equality between groups.

Table B3: Estimating spillover effect

	(1)	(2)	(3)	(4)
Dep. var.:	Reported GDP growth (%)			
Treat x Post	-0.576***	-0.580***	-0.565***	-0.585***
	(0.161)	(0.164)	(0.164)	(0.162)
#Treated neighbors x Post		-0.010		
		(0.071)		
1(#Treated neighbors>0) x Post			0.108	
			(0.263)	
1(#Treated neighbors>Median=2) x Post				-0.133
				(0.178)
Light growth (%)	0.017***	0.017***	0.017***	0.017***
	(0.005)	(0.005)	(0.005)	(0.005)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
County controls x Post	Yes	Yes	Yes	Yes
Neighbor number FE x Post	No	Yes	Yes	Yes
Cluster level	County	County	County	County
Mean dep. var.	10.84	10.84	10.84	10.84
Mean number of neighbors	5.91	5.91	5.91	5.91
Mean number of treated neighbors	1.97	1.97	1.97	1.97

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. #Treated neighbors denote the number of treatment counties among a county's neighbors, where neighbors are defined as counties sharing a common boundary segment with a county. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B4: Estimating spillover effect - robustness

	(1)	(2)	(3)	(4)
Dep. var.:	Reported GDP growth (%)			
Treat x Post	-0.610***	-0.559***	-0.581***	-0.576***
	(0.162)	(0.164)	(0.164)	(0.164)
#Treated neighbors within 50km x Post	-0.094			
	(0.091)			
#Treated neighbors within 100km x Post		0.027		
		(0.036)		
#Treated neighbors (GDP-weighted) x Post			-0.011	
			(0.068)	
#Treated neighbors (population-weighted) x Post				0.003
				(0.069)
Light growth (%)	0.017***	0.017***	0.017***	0.017***
	(0.005)	(0.005)	(0.005)	(0.005)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
County controls x Post	Yes	Yes	Yes	Yes
Neighbor number FE x Post	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County
Mean dep. var.	10.84	10.84	10.84	10.84
Mean number of neighbors	2.40	11.45	5.91	5.91
Mean number of treated neighbors	0.88	4.47	1.97	1.97

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. In columns (1) and (2), neighbors are defined as counties within a certain distance of a county, while in columns (3) and (4) neighbors are defined as counties sharing a common boundary segment with a county and are weighted by their sizes (GDP or population). Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: Robustness to weighting

	(1)	(2)	(3)	(4)
Dep. var.:	Reported GDP growth (%)			
Treat x Post	-0.804**	-0.586***	-0.727**	-0.589***
	(0.314)	(0.160)	(0.317)	(0.161)
Light growth (%)	0.023***	0.017***	0.024***	0.018***
	(0.005)	(0.005)	(0.005)	(0.005)
County controls x Post	No	Yes	No	Yes
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County
Observations	22,998	20,273	23,346	20,273
R-squared	0.268	0.362	0.271	0.365
Mean of dep. var.	10.98	10.84	10.97	10.84

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. Columns (1)-(2) weight the observations using population in 2010 and columns (3)-(4) weight the observations using GDP in 2008. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B6: Robustness to alternative clustering strategies

	(1)	(2)	(3)	(4)
Dep. var.:	Reported GDP growth (%)			
Treat x Post	-0.751**	-0.751**	-0.751**	-0.751**
	(0.312)	(0.347)	(0.336)	(0.335)
Light growth (%)	0.023***	0.023*	0.023***	0.023**
	(0.008)	(0.013)	(0.008)	(0.009)
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster level	Prefecture	Province	2° lon. x 2° lat.	4° lon. x 4° lat.
Number of clusters	311	26	216	74
Observations	23,360	23,360	23,360	23,360
R-squared	0.269	0.269	0.269	0.269
Mean of dep. var.	10.97	10.97	10.97	10.97

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. Standard errors in parentheses are clustered at the county level, the province level, the 2 ° longitude by 2 ° latitude level, and the 4 ° longitude by 4 ° latitude level in columns (1)-(4), respectively. * p <0.1, ** p <0.05, *** p <0.01.

Table B7: Testing the political selection channel

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. var.	1(Turnover)	Years in office	Age	1(Local)	Schooling	1(<Master)	1(Master)	1(PhD)
Treat x Post	-0.001	0.113	-0.075	-0.009	-0.157	0.006	0.003	-0.009
	(0.006)	(0.163)	(0.294)	(0.016)	(0.104)	(0.023)	(0.024)	(0.012)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County	County	County	County	County
Observations	22,695	22,695	10,674	8,243	7,994	7,994	7,994	7,994
R-squared	0.151	0.879	0.769	0.857	0.829	0.826	0.817	0.736
Mean of dep. var.	0.05	8.07	47.68	0.16	16.40	0.57	0.39	0.04

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B8: Testing the political selection channel (cont.)

	(1)	(2)	(3)
Dep. var.		Reported GDP growth (%)	
Treat x Post	-0.751**	-0.755**	-0.731**
	(0.316)	(0.315)	(0.315)
Treat x Post x Post-reform turnover number		0.192	-0.204
		(0.246)	(0.319)
Light growth (%)	0.023***	0.023***	0.022***
	(0.005)	(0.005)	(0.005)
Observations	23,360	23,360	23,346
R-squared	0.269	0.269	0.276
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Post-reform turnover number FE x Year FE	No	No	Yes
Cluster level	County	County	County
Mean of dep. var.	10.97	10.97	10.98
Mean of post-reform turnover number	0.50	0.50	0.49

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B9: Testing the soft information channel

	(1)	(2)	(3)
Dep. var.		Reported GDP growth (%)	
Treat x Post	-0.751**	-0.749**	-0.751**
	(0.316)	(0.317)	(0.317)
Treat x Post x Distance to upper-level govt.		-0.138	
		(0.292)	
Treat x Post x 1(Hometown ties)			0.082
			(3.117)
Light growth (%)	0.023***	0.023***	0.023***
	(0.005)	(0.005)	(0.005)
County FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Distance to upper-level govt. x Year FE	No	Yes	No
1(Hometown ties) x Year FE	No	No	Yes
Cluster level	County	County	County
Observations	23,360	23,239	23,360
R-squared	0.269	0.269	0.269
Mean of dep. var.	10.97	10.97	10.97

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. Distance to upper-level govt. denotes the shortest distance from a county to its overseeing upper-level government. 1(Hometown ties) is a dummy variable equal to 1 if the leaders of a county share the same hometown with that of the leaders in the upper-level government. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B10: Controlling for contemporary reforms

	(1)	(2)	(3)	(4)
Dep. var.:	Reported GDP growth (%)			
Treat x Post	-0.814***	-0.744**	-0.752**	-0.806**
	(0.313)	(0.316)	(0.316)	(0.313)
Light growth (%)	0.021***	0.023***	0.023***	0.022***
	(0.005)	(0.005)	(0.005)	(0.005)
Fiscal PMC	Yes	No	No	Yes
Full PMC	No	Yes	No	Yes
Anticorruption inspection	No	No	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County
Observations	23,360	23,360	23,360	23,360
R-squared	0.271	0.269	0.269	0.271
Mean of dep. var.	10.97	10.97	10.97	10.97

Notes: The unit of observation is county x year. The sample period is 2005-2018. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. Fiscal PMC is a dummy for the fiscal province-managing-county (PMC) reform. Full PMC is a dummy for the full province-managing-county (PMC) reform. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B11: Keywords in each topic

Topic	Chinese keywords	English translation
business attraction	招商引资	attract businesses
	外商直接投资	foreign direct investment
	外资	foreign capital
	对外开放	open up
infrastructure	基础设施	infracstrcture
	基建	shorthand term for infractrcture
	工程建设	project construction
	建设项目	construction project
market reform	改革	reform
	非公有制	non-public ownership
	民营企业	private firms
	私营企业	private firms [variant]
	民企	shorthand term for private firms
	私企	shorthand term for private firms [variant]
policy experimentation	试点	experimental places
	试验区	experimental zones
social welfare	社会保险	social insurance
	社保	shorthand term for social insurance
	养老保险	endowment insurance
	医疗保险	medical insurance
	养老金	pension
	社会保障	social security

Notes: This table lists the keywords in each topic, which is used to model governments' policy preferences.

Table B12: Effect on bank loans and deposits: only banks controlled by local governments

	(1)	(2)	(3)	(4)
		IHS transformation of		
Dep. var.:	Deposits	Loans	Loans to small firms	Number of firms granted loans
Treat x Post	0.203*	0.205*	0.292**	0.167**
	(0.123)	(0.119)	(0.126)	(0.085)
County controls x Post	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Cluster level	County	County	County	County
Observations	8,922	8,922	8,922	8,922
R-squared	0.323	0.328	0.512	0.574

Notes: The unit of observation is county x year. The sample period is 2006-2011. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. Post is a dummy variable indicating years after the reform in 2009. The dependent variable is transformed by inverse hyperbolic sine (IHS) to deal with zeros. Standard errors clustered at the county level are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B13: Effect on corruption convictions: robustness

	(1)	(2)	(3)	(4)
Type of corruption:	All	Other than appropriation and bribery	All after 2014	All bribery and appropriation after 2014
Dep. var.:	IHS (Number of convictions)			
Treat	0.052	0.079	0.016	-0.027
	(0.063)	(0.059)	(0.063)	(0.060)
#Anticorruption inspections	-0.022	-0.007	-0.009	-0.009
	(0.015)	(0.014)	(0.015)	(0.013)
County controls	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes
Observations	1,498	1,498	1,498	1,498
R-squared	0.098	0.031	0.070	0.067
First stage F-stat	2204	2204	2204	2204
Mean of dep. var. (raw)	4.648	1.344	2.933	1.997

Notes: The unit of observation is county. Treat is a dummy variable indicating counties with the survey teams deployed in 2005. The dependent variable is transformed by inverse hyperbolic sine (IHS) to deal with zeros. This table conducts 2SLS estimation with $Treat_c^{1984}$ (dummy for counties with the survey teams in 1984) as instrument for Treat. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.