The One Child Policy and Promotion of Mayors in China*

Juan Carlos Suárez Serrato
Duke University & NBER

Xiao Yu Wang
Duke University & NBER

Shuang Zhang†
University of Colorado Boulder

June 12, 2015

Abstract

We analyze the promotion of local officials in China based on the incentives to implement the One Child Policy. We develop a tournament model where provincial governors promote city mayors on the basis of a non-contractible, self-reported measure of performance (reported birth rate) that is subject to mayor manipulation. We show that if the objective of the central government is partially motivated by screening high ability mayors for higher office, then the performance measure will be predictive of promotion. Moreover, we show that OCP performance will have a smaller effect on promotion in areas which are more competitive, and where performance is more informative of ability. We test these predictions empirically and find that the promotion rule is consistent with a screening motive. Finally, we use a retrospective measure of performance and show that mayors are less likely to misreport their performance in years where population audits take place and that actual performance is less predictive of promotion than reported performance. These results show that, while mayor promotions were consistent with a meritocratic objective, misreporting on performance presented a significant obstacle to the selection of high ability leaders in China, contrary to the the extant belief that China's meritocratic system allowed the country to develop despite a lack of democratic accountability.

*We thank Douglas Almond, Kehinde Ajayi, Prashant Bharadwaj, Pat Bayer, Allan Collard-Wexler, Julie Cullen, Pascaline Dupas, Fred Finan, Rob Garlick, Gopi Goda, Hugh Macartney, Matt Masten, Michael Powell, Orie Shelef, and Daniel Xu for helpful comments. Juan Carlos Suárez Serrato and Shuang Zhang thank Stanford Institute for Economic Policy Research (SIEPR) for financial support. We thank Han Gao and Matt Panhans for providing excellent research assistance.

†Juan Carlos Suárez Serrato: Department of Economics, Duke University, and NBER, jc@jcsuarez.com. Xiao Yu Wang: Department of Economics, Duke University, and NBER, xy.wang@duke.edu. Shuang Zhang: Department of Economics, University of Colorado Boulder, shuang.zhang@colorado.edu.

"计划生育是天底下最难的一项任务!"

"Birth planning is the hardship number one under heaven!"

– A Chinese cadre (Scharping, 2003)

Meritocracy in the Chinese political system is often credited with contributing to China's rapid development, despite a lack of democratic institutions.[1] However, the efficacy of meritocratic systems in selecting high-ability leaders may be degraded by corruption through political connections and by manipulation of performace metrics used in promotion decisions.[2] This paper formally tests whether promotion decisions were consistent with a meritocratic objective (*i.e.,* a desire to screen for high ability leaders) by studying the promotions of Chinese mayors through the lens of the One Child Policy (OCP). We develop and test a model of mayor compensation when output is not contractible and mayor manipulation is a concern. The model generates generate sharp theoretical predictions which empirically distinguish between two contrasting objectives: meritocracy and the desire to control population size. In addition, we study mayor manipulation by using population auditing surveys, and we examine the degree to which misreporting of performance metrics affected the efficacy of the promotion mechanism to screen for high ability leaders.

We first develop a theory of optimal compensation and promotion of local Chinese officials that is inspired by our empirical setting. The model generates conditions under which a tournament mechanism is used by the provincial governments (principals) to incentivize unobserved mayoral effort (agents). We show that the principal's desire to promote the highest ability mayor, in addition to maximizing output production (stronger birth control performance relative to the targest set by the OCP), is a key determinant in adopting a tournament compensation mechanism. We characterize the equilibrium promotion rule and derive sharp empirical predictions which show that, if the principal is sufficiently motivated to promote the mayor with highest ability, then higher output will be less predictive of promotion when (1) the noise in the output measure is greater, and (2) when promotions are more competitive. A key feature of the model is that output is non-contractible; in our empirical setting, mayors self-report birth rates, and these self-reports are not perfectly verifiable. Our theory predicts the behavior of mayors facing audit risk and characterizes the capability of the promotion mechanism in selecting the highest-ability mayor. The model shows that, in order to interpret the empirical relation between performance and promotion, it is essential to recognize that the promotion rule is determined in equilibrium and, therefore, paints a more complete picture of the role of meritocracy in the Chinese government.

We connect the theory to the data by testing these predictions using a unique dataset on the performance and promotion of mayors in China from 1985-2000. Our first set of results tests the prediction that, if the principal places sufficient weight on the screening motive, improved birth

[1] Bell and Li. (2012) provide an overview of meritocracy in the Chinese political system from the point of view of political science. Economists have also studied determinants of cadre promotions, for example, in Maskin et al. (2000) and Li and Zhou (2005).

[2] For evidence on corruption and data manipulation see Shih et al. (2012), Jia et al. (ming), and Fisman et al. (2015)

control performance in implementing the OCP will be predictive of promotion.[3] We first report the correlation between reported OCP performance and promotion and find that mayors with better reported OCP performance are more likely to be promoted. A potential concern is that mayors' unobserved political connections could bias the estimation and lead us to find a spurious positive effect (Shih et al. (2012), Jia et al. (ming)). We deal with this concern in two ways. First, we assuage concerns of individual-specific sources of political connections by using the panel structure of the data and including mayor fixed effects, in addition to prefecture fixed effects, year fixed effects and prefecture-year characteristics.The promotion incentive for birth rate control is not diminished and remains economically and statistically significant. As an example, reducing the reported birth rate by 1 in 1000 people increases the probability of promotion by 10% and is equivalent, in its effect on promotion, to an 8% increase in GDP. We pursue a second identification strategy to isolate variation in OCP performance that is exogenous from time-varying political connections. We use changes in birth rate targets set by the central government as part of five-year plans as an instrument for OCP performance. We provide evidence for the validity of this strategy by showing that new province targets are not set to favor particular mayors, and that mayors are not strategically promoted to avoid future changes in birth rate targets. This identification strategy shows that plausibly exogenous changes in OCP performance have statistically and economically significant effects on the promotion of mayors.

Next, we test the model's comparative static predictions of the effects of different informational and competitive settings on a meritocratic promotion rule. First, we show that OCP performance is more predictive of promotion in provinces where the variance of birth rates is smaller. That is, when the signal is more informative of a mayor's ability, the principal gives a higher weight to OCP performance in the promotion rule. Second, we show that OCP performance is less predictive of promotion in years when promotions are more competitive. The intuition for this result is that, when mayoral effort is increased in response to competition for fewer openings, OCP performance is less informative of ability. This leads the principal to place less weight on OCP performance in the promotion rule. These results provide further evidence that, beyond implementing the OCP, promotion incentives were used to screen for high-ability mayors. Our paper thus builds on the strong but small empirical contract literature using an unusual and interesting setting of analysis: promotions within provincial Chinese governments.

Finally, we study whether the incentive system is successful in screening high-ability mayors. Since mayors are evaluated on the basis of self-reported data, mayors may misreport data. Our theory makes two predictions on this front. First, when the probability of detection of misreporting increases, mayoral effort (and actual output) increases and misreporting decreases. We use population auditing surveys and employ two different measures of birth rate to analyze potential cheating behavior. The first measure is the reported birth rate and the second is the birth rate measured in census microdata. We find that mayors adjust their manipulation of birth rate data in response to these economic incentives; we show that reported birth rates are higher and that mayors are less likely to report a lower birth rate in years when audits are conducted. Second, our model shows

[3] We measure mayors' performance in implementing the OCP by the gap between centrally-set birth rate target and the local birth rate.

that the non-contractibility of OCP performance lessens the ability of the screening mechanism to select for high-ability mayors: the expected ability of the promoted mayor is a combination of the ability of a random promotion scheme and a scheme where OCP performance is contractible. We find empirical support for this prediction: OCP performance from census birth rates is not predictive of promotion, suggesting that promoted mayors do not have higher ability than mayors who are not promoted. This finding also corroborates the result that reported birth rate data are likely manipulated. Rather than attributing these results to corruption or to a lack of meritocracy in the Chinese government, we argue that the non-verifiability of the birth rate forces the provincial government to use a promotion rule that is equivalent to weighting random promotions with positive probability, where this weight increases as the verifiability problem worsens. Overall, our results form a counterpoint to the argument that the success of the Chinese authoritarian government can be attributed to the central role played by meritocratic promotions (Bell and Li. (2012)).

Our paper contributes to the broad literature on the design of incentives in uncertain environments by using evidence from China.[4] While a rich literature in contract theory explores a spectrum of incentive structures when incentives must be based on a noisy signals of effort or ability,[5] the body of empirical evidence testing these theoretical mechanisms is small.[6] Whether screening mechanisms successfully select high ability agents for promotion is still an open question. For example, recent studies find empirical evidence of incentive distortions, including teacher manipulation of test scores (Jacob and Levitt (2003)) and of student composition in test-taking pools (Cullen and Reback (2006)). This paper provides evidence that when performance measures can be manipulated, the efficacy of the screening mechanism is weakened in equilibrium. This finding bridges personnel and political economics with forensic economics, a literature which uncovers evidence of hidden behaviors and corruption, and which studies the role of audits in limiting corruption.[7]

In addition, this paper develops a unified framework for understanding the mechanisms and screening capacity of promotion incentives in the Chinese government. A large literature suggests that promotion incentives play an important role in explaining China's rapid economic growth despite its poor institutions (Maskin et al. (2000), Li and Zhou (2005), Landry (2008), and Xu (2011)). Recently, a few studies have argued that political connections are more important in promotion than performance, that they complement the role of performance, or that ties to localities and local elites may influence the policies of promoted officials (Shih et al. (2012), Jia et al. (ming), and Persson (ming)). Furthermore, a growing literature finds evidence on data manipulation of local governments in reporting accidental deaths (Fisman et al. (2015)) and air pollution (Chen et al.

[4]See Oyer and Schaefer (2011) for a review of incentives in personnel economics.

[5]These structures range from piece rates to tournaments, and the literature examines their consequences in a wide variety of environments, ranging from firms to governments. See Lazear and Rosen (1981), Baker et al. (1994), Holmstrom (1999), Rochet and Stole (2003), for example.

[6]See Baker et al. (1994), Chevalier and Ellison (1999), Prendergast (1999), Chiappori and Salanie (2000),Bertrand and Mullainathan (2001), Kane et al. (2002), Chay et al. (2005), Rockoff et al. (2012), for example.

[7]See Zitzewitz (2012) for a review of recent papers in forensic economics. For example, Fisman and Wei (2004) finds that the difference between shipments reported as exports to Hong Kong but not as imports to China is larger for products where Chinese tariff are higher and smaller for products with high tariff on closely related products. Our paper provides new evidence on wrongdoing behaviors among Chinese mayors: that mayors are less likely to underreport birth rate when the cost of cheating is higher in audit years. In using audits to uncover hidden behaviors, our paper is also related to studies using audits to detect corruption (Olken (2006), Ferraz and Finan (2008), Ferraz and Finan (2011), Bobonis et al. (2013)).

(2012), Ghanem and Zhang (2014)). Our findings suggest that the OCP may have persisted for as long as it did not because the government inherently cared about lowering the fertility rate, but because reported success at reducing the fertility rate was believed to be informative of mayors' ability. However, the effectiveness of promoting based on OCP at selecting high-ability mayors was severely dampened by mayor manipulation. To our knowledge, the role played by the OCP in screening for high ability leaders has not been explored by other researchers. Our findings suggest that the OCP may be an unexpected and hitherto unexplored example of a classic principal-agent theory relating the noisiness of contractible performance measures with incentives and promotion. More broadly, this study sheds lights on the competing roles played by encouraging production of output and extracting information in other important policies around the world.

Finally, our paper contributes to the literature on the consequences of the One Child Policy (see, *e.g.*, Ebenstein (2010), Wei and Zhang (2011), Choukhmane et al. (2014)). We provide new evidence of China's enforcement of the world's largest population control policy at the local level. Finding that mayors manipulate the reported birth rate data suggests that the local implementation of the OCP is an important component in their evaluation. However, the marginal effect of the promotion incentive on the enforcement of birth rate control in equilibrium may be minimal in the presence of mayor manipulation.

The rest of the paper is organized as follows. Section 1 describes institutional details of the OCP and its implementation. Section 2 develops a model of optimal compensation for mayors and discusses testable predictions. We describe our data in Section 3. Section 4 tests the theoretical predictions on the promotion rule and Section 5 provides evidence of data manipulation in response to audit risk and analyzes the ability of the promotion mechanism to select high-ability mayors. Section 6 concludes by discussing the role of manipulation in analyzing equilibrium promotion rules.

1 Institutional Details of the One Child Policy

In 1979, soon after China's Cultural Revolution and after a decade-long economic crisis, Deng Xiaoping expressed the fear that "without birth planning, economic growth would be consumed by population growth." Since then, all economic planning has therefore presupposed success in population control. At the national level, a specific target on population growth was set so that the total population would not exceed 1.2 billion in 2000. Chinese scientists working for the government further developed a projection that showed that, in order to achieve the population target, the optimal fertility level should be one child per woman (Scharping (2003)). This recommendation was incorporated into the family planning policy in the same year and the policy was thereafter known in the West as the One Child Policy (OCP).

Under the OCP, a limit of one child per family is strictly enforced in urban areas, and second-child permits are issued for special exemptions in rural areas and for ethnic minorities. Some exemptions are granted, for example, to couples who are disabled or who live in remote areas. Provinces with a tight policy restrict themselves to common norms for exemptions, while regions with a more relaxed policy may include other criteria. The national policy was relaxed in 1984 to allow rural couples to have a second child if their first-born is a girl.

1.1 Enforcement Mechanisms

A variety of birth control methods have been used to enforce the OCP. Sterilization and insertion of an intrauterine device (IUD) after the first or second birth were implemented on a mass scale. Between 1979 and 1999, the percentage of women of reproductive age who underwent sterilization rose from 21% to 35% (Scharping (2003)). Meanwhile, induced abortions of unauthorized pregnancies have been used as a "remedial measure for making up contraceptive failures." For above-quota births, financial sanctions are the main instrument to enforce the OCP. Depending on the location and time period, the birth of an extra child could cost the family 10%-25% of their annual income for 7-14 years. Other punishments widely used included denial of bonus payments, health and welfare benefits, denials of job promotions or even demotions in urban work units, as well as confiscation of family farm-land in rural areas.

Strong resistance and non-compliance at the grassroots level, especially in rural areas, made it very challenging to enforce the OCP. As documented by Scharping (2003), internal reports issued within the party in the 1980s and 1990s acknowledged that assaults to local birth-planning cadres were frequently provoked by coercion in regard to abortion, sterilizations, and the administration of penalties.

1.2 Mayoral Promotions and the OCP

The central government controls appointments, evaluation, promotion, and demotion of subnational officials in China, and the career paths of these officials are determined by the performance of their jurisdictions (Xu (2011)). The central government directly controls the key positions at the province level and grants the provincial government the power to appoint key officials at the prefecture level. The provincial government stipulates a set of performance criteria for mayors. Economic growth, social stability, and enforcement of the One Child Policy are consistently among the highest priorities (Birney (2014)). In a published list of performance indicators of 104 prefectures in 2000, GDP per capita was used to evaluate economic growth and birth rate was used to evaluate the enforcement of the OCP (Landry (2008)).

A centrally controlled planning system has been developed to monitor the local enforcement of the OCP since the 1980s. At the highest level, the State Planning Commission sets birth plan targets as part of the five-year plans with the goal of meeting the national population goal of 1.2 billion by year 2000. Annual targets on population and birth rate are set at the province level every five years. Only national and provincial targets are set; they are sent to prefectures and further distributed to lower levels as task assignments (Hardee-Cleaveland and Banister (1988)). Prefectures are responsible for local implementation and submit the population and birth data to provinces. Provinces further transmit the numbers to the central government.

Birth control performance is directly linked to cadre evaluation. Hardee-Cleaveland and Banister (1988) document that,

> In Shaanxi in March 1987, the provincial party committee deputy secretary and acting governor "demanded that leaders at all levels should simultaneously grasp two kinds of production - economic production and reproduction - and take measures to do this work

well and firmly. Otherwise, they are not qualified leaders."

1.3 Misreporting in Birth Rate Data

In addition to the birth rate, the birth planning rate, as measured by the percentage of authorized births of the total number of births, is also reported from lower level governments. However, the birth planning rate is not used in the evaluation of mayor performance for two reasons. First, because birth-planning regulations and permits for second and additional births vary across localities, the rate does not indicate the degree to which the policy target is realized. Second, in practice, lower level authorities often report very high birth-planning rates of 98-99%, which are utterly unrealistic and unreliable (Scharping (2003)).

The leadership has been aware that, even for birth rate numbers, there are foreseeable conflicts that impair data quality since the data are reported by officials whose evaluations depend on these data. Population censuses provide the ideal data to be systematically compare with the reported numbers, but they are only conducted approximately every ten years (1982, 1990 and 2000). A mini census for 0.1% of population was conducted in 1995. To further investigate the credibility of reported birth numbers, the State Birth-Planning Commission was charged with conducting national fertility surveys for 0.1% to 0.2% of the population in 1988, 1992, and 1997. These census and national fertility surveys were organized at the province level, where they serve as the main instrument for data validation. As an example, the 1992 fertility survey uncovered an underreporting of 18% in reported birth rates. A particularly striking case of underreporting was found in Guangxi province and the leadership was forced to deliver a written self-criticism. (Scharping (2003)).

Figure 1 displays an official document from Fujian province that links OCP performance to promotion outcomes and details guidelines for local officials with respect to the implementation of the OCP. The first highlighted section states that local officials are responsible for reporting accurate birth rates and other OCP statistics. The second highlighted section states that local officials should ensure the accuracy of the reported numbers and avoid underreporting, misreporting, faking, and failing to report birth rate statistics. Finally, the third highlighted section states that the province government is responsible for investigating violations of these guidelines. If these guidelines are violated, the responsible officials are denied positive credits in their annual evaluation and their records are sent to the personnel department of the province government.

2 A Tournament Model with Non-Contractible Output

This section establishes the theoretical setting, determines the optimal compensation scheme used by the principal, and characterizes properties of the promotion rule as an equilibrium object. The principal maximizes a weighted sum of total output produced and the ability of the agent she promotes.[8] The model generates empirically-testable predictions on the responsiveness of the equilibrium probability of promotion to actual and reported output that depend on the weight that the principal places on the meritocratic objective. We first consider the case where the principal's

[8] Output is not contractible and results partly on the actions taken by agents of heterogeneous ability.

objective function depends purely on total output produced.[9] We then consider the case where the principal also cares about the birth rate achieved by a mayor because it is informative of mayoral ability.[10] The model's predictions allow the econometrician to distinguish between these two cases by studying the responsiveness of the equilibrium probability of promotion as well as comparative statics across noisier and more competitive environments. Our final set of results characterize the degree to which the non-contractibility of output affects the quality of promotion on birth rates as a screening mechanism for ability.

2.1 The Model

Consider a risk-neutral principal and $N \in \{2, 3, 4, ...\}$ risk-neutral agents. Utility for all agents is described by $u(x) = x,\ x \in$. In our setting, the principal is the provincial governor and the agents are the mayors competing for promotion.

The principal maximizes a weighted sum of total output produced and the expected ability, a_i, of the promoted mayor by choosing how to compensate agents for the output, y_i, they produce, as well as which agent i to promote. The principal's objective is written as:

$$\omega E\left[\sum_{i=1}^{N} y_i\right] + (1-\omega)E[a_i | i \textit{ is promoted}]. \tag{1}$$

In our setting, output y_i is the birth rate achieved by mayor i.[11] The model has the following elements:

Output Production vs. Screening Priority: ω captures how much the principal values maximizing total output produced relative to promoting the highest ability mayor. If $\omega = 1$, the principal is indifferent between promoting any of the mayors and focuses on maximizing output. If $\omega = 0$, the principal is indifferent between all levels of output production and focuses on promoting the agent with the highest ability. ω is determined by the central government in our empirical setting and is held constant across regions.

Production: $y_i = e_i + \varepsilon_i$ is the output produced by agent i, where e_i is the unobservable/non-contractible effort exerted by agent i, and $\varepsilon_i \sim \exp(\lambda)$ is noise with mean $E(\varepsilon) = \frac{1}{\lambda}$, variance $V(\varepsilon) = \frac{1}{\lambda^2}$, and is iid across agents.[12]

Moral Hazard: Effort cost to agent a_i is $c(e_i) = \frac{1}{a_i}\exp(e_i)$ for effort level e_i, where a_i is the privately-known ability of agent i. Assume for notational convenience that $a_1 > a_2 > ... > a_N \geq 1$; an agent with a higher i index has lower ability. The principal and the agents know the distribution of abilities in the economy. While the set of abilities $\{a_1, ..., a_N\}$ is known, the principal does not know which ability level "belongs" to which agent. These assumptions ensure that a given level of effort is less costly for higher ability agents, that higher levels of effort are more costly for all agents, and that effort cost increases at an increasing rate.[13]

[9] That is, the provincial governor sets incentives for the birth rate caring only about population control.

[10] Mayoral ability is private information. In this case, output has screening as well as production value.

[11] The principal prefers a lower birth rate, and agents must exert more effort to achieve a lower birth rate. "Higher y is better" is standard, but in the case where output is the birth rate b, we can define output $y \equiv \frac{1}{b}$.

[12] Assume that $\lambda \in (0, 1)$, a parametric assumption for objects to be well-defined: this ensures that $V(\varepsilon) \in (1, \infty)$.

[13] Note the following properties of $c(e_i)$: $\frac{\partial c(e)}{\partial a} < 0$, $\frac{\partial c(e)}{\partial e} > 0$, $\frac{\partial^2 c(e)}{\partial e^2} > 0$.

Lie Detection: The principal cannot directly contract on the true output produced by the mayors. The principal can only compensate based on mayors' reports $\hat{y}_i$ of how much they actually produced (y_i).[14] The principal audits each mayor after they submit private output reports and detects that a mayor is lying with probability:

$$\Pr(a_i \text{ is caught} \mid a_i \text{ is lying}) = \begin{cases} p, & \frac{\hat{y}_i}{y_i} \leq \delta \\ 1, & \frac{\hat{y}_i}{y_i} > \delta \end{cases}, \tag{2}$$

where $\delta \geq 1$ and $p \in [0, 1]$. That is, if agent a_i exaggerates her actual output production too much (beyond δy_i), she will get caught for sure; but if she slightly overreports, she will be caught with some intermediate probability p. The strength of the audit is described both by p and δ; stronger auditing is captured by a higher probability of catching lies (higher p) and by how much an agent can overreport before being caught with certainty (δ closer to 1).[15] Either $\delta = 1$ or $p = 1$ capture the case where output is contractible. Mayors caught misreporting are fired and suffer disutility $F << 0$.

Compensation: The principal chooses between two compensation schemes. The principal can promote an agent based on reported outputs in a *tournament* or the principal can pay a *piece rate*.

Timing: The timing of the game is as follows:

1. Each mayor observes her own private ability a_i and chooses effort level e_i, which is non-observable and non-contractible by the principal.

2. Each mayor's output y_i is realized and is observed by each mayor. Output remains private information.

3. Each mayor submits a private report of output to the principal: $\hat{y}_i$.

4. The principal audits each mayor. If the mayor reported $\hat{y}_i = y_i$, she is truthful and will not be wrongfully charged with lying. If the mayor reports $\hat{y}_i > y_i$, the principal detects the manipulation with the probability described in Equation 2; mayors caught lying are fired and receive $F < 0$.

5. The principal compensates mayors who are not fired.

Equilibria: There are three types of pure strategy equilibria:

1. "Pure lie": all the mayors misreport their output.

2. "Pure truth": all the mayors truthfully report their output.

3. "Partial truth": some mayors misreport and some mayors report truthfully.

We focus on the first equilibrium ("pure lie") as we find it most empirically relevant. We discuss the other two equilibria in Appendix A.6.

[14] See Appendix A.5 for the case where output is contractible.

[15] If output is the birth rate, so that $y = \frac{1}{b}$, then the agent reports $\frac{1}{\hat{b}_i}$, and is caught lying with probability 1 if $\frac{\frac{1}{\hat{b}_i}}{\frac{1}{b_i}} = \frac{b_i}{\hat{b}_i} > \delta$, that is, if $\hat{b}_i < \frac{1}{\delta} b_i$: the agent tries to report achieving a birth rate much lower than she actually achieved.

2.2 Characteristics of the Equilibrium Promotion Rule

The following sections describe the properties of the "pure lie" equilibrium of the model under different preferences of the principal. We first derive a condition on the principal's objective such that tournaments are preferable to piece-rate compensation mechanism. We then characterize the relationship between OCP performance and probability of promotion as well as how this relationship differs under different degrees of competition between agents and across different degrees of noise in the performance metric. The last result characterizes the ability of the tournament model to screen high ability mayors in the presence of manipulated reported output.

The Optimal Compensation Scheme

We first characterize the compensation scheme set by the principal.

Proposition 1. *There exists an $\tilde{\omega} \in (0, 1)$ such that, if $\omega > \tilde{\omega}$, (the principal values output for production relatively more than for screening) the principal uses a piece-rate mechanism and, if $\omega < \tilde{\omega}$ (the principal values output for screening relatively more than for production), the tournament mechanism is optimal.*

Please see Appendix A.2 for a detailed proof.

The intuition behind this result is the following. When the principal's sole concern is output maximization, continuous incentive pressure will exert maximum effort from all agents at a lower cost to the principal. Tournaments have a screening advantage since agents with lesser ability have little chance of winning the tournament and respond by reducing their effort. Effort separation reduces total output but is valuable if the principal values using output to make inferences regarding agents' abilities. When the principal sufficiently values promoting the highest ability agent, the tournament mechanism becomes optimal as it induces bigger differences in effort exerted between agents of higher versus lower abilities given any cost level, thereby increasing the expected ability of the promoted agent. Nonetheless, the manipulability of output limits this information revelation as the principal can only contract on reported output and detects lies imperfectly.

To prove this result, we solve for the first-best effort allocation in our model with misreporting given any $\omega \in [0, 1]$. Total social welfare given ω is $V_\omega = \omega V_1 + (1 - \omega)V_0$, where V_1 denotes total social welfare when $\omega = 1$ and the principal cares purely about maximizing output production, and V_0denotes total social welfare when $\omega = 0$ and the principal cares purely about promoting the highest-ability agent. We show that $V_1^{PR} > V_1^T$, that is, the piece-rate implements the first-best more efficiently than the tournament when the principal cares only about maximizing total output. Next, we show that $V_0^{PR} < V_0^T$, that is, the tournament is more efficient than the piece-rate when the principal cares only about promoting the highest-ability agent ($\omega = 0$). Since $V_\omega$is continuous and increasing in ω, it follows that a value $\tilde{\omega} \in (0, 1)$ exists such that, when $\omega > \tilde{\omega}$, $V_\omega^{PR} > V_\omega^T$ and the principal uses the piece-rate, and when $\omega < \tilde{\omega}$, $V_\omega^T > V_\omega^{PR}$ and the principal uses the tournament.

Equilibrium Properties with Production Objective ($\omega = 1$)

According to Proposition 1, when the principal's sole concern is output production, the principal compensates output (lower birth rate) with a piece-rate and, therefore, the promotion rule is independent of OCP performance.

Proposition 2. *If $\omega = 1$, agents are compensated with a piece-rate and, conditional on agents' abilities, increasing output does not increase mayor a_i's probability of promotion. This is true regardless of the noise (λ) and the competitiveness (N) of the environment.*

Proposition 2 guides the econometrician to compare promotion outcomes for agents with higher and lower reported output. However, the results of Proposition 2 no longer hold whenever this comparison is not conditional on a given agent's ability.

Corollary 1. *If $\omega = 1$ and the principal bases promotions on a dimension other than OCP performance that is positively correlated with ability, agents are compensated via a piece-rate and, unconditional on agents' abilities:*

(i) Increasing output increases mayor a_i's probability of promotion.

(ii) Increasing output has a larger effect (increase) on the probability of promotion in noisier environments (smaller λ).

(iii) Increasing output does not have a differential effect (increase) on the probability of promotion in more competitive environments (larger N).

The proofs of Proposition 2 and Corollary 1 are presented in Appendix A.3.

The intuition for (ii) is as follows. We know from Proposition 2 that the principal compensates agents with a piece-rate when ω =1. Given this, it can be shown that equilibrium effort increases more sharply in ability in noisier environments. This is because, in noisier environments, the principal sets a lower piece-rate (the slope of the wage in reported output is flatter). Lower ability agents decrease their effort differentially more than higher ability agents in response to this lower piece-rate. Since ability is positively correlated with the dimension on which promotion is based, it must be that the impact of having higher ability on the probability of promotion in noisier environments is larger. But we know that ability is also positively correlated with output. Hence, it must be that the observed impact of increasing output on probability of promotion in noisier environments is also larger. Corollary 1 further guides the econometrician who is not be able to separate variation in agent output from variation in agent ability.

Equilibrium Properties with Screening and Production Objective ($\omega < 1$)

Proposition 3. *If $\omega < \tilde{\omega}$, agents are compensated with a bonus in a tournament where the agent with the highest reported output who is not caught lying is promoted. Further, in the equilibrium where all mayors misreport:*

(i) Increasing output increases mayor a_i's probability of promotion

(ii) Increasing output has a larger effect (increase) on the probability of promotion in less noisy environments (larger λ)

(iii) Increasing output has a smaller effect (increase) on the probability of promotion in more competitive environments (larger N)

We present a sketch of the proof to build intuition. The technical details of the proof of Proposition 3 are presented in Appendix A.4.

We know from Proposition 1 that the principal incentivizes the production of output y by awarding a bonus $B > 0$ to the agent with the highest self-reported output who is not caught lying in the audit. Note that if a mayor does choose to lie, her optimal lie is $\hat{y}_i = \delta y_i$. Although the structure of misreporting is the same across agents, the actual reports will be heterogeneous, since y_i is heterogeneous. That is, there exists an optimal degree of inflation which is independent of the individual but the level of the lie will vary by individual.

We now solve for the equilibrium effort of mayor a_i. The mayor solves this problem:

$$\max_{e_i} B \Pr(\hat{y_i} > \hat{y_{-i}}) - \frac{1}{a_i}\exp(e_i).$$

Expected utility for agent a_i is approximated by (see Appendix A.1 for details):

$$\begin{aligned} EU_{i\ lies}^{-i\ lies} \ \simeq \ & pF + (1-p)\exp\left(-\lambda\left[\frac{1}{N-1}\sum_{j\neq i} e_j - e_i\right]\right)E(p,N)B \\ & +(1-p)p^{N-1}B - \frac{1}{a_i}\exp(e_i), \end{aligned} \tag{3}$$

where $E(p,N)$ denotes the expected probability a_i's error is weakly greater than the maximal order statistic for the error in the population of non-fired mayors:

$$\begin{aligned} E(p,N) \ &\equiv \ \left[(1-p)^{N-1}\exp(-\lambda\bar{\varepsilon}_{N-1}) + ... + (1-p)p^{N-2}\exp(-\lambda\bar{\varepsilon}_1)\right] \\ &= \left[(1-p)^{N-1}\exp\left(-\sum_{j=1}^{N-1}\frac{1}{j}\right) + ... + (1-p)p^{N-2}\exp(-1)\right]. \end{aligned}$$

Agent a_i's expected utility in Equation 3 breaks down in an intuitive way:

1. The first term captures the loss from being fired (F) when a_i gets caught lying, which happens with probability p.

2. The second term is the most complex; it captures the gain from promotion (B) in all the cases where a_i does not get caught but various subsets of the other $(N-1)$ lying mayors are caught. All possibilities ranging from "none of the other mayors is caught" to "all but one of the other mayors are caught" are addressed in this term.

   The key observation is that the average effort of the non-detected mayors is always $\frac{1}{N-1}\sum_{j\neq i} e_j$ regardless of how many of the other mayors are detected. This is due to the constant probability of detection. A simple example will illustrate. Suppose there are four mayors, a_1, a_2, a_3, a_4. The mayor a_1 calculates the average effort of the pool of non-fired mayors she will face, in the case that one of the other mayors is caught. This means that she might face $\{a_2, a_3\}$, or $\{a_2, a_4\}$, or $\{a_3, a_4\}$. But she faces each of these pools with equal probability. Thus, the average of the average effort in each of these pools is just $\frac{1}{3}a_2 + \frac{1}{3}a_3 + \frac{1}{3}a_4$–but that is just the average ability of the three other mayors. The expected utility from these contingencies is simplified by factoring $e^{-\lambda\left[\frac{1}{N-1}\sum_{j\neq i} e_j - e_i\right]}$ in Equation 3.[16]

3. The third term addresses the case where mayor a_i is not caught, but all the other mayors are caught. In this case a_i is promoted for sure.

4. The fourth term is the cost of effort to a_i from exerting effort e_i.

Then the first-order condition characterizing optimal effort in the "pure lie" scenario is:

$$FOC_{e_i} : (1-p)B\lambda e^{-\lambda\left[\frac{1}{N-1}\sum_{j\neq i} e_j - e_i\right]} \left[(1-p)^{N-1}e^{-\lambda\bar{\varepsilon}_{N-1}} + ... + (1-p)p^{N-2}e^{-\lambda\bar{\varepsilon}_1}\right] = \frac{1}{a_i}e^{e_i}.$$

Solving yields equilibrium effort:

$$\begin{aligned} e_i(i\ lies, others\ lie) &= \frac{(N-1)}{N(1-\lambda)-1}\log(a_i) - \frac{\lambda}{N(1-\lambda)-1}\sum_{j=1}^{N}\log(a_j) \\ &\quad + \log[(1-p)B\lambda] + \log E(p, N). \end{aligned} \tag{4}$$

The principal foresees the agents' choices and sets B to maximize Equation 1:

$$\begin{aligned} &\max_B \omega \left[\sum_{j=1}^{N} e_i(i\ lies, others\ lie) + \frac{N}{\lambda}\right] \\ &+(1-\omega)E[a_i|i\ is\ promoted] - c(B), \end{aligned}$$

where we substituted equilibrium effort from Equation 4. The principal's first-order condition describes the optimal bonus:

$$FOC_B : \omega\frac{N}{(1-p)B\lambda}(1-p)\lambda - c'(B) = 0.$$

Assuming that $c(B) = B$, the optimal bonus is $B^* = \omega N$; it is increasing in both ω, the principal's relative valuation of output production, and N, the degree of competition between mayors vying for promotion.

We now describe conditions under which an "all lie" equilibrium can be sustained (see Appendix A.6 for technical details). Suppose that mayor a_i decides to report truthfully anticipating that all other mayors will lie. Equation 3 simplifies to:

$$\max_{e_i} \exp\left(-\lambda\left[\frac{1}{N-1}\sum_{j\neq i}\delta e_j - e_i\right]\right) E(p, \lambda, N)B + p^{N-1}B - \frac{1}{a_i}e^{e_i}.$$

[16] Recall that $\bar{\varepsilon}_k$ is the maximal order statistic for a sample of k iid draws from the error distribution $\exp(\lambda)$ and that $\bar{\varepsilon}_k = \left(1 + ... + \frac{1}{k}\right)\frac{1}{\lambda}$.

We solve for the equilibrium effort exerted by a_i when she anticipates that the other agents will all exert $\{e_j\}_{j\neq i}$ and misreport, but she tells the truth:

$$e_{i\ truth}^{others\ lie} = \frac{1}{1-\lambda}\left[\log(a_i\lambda) + \log(B) + \log(E(p,N)) - \frac{\lambda}{N-1}\sum_{j\neq i}\delta e_j\right],$$

while the effort she exerts when she also chooses to misreport is:

$$e_{i\ lie}^{others\ lie} = \frac{1}{1-\lambda}\left[\log(a_i\lambda(1-p)) + \log(B) + \log(E(p,N)) - \frac{\lambda}{N-1}\sum_{j\neq i}\delta e_j\right].$$

Note that $e_{i\ lie}^{others\ lie} < e_{i\ truth}^{others\ lie}$, since the only difference is the $(1-p) < 1$ in the first term of the expression characterizing $a_i's$ effort when she also misreports. Then, her expected utility when she chooses to be truthful and her expected utility when she also chooses to misreport are described by:

$$\begin{aligned} EU_{i\ truth}^{others\ lie} &= e^{-\frac{\lambda}{1-\lambda}\frac{1}{N-1}\sum_{j\neq i}\delta e_j} a_i^{\frac{\lambda}{1-\lambda}} \lambda^{\frac{\lambda}{1-\lambda}} E(p,N)^{\frac{1}{1-\lambda}} B^{\frac{1}{1-\lambda}}(1-\lambda) + p^{N-1}B, \\ EU_{i\ lie}^{others\ lie} &= e^{-\frac{\lambda}{1-\lambda}\frac{1}{N-1}\sum_{j\neq i} e_j} a_i^{\frac{\lambda}{1-\lambda}} \lambda^{\frac{\lambda}{1-\lambda}} E(p,N)^{\frac{1}{1-\lambda}} B^{\frac{1}{1-\lambda}}(1-\lambda) + (1-p)p^{N-1}B + pF. \end{aligned}$$

An "all lie" equilibrium is maintained when $EU_{i\ lie}^{others\ lie} > EU_{i\ truth}^{others\ lie}$ for every i:

$$p^N B - pF < \left[(1-p)^{\frac{1}{1-\lambda}} e^{-\frac{\lambda}{1-\lambda}\frac{1}{N-1}\sum_{j\neq i} e_j} - e^{-\frac{\lambda}{1-\lambda}\frac{1}{N-1}\sum_{j\neq i}\delta e_j}\right] a_i^{\frac{\lambda}{1-\lambda}} \lambda^{\frac{\lambda}{1-\lambda}} E(p,N)^{\frac{1}{1-\lambda}} B^{\frac{1}{1-\lambda}}(1-\lambda). \tag{5}$$

Note that the right-hand side is monotonic in a_i. When the bracketed term is negative, the right-hand side is decreasing in a_i, and the highest-ability agents are the first to prefer to tell the truth. Thus, an "all lie" equilibrium is maintained when even the highest-ability agent prefers to tell the truth. When the bracketed term is positive, the right-hand side is increasing in a_i, and the lowest-ability agents are the first to prefer to tell the truth. Thus, an "all lie" equilibrium in that case is maintained when even the lowest-ability agent prefers to tell the truth. The bracketed term is more likely to be negative as the probability of detection p increases, and the scope for lying, δ, decreases. The intuition for why the highest-ability agent is the first to prefer to tell the truth in this case is the following: the highest-ability agent is the best at producing output, and so has the best shot at having the highest output even if she reports truthfully. Thus, when probability of detection is high, the highest-ability agent is the first for whom it is not worth the probability of getting caught and fired. And, when the scope for lying is low, the highest-ability agent has even more of a shot at having the highest output when she reports truthfully, since the other misreporting agents can't inflate their reports by very much.

By examining the condition in Equation 5, we infer conditions that make an "all lie" equilibrium more likely. We can easily see that as the probability of detection decreases (p decreases), the bonus B increases, the noise level increases (λ decreases), and F becomes less harsh (recall that F, the disutility from being fired, is negative), the "all lie" equilibrium becomes more likely.

We now compare the probability that a_i is promoted when output is and is not contractible. The probability that a_i is promoted when output is not contractible and misreporting is therefore possible is given by:

$$\Pr(a_i \textit{ is promoted}) = \frac{a_i^{\frac{N-1}{N(1-\lambda)-1}-1}}{\left(\sum_{j=1}^N \log a_j\right)^{\frac{\lambda}{N(1-\lambda)-1}}} E(p,N) + p^{N-1}(1-p).$$

By way of comparison, in a model with contractible output with no possibility of lying,[17] the probability that a_i is promoted is given by:

$$\Pr(a_i \textit{ is promoted, contractible}) = \frac{a_i^{\frac{\lambda}{(1-\lambda)}} \left[e^{-\lambda \bar{\varepsilon}_{N-1}}\right]}{\left(\sum_{j=1}^N \log a_j\right)^{\frac{\lambda}{N(1-\lambda)}}}.$$

When output is not contractible, the probability of promotion has a term that does not depend on ability and captures the chance of not being caught: $p^{N-1}(1-p)$. In addition, the ability term has less weight than in the case of contractible output that is captured by the expected maximal order statistic. The lying mayor no longer competes against all other mayors since some of them will get caught. The main takeaway is that, when output is not contractible and we are in the "pure lie" cheating equilibrium, the probability that any mayor a_i is promoted depends less on her ability a_i and more on a randomly-drawn ability than in the "no cheating/output is contractible" scenario.

The comparative statics described in Proposition 3 follow from the equations characterizing equilibrium effort choice by mayors and the equilibrium compensation scheme set by the principal. Appendix A.4 contains technical details and derivations.

Screening Ability when Output is Non-Contractible

Our final result characterizes the degree to which mayor manipulation decreases the screening ability of the compensation mechanism.

Proposition 4. *If $\omega < \tilde{\omega}$, so that the tournament compensation is optimal in the "all lie" equilibrium where all mayors misreport output:*

(i) The expected ability of the promoted mayor when output is not contractible is a weighted sum of the expected ability of the promoted mayor when output is contractible and there is no misreporting and the expected ability of a randomly-drawn mayor.

(ii) In tournaments with weaker audits (lower p), the expected ability of the promoted mayor is closer to a random draw.

(iii) In tournaments where audits are completely uninformative ($p = 0$), the expected ability of the promoted mayor when output is non-contractible is exactly the population average.

We observe this directly by comparing the expressions for the expected ability of the promoted

[17] See Appendix A.5 for the derivation of this case.

mayor when output is contractible given by:

$$E^{contractible}[a_i|i\ is\ promoted] = \frac{\sum_{j=1}^{N} a_j^{\frac{\lambda}{(1-\lambda)}} \exp\left(-\sum_{j=1}^{N-1}\frac{1}{j}\right)}{N\left(\sum_{j=1}^{N}\log a_j\right)^{\frac{\lambda}{N(1-\lambda)}}}$$

and the corresponding expression when output is not contractible but all mayors cheat:

$$E^{non-contractible}[a_i|i\ is\ promoted] = \frac{\sum_{j=1}^{N} a_j^{\frac{N-1}{N(1-\lambda)-1}-1} E(p,N)}{N\left(\sum_{j=1}^{N}\log a_j\right)^{\frac{\lambda}{N(1-\lambda)-1}}} + Np^{N-1}(1-p)\frac{1}{N}\sum_{j=1}^{N} a_j.$$

In the non-contractible case, the expected ability of the promoted mayor includes an extra term not present in the contractible case: $Np^{N-1}(1-p)\frac{1}{N}\sum_{j=1}^{N} a_j$; this term is proportional to the population average and is the expected ability of mayors that are not caught cheating. The key difference in the screening ability of tournament is that, in the case of non-contractible output, the mechanism places positive probability on promoting a random mayor that increases in the weakness of audits (low p). At $p = 0$, the expected ability is exactly the population average.

3 Measuring Promotion and OCP Performance

Our study focuses on the time period 1985-2000. This time period is ideal for studying the relationship between mayor promotions and the implementation of the OCP, as the system that monitors and sets birth targets was built in the 1980s with the ultimate goal of containing population growth by year 2000.

Sample of Mayors

In order to collect Chinese mayoral data, we digitized a complete list of mayors in office from 1985-2000 from two series of hard copy records: *City Gazetteers* published by the gazetteer office of each city and *the City Development Yearbook* published by the Chinese Urban Development Research Council. The list includes the mayor's name and the year and month at the start and end of his/her term. We obtained data on 967 mayors in 258 prefectures and 28 provinces between 1985 and 2000. While data on Chinese political leaders at the provincial level is more commonly accessible, to our knowledge there is no such comprehensive data for Chinese mayors before 2000. Landry (2008) is the only other example using mayor data in years 1990-2000.[18]

Promotion

Promotion is defined as an upward move in the political career. The most natural upward move for a mayor is becoming the party secretary in the same or a different prefecture, which is the

[18] In Landry (2008), promotion is defined as being promoted to the party secretary in the same prefecture or elsewhere. This definition underestimate the likelihood of promotion because mayors could move to higher ranked positions at the province or central level. We use the most complete definition of promotion based on the bureaucratic rank. In Table 1, we show that 15% of promotion of mayors in 1985-2000 was above the prefecture level.

definition of mayor promotion in the existing literature (Landry, 2008). This measure, though convenient, ignores other possible moves above the prefecture level, including provincial governor or vice-governor, minister of central ministries, etc. In our analysis, we define a promotion as an increase in the bureaucratic rank of Chinese officials. A mayor is defined as being promoted if she moves to any of the following positions at the end of her term:

1. Prefecture: party secretary in the same or a different prefecture.
2. Province: provincial governor or vice-governor, party secretary or vice-secretary, party committee member, chairman or vice-chairman of the People's Political Consultative, chairman or vice-chairman of the People's Congress.[19]
3. Central: minister or vice-minister of central ministries.

A mayor is not promoted if she continues as mayor, moves to positions of the same bureaucratic rank, or exits the political career. First, one could continue as mayor in the same or a different prefecture. In our data, forty mayors served in two prefectures. If one is transferred from the first city to the second, she is not promoted in the first city and her promotion status in the second city depends on her move after serving the second time. Second, one may be promoted to positions in the provincial government that have the same bureaucratic rank as mayor: director or vice-director of provincial departments, assistant to the provincial governor, etc. Finally, one could leave the political career by working in industry, retiring, etc. See Appendix B for details on the measurement of promotion in the data.

OCP Performance

We digitized birth rate targets from *Provincial Five-year Plans* in 1985, 1990, and 1995. Targets for policy outcomes are set every five years. For example, a province's 1985 plan sets the target for annual birth rate in 1986-1990.

We use two measures of the birth rate to compare the reported OCP performance with actual performance: 1) official birth rate from published data that are reported to provincial governments, and 2) birth rate from microdata of 1990 and 2000 population censuses that are not observed by the provincial government on a yearly basis. Official birth rate data at the prefecture level are recorded in *City Statistical Yearbooks*. A mayor's reported OCP performance is measured by comparing the reported birth rate with the target from the corresponding fiveyear plan. The lower the reported birth rate relative to the target, the better the mayor is doing at implementing the OCP in order to control the birth rate. Unfortunately, not all prefectures publish data on birth rates consistently in 1985-2000. On average, 80% of prefectures report birth rate data, except in 1988, when when no

[19] An alternative is to define promotion to province-level positions based on administrative division, i.e., province is a higher administrative division than prefecture. However, it is more controversial than the definition based on bureaucratic rank. In our definition, if a mayor becomes the director of a department in the provincial government that has the same bureaucratic rank as mayor, he/she is not defined as being promoted. We also examine the robustness of our analysis by defining promotion based on administrative division, e.g., a move to the director of any provincial department is a promotion.

prefectures published birth rate data.[20] In our sample of mayors, 754 out of 967 are matched with the official birth rate data.

We measure actual OCP performance by the gap between the birth rate target and the birth rate from census data. We compute the birth rate retrospectively using microdata from the 1990 and 2000 censuses; these data are observed by provincial government only in census years. Birth rates in 1986-1989 come from the 1990 census and those in 1990-2000 come from the 2000 census. The main concern regarding the use of census data is the potential for internal migration, since prefecture of birth is not observed for migrants. Migration was tightly restricted under the *Hukou* system until its relaxation in the 1990s. Figure A.1 shows the percentage of migrants in 1982-2000 from census and population surveys. The migration rate remained under 2% in the 1980s and slowly increased to 4% in 1995. The most significant increase was between 1995 and 2000, and the migration ratio reached 11% in 2000. We use the best available information on migration in the census to account for migration in measuring actual birth rate. We discuss how we measure birth rate from census data and account for migration in Appendix B.

Summary Statistics

Our main analysis sample includes 756 mayors in 216 prefectures and 28 provinces. The number of prefectures in a province varies from 1 to 21.[21] Table 1 reports the summary statistics at both the mayor level and mayor-year level. 52% of mayors were promoted to a higher-ranked position, with 44% promoted to party secretary in the same or a different prefecture, 7% to leadership at the province level, and 2% to central ministries. Among all promotions, 15% moved above the prefecture level, suggesting a substantial underestimation of mayor promotion by the definition in previous studies. On average, mayors spent 3.8 years in office. Figure 2a shows the distribution of years in office. Most mayors were in office from two to five years. The turnover rates are especially high in the second and third years. Tenure at promotion has similar properties and is graphed in Figure 2b.

A key variable of interest is OCP performance, which is measured by the gap between the birth rate target and a given measure of birth rate. The average reported birth rate is 7.6 per 1,000 population, while the birth rate computed using census data is 8.2. Both are lower than the target average of 10.6. On average, the reported birth rates are 3 births per 1,000 people below target. 84% of mayors reported birth rates lower than their assigned target. In comparison, birth rates from census data suggest that only 75% of mayors were below their specified target. Figure 3 plots the empirical cumulative distribution functions of the two OCP performance measures and shows that misreporting occurs at most points in the distribution of outcomes. In our analysis, we use changes in targets as instruments for OCP performance. Figure 4 presents (a) the average birth rate target across provinces, as well as (b) the number of provinces that experienced an increase in the target in each of the 5-year plans.

Data on annual nominal GDP at the prefecture level in 1985-2000 come from *City Statistical Yearbooks.* We use the nominal GDP and national current price index (CPI) to compute real GDP.

[20] Column 2 of Table A.1 summarizes the number of prefectures with and without birth rate data by year.

[21] Figure A.2 plots the histogram of the number of cities per-province.

The average real GDP in our sample is 8424 million RMB. Finally, we compiled prefecture-year controls from *City Statistical Yearbooks* including population, percentage of urban population, and government investment.

4 Inferring the Principal's Objective from the Promotion Rule

We connect the theory to the data by testing empirical predictions from our model in Section 2. The first set of predictions analyzes the effect of reported OCP performance on the promotion of mayors. These predictions use comparative static results on the noisiness of the output measure and on the level of competitiveness across provinces to provide empirical characterizations of the equilibrium promotion rule. Section 5 analyzes additional predictions on the screening ability of the tournament model as well as the manipulation behavior of mayors.

Prediction 1: *If the principal only cares about increasing output production ($\omega = 1$), then we should observe that increasing output (OCP performance) does not affect the probability of promotion.*

This prediction suggests that comparing promotion outcomes across mayors with different levels of OCP performance will allow us to infer whether the principal's objective has a significant meritocratic motive. Unfortunately, if the principal promotes mayors on an alternative metric that is positively correlated with OCP performance, one could erroneously conclude that the mayor follows a meritocratic promotion rule. Prediction 2 provides further guidance in this case.

Prediction 2: *Unconditional on unobserved margins on which a mayor gets promoted, if the principal only cares about increasing output production ($\omega = 1$), then we should observe:*

1. *Increasing output (OCP performance) increases the probability of promotion.*

2. *Increasing output (OCP performance) has a larger positive impact on a_i's probability of promotion in noisier environments (larger λ).*

3. *Increasing output (OCP performance) does not have a differential impact on a_i's probability of promotion in more vs. less competitive environments.*

Thus, by comparing the effect of OCP performance on promotion outcomes across provinces with different levels of competition and noise in the output variable, we can test whether a positive effect of OCP performance on promotion is due to alternative production metrics or to a meritocratic objective. Finally, Prediction 3 shows that the settings in the first two predictions are empirically distinguishable from the case of a meritocratic promotion rule.

Prediction 3: *If the principal cares about meritocracy, that is, promoting the highest-ability agent, in addition to increasing output production ($\omega < \tilde{\omega}$), then we should observe:Increasing output (OCP performance) increases the probability of promotion.*

1. *Increasing output (OCP performance) has a smaller positive impact on a_i's probability of promotion in noisier environments (larger λ).*

2. *Increasing output (OCP performance) has a smaller positive impact on a_i's probability of promotion in more competitive environments (larger N).*

4.1 Empirical Specification and Identification

Our first specification examines whether better reported OCP performance increases a mayor's probability of promotion using a linear probability model:

$$Promoted_{icpt} = \beta_1 OCP^{reported}_{cpt} + X_{icpt}\gamma + \mu_i + \eta_{cp} + \lambda_t + \varepsilon_{icpt}, \tag{6}$$

where i denotes the mayor, c the prefecture, p the province, and t the year. The dependent variable, $Promoted_{icpt}$ is a binary outcome that is equal to 1 if mayor i in prefecture c of province p is promoted in year t and 0 otherwise. The key regressor of interest is reported OCP performance, measured as $OCP^{reported}_{cpt} = Target_{pt} - Birth\ Rate^{reported}_{cpt}$. Superior performance in implementing the OCP, measured by $OCP^{reported}_{cpt}$, corresponds to a lower reported birth rate compared to the target. X_{icpt} is a vector of time-varying attributes of mayor i or prefecture c in year t, including the mayor's tenure, and prefecture-year log of real GDP, log of population, percentage of urban population, log of investment, and migration controls. Mayor fixed effects, μ_i, account for all time-invariant characteristics of the mayor i. Year fixed effects, λ_t, control for all national changes over time. Moreover, we are able to control for prefecture fixed effects, η_{cp}, because some mayors served two different prefectures. Standard errors are clustered at the province-year level.

Our second specification examines whether the effect of OCP performance on promotion differs by noisiness and competitiveness of the environment. To measure the noisiness of OCP performance as a signal of effort and ability, we use the standard deviation of birth rates by province in the census data.[22] To measure competitiveness, we use the tenure of positions above the bureaucratic rank of mayors. In provinces with longer tenure at these upper-level positions (and thus less turnover), mayors have fewer opportunities of being promoted and must work harder at proving themselves worthy of promotion. In order to implement this strategy, we digitized hard-copy records on the term information of all province-level officials ranked higher than mayors. The average tenure of provincial officials ranges from 3 to 6 years across provinces (see the distribution in Figure A.3). In both cases, we assume that noisiness and competitiveness vary at the province level. These are natural assumptions since promotions for mayors are determined at the province level. We estimate Equations 7 and 8 separately:

$$Promoted_{icpt} = \beta_1 OCP^{reported}_{cpt} + \beta_2 OCP^{reported}_{cpt} * Noise_p + X_{icpt}\gamma + \mu_i + \eta_{cp} + \lambda_t + \varepsilon_{icpt} \tag{7}$$

$$Promoted_{icpt} = \beta_1 OCP^{reported}_{cpt} + \beta_3 OCP^{reported}_{cpt} * Comp_p + X_{icpt}\gamma + \mu_i + \eta_{cp} + \lambda_t + \varepsilon_{icpt}. \tag{8}$$

Equations 6-8 directly correspond to empirical Predictions 1-3. Prediction 1 suggests that $\beta_1 = 0$; Prediction 2 suggests that $\beta_1 > 0$, $\beta_2 > 0$ and $\beta_3 = 0$; and Prediction 3 suggests that $\beta_1 > 0$, $\beta_2 < 0$ and $\beta_3 < 0$.

[22] In order to interpret this heterogeneity in output as noise, we make the assumption that the distribution of abilities across provinces is held constant.

Identifying assumptions

There are three main concerns in identifying β_1 , β_2, and β_3. First, a mayor could be promoted on unobserved non-OCP performance margins that are correlated with ability. This concern is directly testable. If this is the case, we should find that Prediction 2 fits the data best, that is, $\beta_1 > 0$, $\beta_2 > 0$ and $\beta_3 = 0$.

Second, if we find that $\beta_1 > 0$, an alternative interpretation could be signaling. As commonly observed among Chinese top leaders, high ability mayors could be assigned to "problem prefectures" with larger challenges to showcase their performance. This concern is addressed by including mayor fixed effects, μ_i, that control for time-invariant ability that could affect the initial placement.

Third, unobserved political connections are an important concern in the existing literature. Again, mayor fixed effects, μ_i, capture time-invariant connections that could affect initial placement. However, if political connections could be lost or expanded over time, the change in connections could be confounding. Jia (2014) finds that shuffling of politicians in the central government changes the connections of provincial leaders. Using an instrument variable approach below, we take advantage of variation in the reported OCP performance that is uncorrelated with either change in connections or other changes in unobserved margins on which a mayor is promoted.

Instrument Variable for Reported OCP Performance

We leverage the fact that birth rate targets are set at the province level by the central government in five-year plans. Changes in direction and magnitude of these targets generate "surprise changes" in reported OCP performance among mayors in office. For example, if there is an increase in the birth rate target, it is easier for mayors to get closer to the target and thus achieve a better OCP performance. We use the change in the birth rate target to instrument for reported OCP performance. Figure 4a shows the average birth target from 1985 to 2000. Birth rate targets were changed twice during these fifteen years; the 1990 plan saw an average increase from 1986-90 to 1991-95 and the 1995 plan saw an average decrease from 1991-95 to 1996-2000. For a given plan, however, there was substantial variation in whether a province experienced an increase or a decrease in target. Figure 4b shows that 20 provinces had an increase in the target in the 1990 plan, while only 8 provinces had an increase in the target in the 1995 plan.

Two identifying assumptions underpin this instrumental variable approach. First, we assume that the new province target is not set to favor a particular mayor. While this assumption is plausible given that most provinces have 5 or more prefectures, we analyze the potential for this concern. For example, if politically connected mayors performed worse in birth control than unconnected ones, the birth rate target might be raised to help connected mayors improve their performance. Although an increase in target could favor every mayor within a province, only those staying in office after the change experience the benefit. Thus, we examine whether mayors with worse OCP performance before an increase in target are more likely to stay in office after the plan year. Meanwhile, we do not expect that worse performers are more likely to stay after a decrease in target. The second assumption underpinning our analysis is that expected changes in birth rate targets do not change the timing of promotion decisions. If the change in target could be anticipated by connected mayors, there could be selection in promotion before the target change. To test this concern, we examine whether future changes in the target affect promotion and find no evidence of this concern.

4.2 Does Reported OCP Performance Increase the Probability of Promotion?

We begin by showing the correlation of reported OCP performance and promotion in Figure 5. Reported OCP performance is measured by the difference between the reported birth rate and the birth rate target. A positive and larger gap corresponds to improved OCP performance. The X-axis represents the residualized OCP performance and the Y-axis represents residualized promotion probability, where we control for person, city, and year fixed-effects. Panel A of Figure 5 shows the relation in the subsample where the reported birth rate is above target (negative OCP performance); we do not observe any correlation between reported OCP performance and promotion. Panel B of Figure 5 plots the relation when the reported birth rate is equal to or below target (non-negative OCP performance). In this case, OCP performance is positively correlated with probability of promotion. This is consistent with anecdotal evidence that local officials are rewarded for their OCP performance only if they have met the birth rate target.

We focus our regression analysis in the subsample with non-negative OCP performance.[23] The baseline regression results in Table 2 show similar findings to those presented in Panel B of Figure 5. Consistent with the graphical presentation, estimates from column (1) through (4) all show that mayors with better OCP performance are more likely to be promoted. In this and other tables, we focus on the results from the richest specification in column (4). Decreasing birth rates by 1 per 1000 increases the chance of promotion by 1.4 percentage points, or around 10% of the probability of promotion. To compare with GDP, we also show the estimate of log (GDP) and find that increasing GDP by 1% increases the chance of promotion by 17.5 basis points. These estimates suggest an economically large effect of OCP performance compared to economic growth, since decreasing the birth rate by 1 per 1000 is equivalent in its effect on promotion to an 8% increase in GDP.

Relative to the predictions of our model, we find that $\beta_1 > 0$, which rules out Prediction 1. To test Predictions 2 and 3, we further examine whether the effects of OCP on promotion vary by the noisiness and competitiveness of the environment.

4.3 Do Signal Noise and Competitiveness Affect the Promotion Rule?

Table 3 presents results from estimating Equation 7. Column (1) is similar to column (4) in Table 2 except for not including GDP, and the estimate of reported OCP performance remains unchanged. Column (2) includes an interaction of reported OCP performance and the standard deviation of actual birth rates. It indicates that the marginal effect of increased OCP performance on promotion is decreasing in the noisiness of actual birth rates, that is, $\beta_2 < 0$. In column (3), log GDP is included to match our preferred specification of column (4) in Table 2. We focus on the results in column (3) and use these estimates to produce the plot in Figure 6a. The X-axis is the noisiness measure in 100 quantiles; the larger the measure, the noisier the signal. The Y-axis is the predicted percentage change in the probability of promotion of increasing OCP performance by 1 per 1000. In the visual

[23] We also present the estimates using the full sample in Appendix Table 2. In all specifications, we do not find that the OCP performance is significantly correlated with promotion.

presentation, the promotion incentive on OCP performance continuously gets smaller as the signal becomes noisier. In provinces where the signal is the noisiest, the effect of OCP performance on promotion is null. By contrast, in provinces in the 10th-percentile of the distribution of noisiness, an increase in OCP performance leads to a 20% increase in the probability of promotion.

Table 4 shows results from estimating Equation 8. Column (1) presents a baseline regression similar to column (4) in Table 2 except for not including GDP. Column (2) adds an interaction of reported OCP performance and the competitiveness measure. We find that, in provinces where promotions are more competitive, the marginal effect of reported OCP performance on promotion is smaller, that is, $\beta_3 < 0$. Column (3) includes log GDP. Using estimates in column (3), we plot Figure 6b. This graph shows that the higher the degree of competition, the smaller is the effect of OCP performance on promotion.

To summarize, we find that $\beta_1 > 0$, $\beta_2 < 0$ and $\beta_3 < 0$. Prediction 3 fits the data best; Predictions 1 and 2 are ruled out. These results show that provincial governors are instructed by the central government to value OCP performance both because population control is inherently valued and because doing so may select high ability mayors for promotions.

4.4 An Instrumental Variable for Reported OCP Performance

A potential concern is that our estimates could be biased if unobserved changes in political connections are correlated with reported OCP performance and the probability of promotion. We now turn to an instrument variable approach, using the variation in the reported OCP performance that is generated by changes in the province-level birth rate target. Column (1) of Table 5 presents the first stage estimates. An increase in birth rate target is positively correlated with reported OCP performance. The estimate is very precise and statistically significant at the 1 percent level.[24] Column (2) reports OLS estimates in this sample that are similar to column (4) of Table 2. Column (3) presents 2SLS estimates, with an F-value of the first stage of 67, indicating a strong first stage. The 2SLS estimate is slightly larger than the OLS estimate but statistically indistinguishable. Decreasing birth rates by 1 per 1000 increases the chance of promotion by 2 percentage points. Since the average promotion of the sample in Table 5 is slightly larger, this represents an increase of 14% in the probability of promotion, which is economically similar to the implied estimate in the OLS regression. However, these findings might suggest that unobserved changes in connection could lead to downward bias of the promotion incentive, consistent with the observation that well-connected candidates for promotion are assigned to "problem places" with larger challenges.

As discussed above, we explore potential concerns about the validity of the IV approach. We have investigated each one of these concerns and present results in Table A.3. First, if the target is increased to favor connected mayors who performed worse, they might be more likely to stay in office after an increase in target. In Panel A, we split the sample into two parts. Column (1) uses the subsample in provinces and years with an increase in birth rate target, and column (2) uses the subsample in provinces and years with a decrease in birth rate target. In both subsamples, we fail to find evidence that the OCP performance prior to a target change is correlated with whether

[24] Note that the sample is smaller than that in column (4) of Table 2 because the birth rate target data are unavailable in a few years in a few provinces.

they stay in office after the change. Second, if a change in the birth rate target is anticipated by connected mayors, there could be selection on whether they are promoted prior to the change. In Panel B, we do not find a statistically significant correlation between the next change in birth rate target and promotion, suggesting that the target change is not set to favor some (connected) mayors. Finally, one might be concerned that mayors respond to a decrease in target by reporting lower birth rates. Panel C indicates that changes in birth rate target do not significantly change the difference between the birth rate from census data and reported birth rate, which rules out the last concern.

We also test whether the effect of OCP performance on promotion differs by noisiness and competitiveness of the environment using the IV approach. In Table A.4, the OCP performance is instrumented by the change in birth rate target and the interaction of OCP performance and standard deviation of actual birth rate is instrumented by the interaction of the change in birth rate target and standard deviation of actual birth rate. Consistent with the results in Table 3, the OCP performance has a smaller effect on promotion in noisier environments. In Table A.5, we compare the promotion incentive in regions with high versus low levels of competition (based on the median of the average tenure of upper-level officials). We instrument for low competition interacted with OCP performance using low competition interacted with the change in target and high competition interacted with OCP performance using high competition interacted with the change in target. The 2SLS results are consistent with the baseline results that the promotion incentive becomes smaller in more competitive environments.

5 Empirical Evidence of Manipulation and Screening Ability

The previous section show that reported OCP performance has a positive and substantively large effect on the probability that a given mayor is promoted. Moreover, tests of comparative statics further support the view that the promotion rule is consistent with a meritocratic objective. However, it is still an open question whethersuch a tournament mechanism with non-contractible output is able to screen high ability mayors and whether the population audits have an effect on misreporting behavior. Our fourth prediction allows us to distinguish between output being contractible for the principal (no cheating) and output being non-contractible (cheating).

Prediction 4: *If output y, the birth rate, is contractible, then audit years should not affect the degree of manipulation.*

Finally, our fifth prediction describes how the expected ability of the promoted mayor should differ when output is contractible (no cheating) and when output is non-contractible, when audits are weak.

Prediction 5: *If audits are weak, then when output (the birth rate) is contractible, the ability of the promoted mayor is higher than the population average. When output (the birth rate) is non-contractible, and mayors are therefore manipulating reported birth rates, the ability of the promoted mayor is close to the population average.*

5.1 Empirical Specification and Identification

As shown in Figure 3, census data indicate that birth rates are higher than in the reported data for most levels of reported birth rates. Prediction 4 suggests that, if no manipulation is taking place and mayors are truthfully revealing their performance, the difference between these two data sources should not depend on whether an audit is taking place. As discussed in Section 1.3, the central government uses population census and national fertility surveys to investigate the actual birth rates and the credibility of reported birth rates, which are organized at the province-level. The audit year is the year before the census or fertility survey when the actual birth rates are fully observed. Equation 9 tests whether the difference between reported OCP performance and the OCP performance measure from the census is smaller one year prior to the census or national fertility survey:

$$OCP_{cpt}^{census} - OCP_{cpt}^{reported} = \delta Audit_t + f(t) + X_{icpt}\gamma + \mu_i + \eta_{cp} + \varepsilon_{icpt}, \tag{9}$$

where the binary variable $Audit_t$ is equal to 1 in years 1987, 1989, 1991, 1994, 1996, and 1999. We include a flexible year trend $f(t)$. If the difference indeed suggests data manipulation, we should observe that $\delta < 1$. If this is the case, we further examine whether the decrease in the difference from these two data sources comes from higher reported birth rates, suggesting less manipulation, or from lower actual birth rates, indicating actual improvement in OCP enforcement.

Finally, Prediction 5 suggests that in the presence of misreporting in the reported OCP performance, promoted and unpromoted mayors could be similar in their actual OCP performance. We test this hypothesis in Equation 10:

$$Promoted_{icpt} = \beta_4 OCP_{cpt}^{census} + X_{icpt}\gamma + \mu_i + \eta_{cp} + \lambda_t + \varepsilon_{icpt}, \tag{10}$$

where the key regressor of interest is the OCP performance measure from census data, measured as $OCP_{cpt}^{census} = Target_{pt} - Birth\,Rate_{cpt}^{census}$. The lower the census birth rate compared to the target, the better the actual OCP performance. Our theory predicts that $\beta_4 \approx 0$ in equilibrium.

5.2 Effects of Audits on Output and Data Manipulation

We find that Prediction 3 in our theory fits the data best. In this world, the provincial governor uses OCP performance to screen high ability mayors and ability is unobserved, so that a positive amount of misreporting may arise in equilibrium. However, the difference between the reported OCP performance and the actual OCP performance from the census could indicate data misreporting, or measurement error. To examine whether the difference indeed captures cheating, we compare the difference in audit years versus non-audit years by estimating Equation 9.

In column (1) of Table 6, we find that, in audit years, the difference between reported OCP performance and actual OCP performance from the census becomes smaller, that is, $\delta < 1$. This is consistent with Prediction 4 and is evidence that mayors manipulate reported birth rates. We further examine whether the decrease in the difference from these two data sources comes from higher reported birth rates, suggesting less manipulation, or lower actual birth rates, indicating actual improvement in OCP enforcement. In columns (2) and (3), there is suggestive evidence that 62% of the decrease in the difference is attributable to higher reported birth rates in audit years, while 38% is attributable to mayors exerting more effort to lower the actual birth rate. Finally, if lower birth rates and higher GDP growth rate are substitutes, we would expect that, in audit years when mayors reported higher birth rates (thus doing worse on the OCP dimension), mayors might work harder to improve GDP or report higher GDP numbers. We find that this is indeed the case in column (4).

5.3 Screening Ability in the Presence of Misreporting

Prediction 5 suggests that, in the presence of misreporting, the promotion mechanism is closer to simply choosing a mayor at random. We estimate Equation 10 and present results in Table 7. To enable comparison with Table 2, we use the subsample with non-negative reported OCP performance. From column (1) through column (4), results from all specifications suggest that that actual OCP performance is not significantly predictive of promotion. These findings imply that promoted mayors do not have significantly higher ability in lowering actual birth rate than mayors who are not promoted. This is supportive of Prediction 5.

We have utilized the best information available in the census to account for migration in Table 7. Furthermore, we investigate potential measurement error from not directly observing migration before 1995 (when the migration rate was below 4%). We repeat the specification in column (4) of Table 7, with various incomplete controls for migrationand using the same subsample, and report the results in column (1) through column (3) in Table A.6. Column (1) does not account for migration in 1995 and 2000 and does not include migration controls in 1985-1994 interacted with time fixed effects. Column (2) does not consider migration in 1995-2000 but does include migration controls in 1985-1994 interacted with time fixed effects. Column (3) accounts for migration in 1995 but does not include migration controls in 1985-1994 interacted with time fixed effects. There is very little difference in these estimates and the estimates in column (4) of Table 7, which include a broad set of controls for migration. Using the full sample in columns (4) through (6), we conduct the same exercise and compare the estimates with those in column (4) of Table A.6. These results provide supportive evidence that the lack of correlation between actual OCP performance and promotion is not due to attenuation bias from measurement error, but rather from the impaired capacity of the tournament mechanism to screen for ability when output is non-contractible and mayor manipulation is a concern.

6 Conclusion

This paper analyzes the role of meritocracy in determining the promotion of mayors in China. We document that, despite potential for corruption through political connections, promotion rules are partly driven by perceived performance in implementing the OCP. Moreover, we show that the relationship between performance and promotion is determined by a desire to screen high-ability mayors for higher office. While we confirm that observed promotion decisions are consistent with a meritocratic objective, the efficacy of this screening mechanism is weakened by mayors' ability to manipulate reported outcomes. Empirically, we find that mayors manipulate less in audit years (that is, when monitoring is increased), which is consistent with the importance of OCP as a performance metric. Nonetheless, we find that audits are not able to resolve the fundamental problem of non-contractible output.

The combination of theory and empirical analysis makes our findings particularly compelling and demonstrates the importance of interpreting empirical results through the lens of a rigorous model of incentives . Without guidance from our model, the applied econometrician could arrive at the mistaken conclusion that meritocracy was not a driving force in the Chinese government, as promoted mayors do not appear to be of higher ability than mayors who are not promoted. Thus, while the manipulability of outcomes and the inability to monitor perfectly indeed weaken the realized quality of the screening, by testing other, more subtle predictions of our model, we are able to separate the desire to implement the OCP from the meritocratic objective, a distinction which previous studies of promotion based on other measures of performance were not able to address.

We conclude by noting that, while critics of the implementation of the OCP point toward local government promotion incentives as a cause for human rights abuses, including forced abortions and sterilizations (see, *e.g.*, Wong (2012)), the alternative piece-rate compensation mechanism would likely lead to more such cases as effort from lower ability mayors would likely increase. Moreover, one consequence of the non-contractilibility and manipulation of birth rates is that marginal incentives may have little or no effect on actual birth rates, even though the policy on the whole may lead to human rights abuses.

References

Baker, G., M. Gibbs, and B. Holmstrom (1994). The internal economics of the firm: Evidence from personnel data. *The Quarterly Journal of Economics 109*(4), 881–919.

Bell, D. and E. Li. (November 11, 2012.). In defence of how china picks its leaders. *The Financial Times*.

Bertrand, M. and S. Mullainathan (2001). Are ceos rewarded for luck? the ones without principals are. *Quarterly Journal of Economics 116*(3), 901–932.

Birney, M. (2014). Decentralization and veiled corruption under china's rule of mandates. *World Development 53*, 55–67.

Bobonis, G., L. R. C. Fuertes, and R. Schwabe (2013). Monitoring corruptible politicians.

Chay, K. Y., P. J. McEwan, and M. Urquiola (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review 95*(4), 1237–1258.

Chen, Y.-Y., G. Z. Jin, N. Kumar, and G. Shi (2012). Gaming in air pollution data? lessons from china. *The B.E. Journal of Economic Analysis & Policy 13*(3).

Chevalier, J. and G. Ellison (1999). Career concerns of mutual fund managers. *Quarterly Journal of Economics 14*(2), 389–432.

Chiappori, P. and B. Salanie (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy 108*, 56–78.

Choukhmane, T., N. Coeurdacier, and K. Jin (2014). The one child policy and household saving.

Cullen, J. B. and R. Reback (2006). *Improving School Accountability: Check-Ups or Choice, Advances in Applied Microeconomics 14*, Chapter Tinkering Toward Accolades: School Gaming Under a Performance Accountability System, pp. 1–34. Amsterdam: Elsevier Science.

Ebenstein, A. (2010). The "missing girls" of china and the unintended consequences of the one child policy. *The Journal of Human Resources 45*(1), 87–115.

Ferraz, C. and F. Finan (2008). Exposing corrupt politicians: The effects of brazil's publicly released audits on electoral outcomes. *The Quarterly Journal of Economics 123*(2), 703–745.

Ferraz, C. and F. Finan (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review 101*(4), 1274–1311.

Fisman, R., , and Y. Wang (2015). The economics of death ceilings.

Fisman, R. and S.-J. Wei (2004). Tax rates and tax evasion: Evidence from 'missing imports' in china. *Journal of Political Economy 112*(2), 471–96.

Ghanem, D. and J. Zhang (2014). Effortless perfection: Do chinese cities manipulate air pollution data? *Journal of Environmental Economics and Management 68*, 203–225.

Hardee-Cleaveland, K. and J. Banister (1988). Fertility policy and implementation in china, 1986-88. *Population and Development Review 14*(2), 245–286.

Holmstrom, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies 66*(1), 169–182.

Jacob, B. and S. Levitt (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics 118*, 843–877.

Jia, R.-X. (2014). Pollution for promotion. mimeo.

Jia, R.-X., M. Kudamatsu, and D. Seim (forthcoming). Political selection in china: The complementary role of connections and performance. *Journal of the European Economic Association*.

Kane, T., D. G. D. Staiger, and H. Ladd (2002). *Brookings Papers on Education Policy*, Chapter Volatility in School Test Scores: Implications for Test-based Accountability Systems, pp. 235–283.

Landry, P. (2008). *Decentralized Authoritarianism in China: The Communist Party's Control of Local Elites in the Post-Mao Era.* Cambridge and New York: Cambridge University Press.

Lazear, E. and S. Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy 89*(5), 841–864.

Li, H.-B. and L.-A. Zhou (2005). Political turnover and economic performance: The incentive role of personnel control in china. *Journal of Public Economics 89*(9-10), 1743–62.

Maskin, E., Y.-Y. Qian, and C.-G. Xu (2000). Incentives, information, and organizational form. *Review of Economic Studies 67*(2), 359–78.

Olken, B. (2006). Monitoring corruptible politicians: Evidence from a field experiment in indonesia. *Journal of Political Economy 115*(2), 200–249.

Oyer, P. and S. Schaefer (2011). *The Handbook of Labor Economics*, Volume 4b, Chapter Personnel Economics: Hiring and Incentives, pp. 1769–1823.

Persson, P. n. E. Z. (forthcoming). The limits of career concerns in federalism: Evidence from china. *Journal of the European Economics Association*.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature 37*, 7–63.

Rochet, J.-C. and L. Stole (2003). *Advances in Economic Theory, 7th World Congress*, Chapter The Economics of Multidimensional Screening. Cambridge University Press.

Rockoff, J., D. Staiger, T. Kane, and E. Taylor (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *American Economic Review 102*(7), 3184–3213.

Scharping, T. (2003). *Birth Control in China 1949-2000: Population Policy and Demographic Development*. London and New York: Routledge-Curzon.

Shih, V., C. Adolph, and M.-X. Liu (2012). Getting ahead in the communist party: Explaining the advancement of central committee members in china. *American Political Science Review 106*, 166–87.

Wei, S.-J. and X. Zhang (2011). The competitive saving motive: Evidence from rising sex ratios and saving rates in china. *The Journal of Political Economy 119*(3), 511–564.

Wong, E. (2012, July, 22). Reports of forced abortions fuel push to end chinese law. *The New York Times*.

Xu, C.-G. (2011). The fundamental institutions of china's reforms and development. *Journal of Economic Literature 49*(4), 1076–1151.

Zitzewitz, E. (2012). Forensic economics. *Journal of Economic Literature 50*(3), 731–769.

Figure 1: Official Document From Fujian Province

福建省人口和计划生育工作
责任考核奖惩办法(试行)

第一条 为了稳定低生育水平，提高人口素质，确保人口和计划生育工作目标的实现，为建设海峡西岸经济区和我省全面建设小康社会创造良好的人口环境，根据《中华人民共和国人口与计划生育法》、《中共中央国务院关于加强人口与计划生育工作稳定低生育水平的决定》和《福建省人口与计划生育条例》及其他有关规定， 制定本办法。

第二条 市、县(区)、乡镇(街道)党委、政府(办事处)， 省、市、县(区)党委、政府工作部门，村(居)党组织和村(居) 民委员会、企事业单位、各社会团体和相关责任人，在人口和计划生育工作中应给予奖励或追究责任的，适用本办法。

第三条 对在人口和计划生育工作中作出显著成绩的单位和个人，给予精神和物质奖励。成绩突出的，分别对党政主要领导、分管副职和人口计生委(局)；主任(局长)给予奖励，并发给奖金。

第四条 有下列情形之一的，对有关单位和相关责任人应予以追究责任：

(一)未完成年度人口和计划生育工作主要责任目标的。

市、县(区)党委、政府及计生行政部门的主要责任目标为六项指标：1、出生人口政策符合率；2、出生人口性别比；3、统计准确率；4、群众满意率；5、计划生育基本知识知晓率；6、工作经费投入。

其他单位的主要责任目标为上述指标的第1、5、6项。

(二)违反国家人口计生委规定的“七个不准”，发生重大恶性案件的；

(三)在一年内对实行计划生育的公民兑现政策规定的奖励低于当年度应奖励总量的70%的；

(四)虚报、瞒报、漏报、伪造、篡改或者拒报人口和计划生育统计数据，报表统计准确率低于95%的；

(五)在经费投入上弄虚作假，截留、挪用、克扣计划生育经费的；

(六)对违反计划生育政策的党员和干部及村(居)民委员会成员未按有关规定处理，造成恶劣影响的；

(七)有关部门没有履行计划生育综合治理职责的，或制定出台的政策规定不利于计划生育工作开展的；

(八) 因工作不力，造成计划生育工作严重滑坡的；被计划生育领导小组黄牌警告或单列管理的；

(九)机关、企事业单位人员违反生育政策被发现后未及时处理的；

(十)村(居)班子软弱涣散或处于瘫痪状态，又未及时调整整顿，造成计划生育工作无人管理或失控的；

(十一)人口和计划生育部门及其工作人员不依法履行管理和服务职责的。

对上述情形按如下办法认定：

出现第(一)项情形的， 以上级人口和计划生育部门对其年度人口和计划生育工作目标责任制考核的结果为准。

出现第(二)至第(十一)项情形的，由上级人口和计划生育行政部门负责核实。涉及到需要进行人事调整或党纪；政纪处分的，会同组织人事部门或纪检监察机关核实。

第五条 被列为追究责任的单位，取消其当年度综合性荣誉称号和综合性评先评奖资格。其应负责任的党政主要领导或法定代表人、分管领导及直接责任人， 当年度考核中不能评为称职及以上等次，情节严重的要给予免职、责令辞职，并给予相应的党纪政纪处分。对应负责任而已调离岗位的，一年内要跟踪处理。

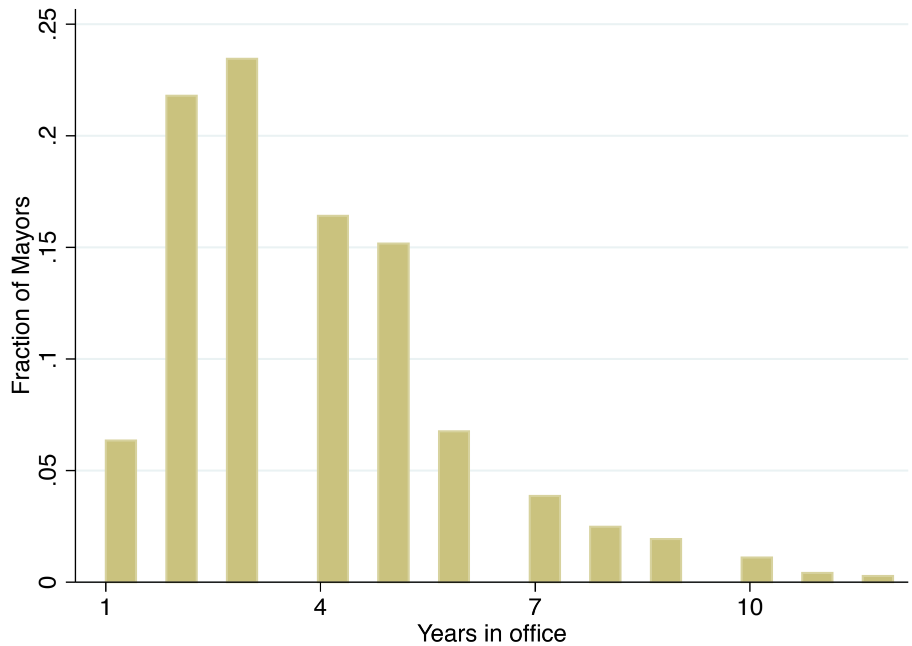
第六条 出现第四条(一)、 (六)、 (八)、(十)项规定情形的，对单位和党政主要领导或法定代表人；出现第四条其它项规定情形的，对有关部门及其分管领导和直接责任人，均按第五条规定给予相应的责任追究。

第七条 人口和计划生育行政部门负责对年度人口和计划生育目标管理责任制执行情况进行考核。考核中要注重检查单位和领导干部履行职责的情况。考核结果要向同级党委、政府或人口和计划生育领导小组和上级计生行政部门报告，提出实行计划生育工作奖励或责任追究的建议。向时，向同级组织人事部门和纪检监察机关通报有关情况。

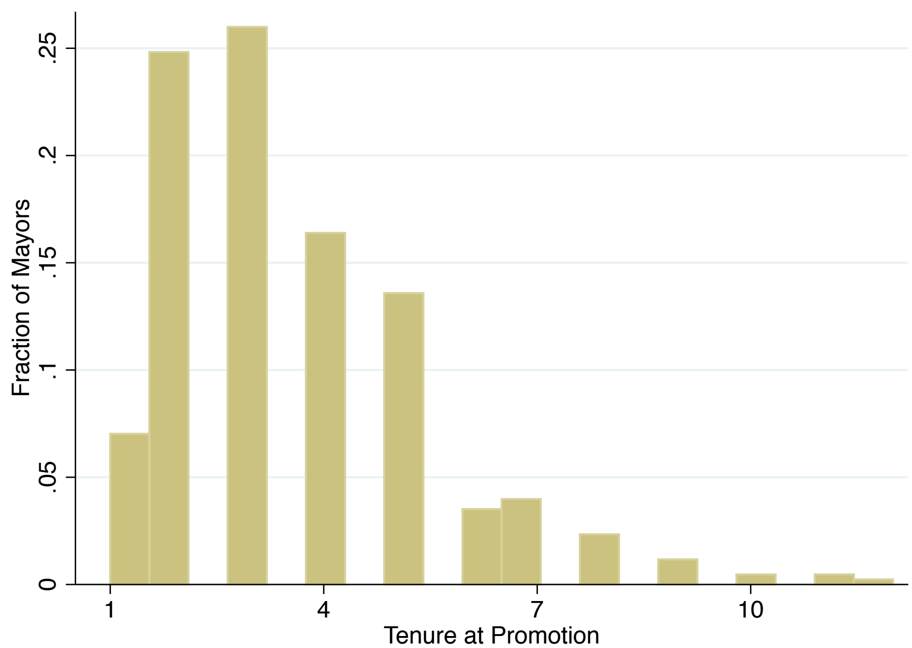
Note: This document from Fujian Province outlines guidelines for local officials on the One Child Policy and links performance to promotion outcomes. The first highlighted section states that local officials are responsible for reporting accurate birth rates and other OCP statistics. The second highlighted section states that local officials should ensure the accuracy of the reported numbers and avoid underreporting, misreporting, faking and missing birth rate statistics. Finally, the third highlighted section, mentions that the province government is responsible for investigating violations of these guidelines. If they are violated, the responsible officials are denied positive credits in their annual evaluation, and their records are sent to the personnel department of the province government. Source: http://yz.zfxxgk.gov.cn/ShowArticle.asp?ArticleID=75204

Figure 2: Histogram of Years in Office

(a) Histogram of Years in Office

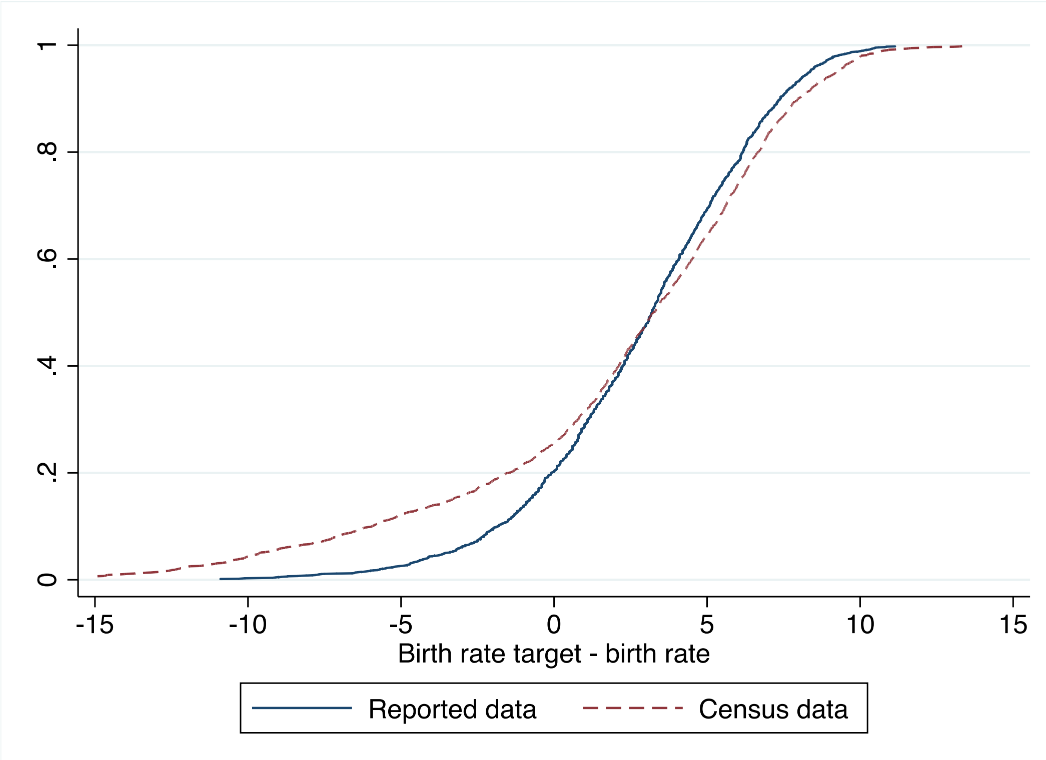


(b) Tenure at Promotion



Notes: Figure 2a shows the distribution of years in office per mayor. Figure 2b shows the distribution of tenure at promotion of mayors who were promoted.

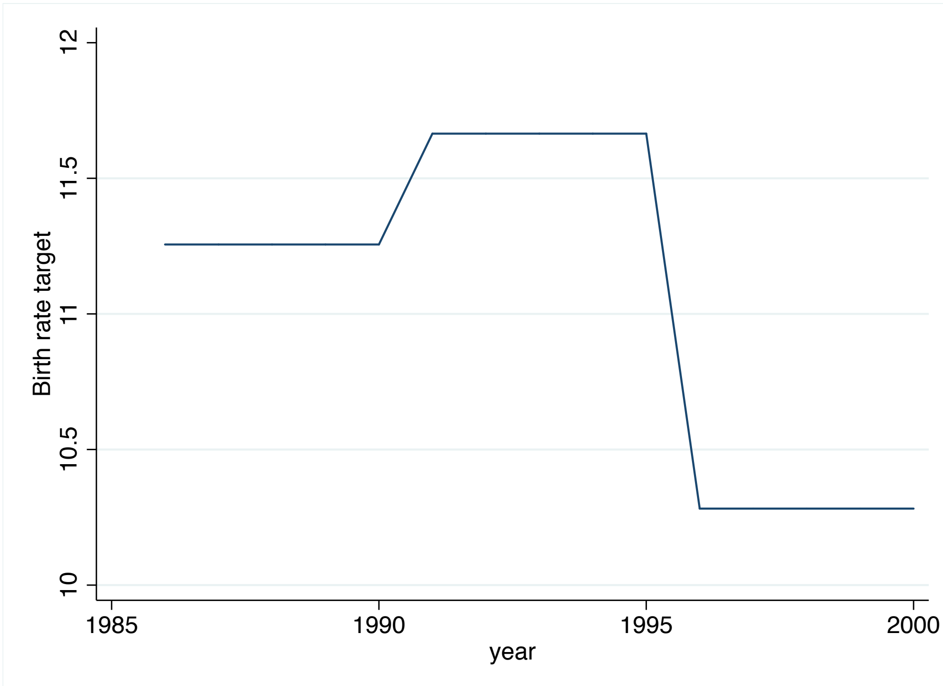
Figure 3: Distribution of the OCP performance measure (birth rate target - birth rate)



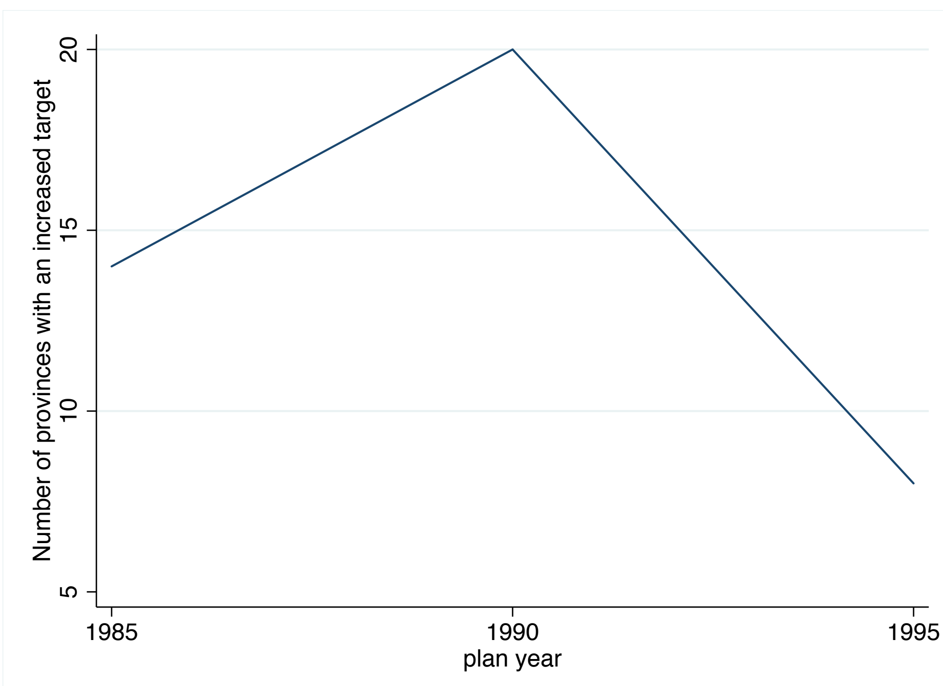
Notes: Figure 3 plots the cumulative distribution functions of the two OCP performance measures (birth rate target - birth rate), based on reported data and census data respectively.

Figure 4: Birth Rate Targets

(a) Birth Rate Target

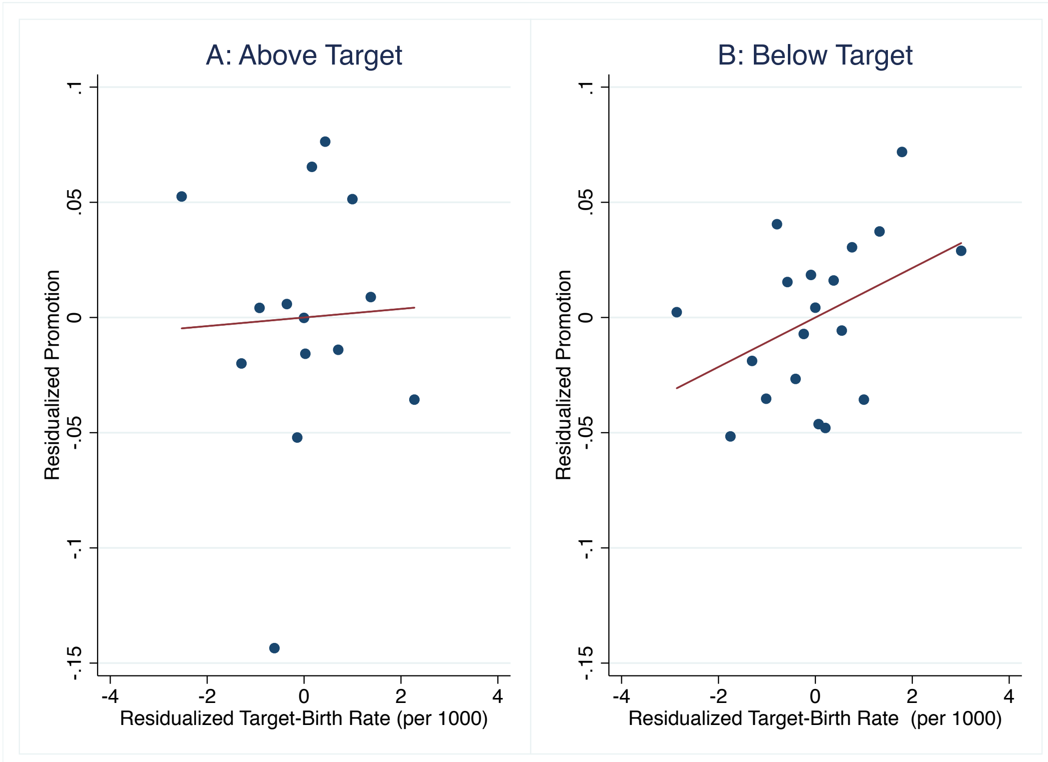


(b) Number of Provinces with an Increased Target



Notes: Figure 4a shows the average birth rate target across provinces set in five-year plans. Figure 4b presents the number of provinces that experienced an increase in the target in each of the five-year plans during 1985-1995.

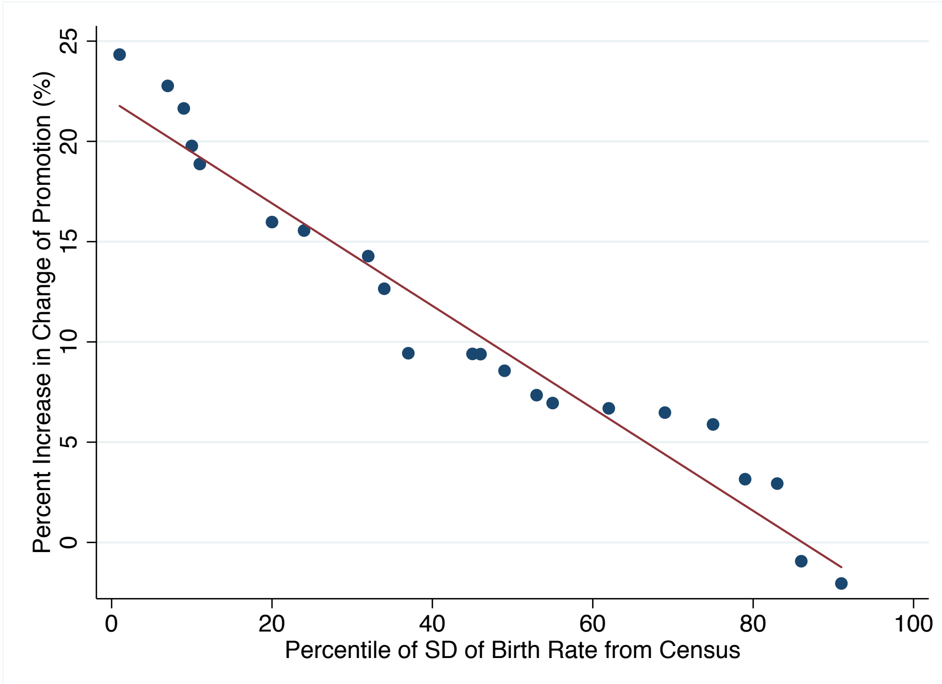
Figure 5: OCP Performance and Promotion



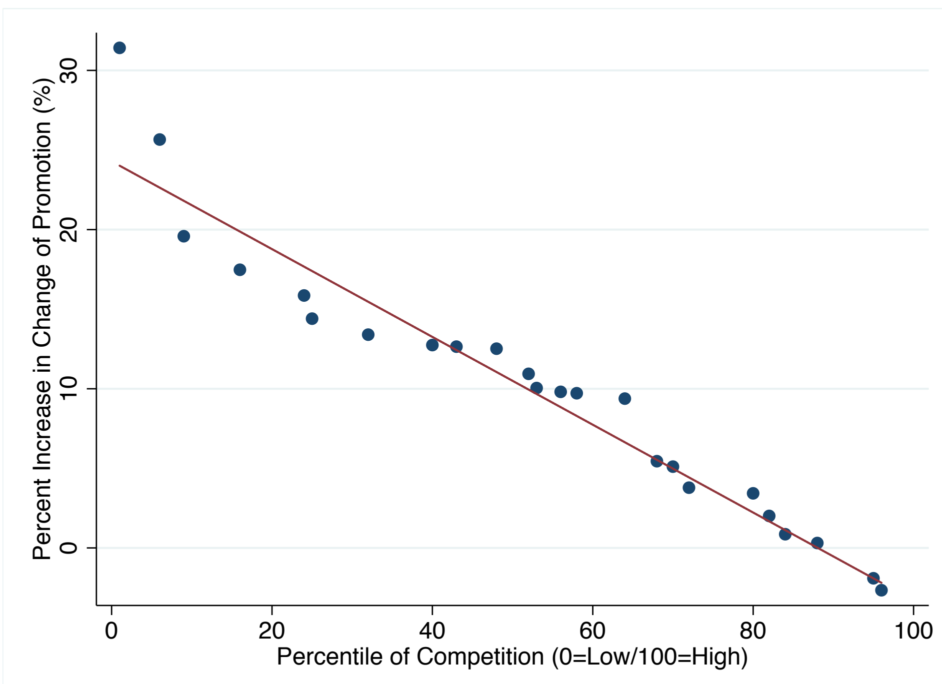
Notes: In Figure 5, the X- axis represents the residualized OCP performance and the Y-axis represents residualized promotion probability where we control for person, city, and year fixed-effects. OCP performance reported is birth rate target minus reported birth rate. Panel A shows the relation in the subsample where the reported birth rate is above target (negative OCP performance). Panel B plots the relation when the reported birth rate is equal to or below target (non-negative OCP performance).

Figure 6: Heterogeneous Effects of OCP on Promotion

(a) OCP Promotion and Outcome Noise



(b) OCP Promotion and Agent Competition



Notes: In Figure 6a, the X-axis is the noisiness measure in 100 quantiles; the larger the measure, the noisier the signal. The Y-axis is the predicted percentage change in the probability of promotion of increasing OCP performance by 1 per 1000. The figure is plotted using estimates in column (3) of Table 3. In Figure 6b, the X-axis is the competitiveness measure in 100 quantiles; the larger the measure, the more competitive. The Y-axis is the predicted percentage change in the probability of promotion of increasing OCP performance by 1 per 1000. The figure is plotted using estimates in column (3) of Table 4.

Table 1: Summary Statistics

	(1) Mayor			(2) Mayor-year		
	Mean	Std. Dev.	Obs	Mean	Std. Dev.	Obs
Promotion						
Promoted	0.52	0.50	756	0.14	0.35	2249
Prefecture party secretary	0.44	0.50	756	0.12	0.32	2249
Province government or party leaders	0.07	0.25	756	0.02	0.14	2249
Central ministries	0.02	0.13	756	0.004	0.07	2249
Tenure (year)	3.76	2.00	756	2.76	1.86	2249
Birth rate (per 1,000 population)						
Recorded birth rate	7.61	3.20	756	7.62	3.73	2249
Birth rate from census	8.15	5.37	739	8.40	5.76	2060
Birth rate target	10.60	2.28	756	10.56	2.48	2249
OCP performance						
Reported birth rate is below target (%)	0.84	0.25	756	0.80	0.40	2249
Birth rate target - reported birth rate	3.00	3.11	756	2.95	3.72	2249
Birth rate from census is below target (%)	0.75	0.38	739	0.74	0.44	2060
Birth rate target - birth rate from census	2.47	5.07	739	2.20	5.50	2060
Real GDP (million RMB)	8360	11139	756	9363	12067	2249
Log (GDP)	3.87	1.04	756	4.00	1.06	2249
Prefecture-year controls						
Population (1,000)	4931	38538	735	6030	112008	2162
Percentage of urban population	0.32	0.17	735	0.32	0.18	2162
Investment (million RMB)	4001	8798	735	4506	10332	2160

Table 2: OLS Regression of Promotion on Reported OCP Performance

	Promotion=1			
	(1)	(2)	(3)	(4)
OCP performance reported	0.009**	0.013**	0.013**	0.014**
	[0.004]	[0.006]	[0.007]	[0.007]
Log (GDP)	-0.013	0.145**	0.162**	0.175**
	[0.011]	[0.058]	[0.064]	[0.077]
mean (promotion)		0.14		
mean (OCP performance)		4.3		
Observations	1660	1660	1660	1660
R-squared	0.079	0.556	0.561	0.564
Year FE	X	X	X	X
Person FE		X	X	X
City FE			X	X
Prefecture-year controls				X

Notes: OCP performance reported is birth rate target minus reported birth rate. Tenure fixed effects are controlled for in all regressions. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 3: Heterogeneous Effects of Reported OCP Performance on Promotion by Signal Noise

	Promotion=1		
	(1)	(2)	(3)
OCP performance reported	0.014**	0.062**	0.059**
	[0.007]	[0.028]	[0.027]
OCP performance reported*SD of actual birth rate		-0.009*	-0.009*
		[0.005]	[0.005]
Log(GDP)			0.169**
			[0.077]
Observations	1,660	1,660	1,660
R-squared	0.562	0.564	0.566
Year FE	X	X	X
Person FE	X	X	X
City FE	X	X	X
Prefecture-year controls	X	X	X

Notes: OCP performance reported is birth rate target minus reported birth rate. Tenure fixed effects are controlled for in all regressions. SD of actual birth rate is measured at the province-level and is absorbed by city fixed effects. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 4: Heterogeneous Effects of Reported OCP Performance on Promotion by Competitiveness

	Promotion=1		
	(1)	(2)	(3)
OCP performance reported	0.014**	0.082**	0.084**
	[0.007]	[0.040]	[0.039]
OCP performance reported*Competitiveness		-0.014*	-0.015*
		[0.008]	[0.008]
Log(GDP)			0.179**
			[0.077]
Observations	1,660	1,660	1,660
R-squared	0.561	0.563	0.566
Year FE	X	X	X
Person FE	X	X	X
City FE	X	X	X
Prefecture-year controls	X	X	X

Notes: OCP performance reported is birth rate target minus reported birth rate. Tenure fixed effects are controlled for in all regressions. Competitiveness at the province-level is measured by average tenure of upper-level officials and is absorbed by city fixed effects. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 5: Instrument Variable Regression of Promotion on Reported OCP Performance

	OCP performance reported	Promotion=1	
	(1)	(2)	(3)
	First sage	OLS	2SLS
OCP performance reported		0.013**	0.021*
		[0.007]	[0.012]
Change in population target	0.533***		
	[0.065]		
Log (GDP)	0.169	0.139*	0.137**
	[0.459]	[0.080]	[0.061]
Kleibergen-Paap rk Wald F statistic			67
Observations	1,492	1,492	1,492
R-squared	0.796	0.578	
Year FE	X	X	X
Person FE	X	X	X
City FE	X	X	X
Prefecture-year controls	X	X	X

Notes: OCP performance reported is birth rate target minus reported birth rate. Tenure fixed effects are controlled for in all regressions. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 6: Effects of Population Audits

	(1) OCP performance from Census - OCP performance reported	(2) OCP performance reported	(3) OCP performance from census	(4) Log (GDP)
Audit Year	-0.335***	0.207	-0.127	0.167***
	[0.125]	[0.130]	[0.170]	[0.021]
Observations	1,535	1,535	1,535	1,535
R-squared	0.844	0.815	0.924	0.984
Cubic Year Trend	X	X	X	X
Person FE	X	X	X	X
City FE	X	X	X	X
Prefecture-year controls	X	X	X	X

Notes: Audit years include the year before the census year in 1990 and 1995, and the year before the national fertility survey in 1988, 1992, and 1997. OCP performance from census is birth rate target minus birth rate from census data. Tenure fixed effects are controlled for in all regressions. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 7: OLS Regression of OCP Performance from Census on Promotion

	Promotion=1			
	(1)	(2)	(3)	(4)
OCP performance from Census	0.005*	0.003	0.005	0.005
	[0.002]	[0.007]	[0.007]	[0.007]
Log (GDP)	-0.015	0.138**	0.140**	0.138*
	[0.011]	[0.060]	[0.063]	[0.077]
Observations	1,538	1,538	1,538	1,538
R-squared	0.073	0.566	0.570	0.574
Year FE	X	X	X	X
Person FE		X	X	X
City FE			X	X
Prefecture-year controls				X

Notes: OCP performance from census is birth rate target minus birth rate from census data. Tenure fixed effects are controlled for in all regressions. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Online Appendices Not For Publication

A Appendix - Theory

A.1 Approximation of the probability that a_i produces the maximum output

We approximate $\Pr(y_i \ is \ \max|e_1, ..., e_N)$ in the following way. (Without an approximation, this expression is intractable. Additionally, we believe it is unlikely our agents are calculating extremely complex probabilities, so an approximation is also more realistic.) Recall that the expected value of the maximal order statistic for N iid draws from an exponential distribution with parameter λ is:

$$E[\varepsilon_{(N)}] = \left(1 + \frac{1}{2} + ... + \frac{1}{N}\right)\frac{1}{\lambda}.$$

We approximate the probability that y_i is the maximal output in the following way:

$$\begin{aligned}\Pr(y_i \ is \ \max|e_1, ..., e_N) &\approx \Pr\left(y_i \geq \frac{1}{N-1}\sum_{j\neq i} y_j + \left(1 + \frac{1}{2} + ... + \frac{1}{N-1}\right)\frac{1}{\lambda}\right) \\ &= \Pr\left(\varepsilon_i \geq \frac{1}{N-1}\sum_{j\neq i} e_j - e_i + \left(1 + \frac{1}{2} + ... + \frac{1}{N-1}\right)\frac{1}{\lambda}\right),\end{aligned}$$

where $\frac{1}{N-1}\sum_{j\neq i} e_j + E[\varepsilon_{(N-1)}]$ is the expected value of the sample maximum of the realized outputs of the other (N-1) agents: $\hat{y}_i = e_i + \hat{\varepsilon}_i$.

Recall that if $\varepsilon \sim \exp(\lambda)$, then the cdf is $F_\varepsilon(x) = 1 - e^{-\lambda x}, \ x > 0$. Thus, this probability is:

$$\begin{aligned}\Pr(y_i \ is \ \max|e_1, ..., e_N) &\approx 1 - F_{\exp(\lambda)}\left(\frac{1}{N-1}\sum_{j\neq i} e_j - e_i + \left(1 + \frac{1}{2} + ... + \frac{1}{N-1}\right)\frac{1}{\lambda}\right) \\ &= e^{-\lambda\left[\frac{1}{N-1}\sum_{j\neq i} e_j - e_i + \left(1+\frac{1}{2}+...+\frac{1}{N-1}\right)\frac{1}{\lambda}\right]}\end{aligned}$$

A.2 Optimality of Tournament Mechanism (Proof of Proposition 1)

The approach of the proof is the following. First, we solve for the first-best effort allocation given any $\omega \in [0, 1]$. Total social welfare given ω is $V_\omega = \omega V_1 + (1-\omega)V_0$, where V_1 denotes total social welfare when $\omega = 1$ and the principal cares purely about maximizing output production, and V_0 denotes total social welfare when $\omega = 0$ and the principal cares purely about promoting the highest-ability agent. We show that $V_1^{PR} > V_1^T$, that is, the piece-rate implements the first-best more efficiently than the tournament when the principal cares only about maximizing total output ($\omega = 1$). Next, we show that $V_0^{PR} < V_0^T$, that is, the tournament is more efficient than the piece-rate when the principal cares only about promoting the highest-ability agent ($\omega = 0$). Since V_ω is continuous and increasing in ω, it follows that a value $\tilde{\omega} \in (0, 1)$ exists such that, when $\omega > \tilde{\omega}$, $V_\omega^{PR} > V_\omega^T$ and the principal uses the piece-rate, and when $\omega < \tilde{\omega}$, $V_\omega^T > V_\omega^{PR}$ and the principal uses the tournament.

Solving for the first-best effort allocation given $\omega \in [0,1]$.

First, we solve for the first-best when $\omega = 1$. $\{e_i^{FB,\omega=1}\}$ maximizes total expected output, net of total cost of effort expended:

$$\max \sum_{i=1}^{N} e_i + \frac{N}{\lambda} - \sum_{i=1}^{N} \frac{1}{a_i}\exp(e_i)$$
$$FOC_{e_i} \quad : \quad 1 - \frac{1}{a_i}\exp(e_i) = 0$$
$$e_i^{FB,\omega=1} \quad = \quad \log(a_i)$$

Thus, total effort in the first-best when $\omega = 1$ is:

$$\sum_{i=1}^{N} e_i^{FB,\omega=1} = \sum_{i=1}^{N} \log(a_i)$$

What about when $\omega = 0$? In this case, $\{e_i^{FB,\omega=0}\}$ maximizes the expected ability of the promoted agent, net of total cost of effort expended:

$$\max E[a_i | i\, promoted] - \sum_{i=1}^{N} \frac{1}{a_i}\exp(e_i).$$

Intuitively, it should be that $e_j^{FB,\omega=0} = 0$ for $j = 2, ..., N$, and $e_1^{FB,\omega=0}$maximizes the following:

$$\max a_1 Pr(1\, promoted | e_1^{FB,\omega=0}) + \frac{1}{N-1}\sum_{j\neq i} a_j (1 - Pr(1\, promoted | e_1^{FB,\omega=0})) - \frac{1}{a_1}\exp(e_1^{FB,\omega=0}).$$

Solving yields:

$$e_1^{FB,\omega=0} = \frac{1}{1-\lambda}\left[\log(a_1) + \log\left(a_1 - \frac{1}{N-1}\sum_{j\neq i} a_j\right) + \log(\lambda) - \left(1 + \cdots + \frac{1}{N-1}\right)\right].$$

Thus, we know that the first-best effort allocation given any $\omega \in [0,1]$ is:

$$e_1^{FB,\omega} = \omega \log(a_1) + (1-\omega)\tfrac{1}{1-\lambda}\left[\log(a_1) + \log\left(a_1 - \tfrac{1}{N-1}\textstyle\sum_{j\neq i} a_j\right) + \log(\lambda) - \left(1 + \cdots + \tfrac{1}{N-1}\right)\right]$$
$$e_j^{FB,\omega} = \omega \log(a_j),\ j = 2, ..., N.$$

Piece-rates are Optimal when $\omega = 1$.

Can the piece-rate or the tournament mechanism implement the first-best when all mayors misreport and $\omega = 1$?

The piece rate mechanism ($\omega = 1$): the principal pays the agent $s(\hat{y}_i) = \alpha \hat{y}_i$.[25]An agent's

[25] We assume that the outside option to any agent of being unemployed is very bad, so we don't worry about IR

effort choice in the "all lie" equilibrium, where an agent who is caught lying gets fired, is characterized by:

$$\max_{e_i} pF + (1-p)\left[\alpha\delta e_i + \frac{\alpha\delta}{\lambda}\right] - \frac{1}{a_i}\exp(e_i).$$

This yields:

$$e_{i,PR}^{lie} = \log((1-p)\delta\alpha) + \log(a_i).$$

Thus, the piece-rate mechanism can induce first-best effort even when all agents are known to misreport:

$$\alpha^{FB} = \frac{1}{(1-p)\delta}.$$

Note that the unobservability of ability is not a problem, since this piece-rate does not depend on a_i. The total cost of effort when the piece rate implements the first-best is thus:

$$\sum_{i=1}^{N} c(e_{i,PR}^{FB}) = N$$

The tournament mechanism ($\omega = 1$): now suppose that the principal incentivizes the production of output y by awarding a bonus $B > 0$ based on reported output. Given that this is the mechanism the principal uses, it is optimal for her to award B to the agent with the highest self-reported output (because the principal wants to reward higher output production).

Note that if a mayor does choose to lie, her optimal lie is $\hat{y}_i = \delta y_i$, given the monitoring framework set by the model. Although the structure of misreporting is the same across agents, the actual reports will be heterogeneous, since y_i is heterogeneous. That is, there exists an optimal degree of inflation which is independent of the individual, but the level of the lie will vary by individual.

Then, what is the set of equilibrium misreports? (Here, we focus on the “all lie” equilibrium–see Appendix A.6 for a discussion of the different equilibria and why we focus on “all lie”.) The agent solves:

$$\max_{e_i} B \Pr(\hat{y}_i > \hat{y_{-i}}) - \frac{1}{a_i}\exp(e_i).$$

Then an agent a_i's expected utility is (see Appendix A.1 for details on the approximation):

$$\begin{aligned} EU_{i\ lies}^{-i\ lies} &\simeq pF + (1-p)\exp\left(-\lambda\left[\frac{1}{N-1}\sum_{j\neq i} e_j - e_i\right]\right) E(p,N)B \\ &\quad +(1-p)p^{N-1}B - \frac{1}{a_i}\exp(e_i), \end{aligned}$$

where $E(p, N)$ denotes the expected probability a_i's error is weakly greater than the maximal order

constraints here. Everybody is risk-neutral, so there's no secondary (insurance) effect of not having a constant transfer as a component of compensation, and we don't need the IR constraint to pin down α_i.

statistic for the error in the population of non-fired mayors:

$$\begin{aligned} E(p,N) &\equiv \left[(1-p)^{N-1}\exp(-\lambda\bar{\varepsilon}_{N-1}) + ... + (1-p)p^{N-2}\exp(-\lambda\bar{\varepsilon}_1)\right] \\ &= \left[(1-p)^{N-1}\exp\left(-\sum_{j=1}^{N-1}\frac{1}{j}\right) + ... + (1-p)p^{N-2}\exp(-1)\right]. \end{aligned}$$

Agent a_i's expected utility breaks down in an intuitive way:

1. The first term addresses the case where a_i gets caught lying, which happens with probability p. If she gets caught, she is fired and receives $F < 0$.

2. The second term is the most complex: it addresses the case where a_i does not get caught, but various subsets of the other $(N-1)$ lying mayors are caught. All possibilities ranging from "none of the other mayors is caught" to "all but one of the other mayors are caught" are addressed in this term.

   The key observation is that the average effort of the non-detected other mayors is always $\frac{1}{N-1}\sum_{j\neq i} e_j$, regardless of how many of the other mayors are not detected. This is because the probability of detection is always p for each lying mayor. A simple example will illustrate. Suppose there are four mayors, a_1, a_2, a_3, a_4. The mayor a_1 calculates the average effort of the pool of non-fired mayors she will face, in the case that one of the other mayors is caught. This means that she might face $\{a_2, a_3\}$, or $\{a_2, a_4\}$, or $\{a_3, a_4\}$. But she faces each of these pools with equal probability. Thus, the average of the average effort in each of these pools is just $\frac{1}{3}a_2 + \frac{1}{3}a_3 + \frac{1}{3}a_4$–but that's just the average ability of the three other mayors.

   Thus, we can "factor out" $e^{-\lambda\left[\frac{1}{N-1}\sum_{j\neq i} e_j - e_i\right]}$.[26]

3. The third term addresses the case where our mayor a_i is not caught, but all the other mayors are caught. Then our mayor is promoted for sure.

4. The fourth term is the cost of effort of a_i from exerting effort e_i.

Then the first-order condition characterizing optimal effort in the "pure lie" scenario is:

$$FOC_{e_i} : (1-p)B\lambda e^{-\lambda\left[\frac{1}{N-1}\sum_{j\neq i} e_j - e_i\right]}\left[(1-p)^{N-1}e^{-\lambda\bar{\varepsilon}_{N-1}} + ... + (1-p)p^{N-2}e^{-\lambda\bar{\varepsilon}_1}\right] = \frac{1}{a_i}e^{e_i}.$$

Solving yields equilibrium effort:

$$\begin{aligned} e_i(i\ lies, others\ lie) &= \frac{(N-1)}{N(1-\lambda)-1}\log(a_i) - \frac{\lambda}{N(1-\lambda)-1}\sum_{j=1}^{N}\log(a_j) \\ &\quad + \log[(1-p)B\lambda] + logE(p,N). \end{aligned}$$

[26] Recall that $\bar{\varepsilon}_k$ is the maximal order statistic for a sample of k iid draws from the error distribution, $\exp(\lambda)$. Recall that $\bar{\varepsilon}_k = \left(1 + ... + \frac{1}{k}\right)\frac{1}{\lambda}$.

Then, the tournament mechanism implements the first-best when all mayors misreport by setting:

$$B^{FB} = \frac{1}{(1-p)\lambda E(p,N)}.$$

The total cost of effort when the tournament implements the first-best is thus:

$$\begin{aligned}\sum_{i=1}^{N} c(e_{i,B}^{FB}) &= \frac{\sum_{i=1}^{N}\left(a_i^N\right)^{\frac{\lambda}{N(1-\lambda)-1}}}{\prod\limits_{i=1}^{N} a_i^{\frac{\lambda}{N(1-\lambda)-1}}} \\ &= \sum_{i=1}^{N}\left[\frac{a_i * \dots * a_i}{a_1 * \dots * a_N}\right]^{\frac{\lambda}{N(1-\lambda)-1}}.\end{aligned}$$

The piece-rate vs. the tournament ($\omega = 1$): the piece-rate is more efficient at implementing the first-best if and only if:

$$\frac{1}{N}\sum_{i=1}^{N}\left[\frac{a_i * \dots * a_i}{a_1 * \dots * a_N}\right]^{\frac{\lambda}{N(1-\lambda)-1}} > 1$$

where

$$\frac{\lambda}{N(1-\lambda)-1} = \begin{cases} -1, & \lambda = 1 \\ \infty, & \lambda = \frac{N-1}{N} \\ 0, & \lambda = 0 \end{cases}.$$

Define a sequence $\{s_i\}_{i=1}^{N}$, where $s_i = a_i^{\frac{\lambda N}{N(1-\lambda)-1}} > 0\,\forall i$. Then the inequality we want to show can be re-expressed as:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{s_i}{\left(\prod_{i=1}^{N} s_i\right)^{\frac{1}{N}}} > 1 \iff$$

$$\frac{1}{N}\sum_{i=1}^{N} s_i > \left(\prod_{i=1}^{N} s_i\right)^{\frac{1}{N}}.$$

But this last holds by the AM-GM inequality, which states that, for any sequence of non-negative real numbers, the arithmetic mean is greater than the geometric mean as long as all the terms are not equal (and holds at equality iff all the terms are equal).

So, when $\omega = 1$, the maximum social welfare achieved the piece-rate dominates the maximum social welfare achieved by the tournament: $V_1^{PR} > V_1^{T}$.

Tournaments are Optimal when $\omega = 0$.

We know that in this case, the principal only cares about maximizing the probability of choosing the highest-ability agent a_1, fixing the level of total cost of effort exertion. To show that the principal uses the tournament in this case, we will show that, fixing any level of total cost of effort exertion, under the tournament mechanism, all the higher ability agents (agents whose ability is above the average) have a higher probability of getting promoted, and all the lower ability agents (agents whose ability is below the avreage) have a lower probability of getting promoted, than under the piece-rate mechanism. Hence, the expected ability of the promoted agent is higher under the tournament mechanism than under the piece-rate mechanism given any level of total cost of effort exertion, and the principal therefore uses the tournament when $\omega = 0$.

Recall that, for any piece-rate α, effort is:

$$e_{i,PR}^{lie} = \log((1-p)\delta\alpha) + \log(a_i)$$

so that the probability that agent a_i is promoted is:

$$Pr(a_i\, highest\, output) \approx exp\left(-\lambda\left[\frac{1}{N-1}\sum_{j\neq i} log(a_j) - log(a_i) + \left(1 + ... + \frac{1}{N-1}\right)\lambda\right]\right)$$

while under the tournament mechanism, for any bonus B, effort is:

$$e_{i,T}^{lie} = \frac{(N-1)}{N(1-\lambda)-1}\log(a_i) - \frac{\lambda}{N(1-\lambda)-1}\sum_{j=1}^{N}\log(a_j) + \log[(1-p)B\lambda] + logE(p,N)$$

so that the probability that agent a_i is promoted is:

$$Pr(a_i\, highest\, output) \approx exp\left(-\lambda\left[\frac{(N-1)}{N(1-\lambda)-1}\left(\frac{1}{N-1}\sum_{j\neq i}\log(a_j) - \log(a_i)\right) + \left(1 + ... + \frac{1}{N-1}\right)\lambda\right]\right)$$

But note that, because $\frac{1}{N-1}\sum_{j\neq i} log(a_j) - log(a_i) < 0$ precisely for the above-average ability agents, and $\frac{1}{N-1}\sum_{j\neq i} log(a_j) - log(a_i) > 0$ precisely for the below-average ability agents, it is the case that:

$$Pr(a_i\, promoted)^T > Pr(a_i\, promoted)^{PR}\, for\, a_i > avg(a_i)$$
$$Pr(a_i\, promoted)^T < Pr(a_i\, promoted)^{PR}\, for\, a_i < avg(a_i)$$

Note that this is *regardless* of the bonus B or the piece-rate α. Thus, for every possible piece-rate α, there exists a bonus B such that total cost of effort exertion is equalized, but the higher ability agents have a higher probability of getting promoted, and the lower ability agents have a lower probability of getting promoted. Therefore, the principal uses the tournament mechanism when $\omega = 0$.

Optimality when $\omega \in (0,1)$.

Given interior $\omega \in (0,1)$, if the principal is choosing between using a piece-rate and using a tournament, the principal will choose to use the piece rate if and only if:

$$\omega V_1^{PR} + (1-\omega)V_0^{PR} > \omega V_1^T + (1-\omega)V_0^T.$$

But $V_\omega = \omega V_1 + (1-\omega)V_0$ is continuous and increasing in ω, and we have shown that $V_1^{PR} > V_1^T$, while $V_0^{PR} < V_0^T$. Thus, there exists a threshold $\tilde{\omega} \in (0,1)$ such that, when $\omega > \tilde{\omega}$, the principal uses the piece-rate , while when $\omega < \tilde{\omega}$,the principal uses the tournament.

A.3 Proof of Proposition 2 and Corollary 1

Let $\omega = 1$. We know from Proposition 1 that the equilibrium compensation scheme is a piece-rate.

Proposition 2 is straightforward. If the principal values output only for production (not for screening), and the equilibrium compensation scheme to reward the birth-rate is the piece-rate, then we should not expect OCP to impact the probability of promotion (since promotion must be based on something that isn't OCP).

There are two possibilities: either the dimension on which the principal promotes is not related to ability at all, or there is a correlation. Proposition 2 addresses the former case, Corollary 1 the latter.

If the dimension on which the principal promotes is not related to ability, then OCP (which is correlated with ability) should not have an impact on the probability of promotion, and this non-effect should not vary by noisiness or competitiveness of environment.

If the dimension on which the principal promotes is related to ability (but isn't OCP), then if we observe that better OCP performance corresponds with increased probability of promotion, it must be that the non-OCP dimension on which promotion is based is positively correlated with ability.

But then we should observe that better OCP performance has a larger positive effect on the probability of promotion in environments that are more noisy, and the effect should not differ by competitiveness of environment. Why is this?

Recall that in the "all lie" equilibrium, the effort exerted by a_i is (see Appendix A.2):

$$e_{i,PR}^{lie} = \log((1-p)\delta\alpha) + \log(a_i)$$

Note that if $(1-p)\delta < 1$, all agents tell the truth. A necessary condition for agents to lie is therefore $(1-p)\delta > 1$ (probability of detection low enough, scope for lying high enough). This isn't sufficient (we also need F, the disutility from being fired, to not be too negative, and so on), but this is the primary condition.

Then, the principal solves:

$$\max_\alpha (1-\alpha)\left(\log[(1-p)\delta\alpha] + \log(a_i) + \frac{1}{\lambda}\right)(1-p).$$

This yields:

$$\alpha^* : 1 - \alpha\left(1 + \frac{1}{\lambda}\right) - \alpha \log[(1-p)\delta\alpha] - \alpha \log(a_i) = 0$$

Then we can calculate the following important comparative statics:

$$\begin{aligned}
\frac{\partial \alpha^*}{\partial a_i} &= -\frac{\alpha}{a_i} \frac{1}{2 + \frac{1}{\lambda} + \log[(1-p)\delta\alpha a_i]} < 0 \\
\frac{\partial \alpha^*}{\partial \lambda} &= \frac{\alpha \frac{1}{\lambda^2}}{\left[1 + \frac{1}{\lambda} + \log[(1-p)\delta\alpha a_i] + \frac{\alpha_i}{\alpha_i} + \log(a_i)\right]} > 0
\end{aligned}$$

Thus, the incentive strength of the piece rate is lower for higher ability mayors, and is lower in noisier environments.

To be specific:

1. *OCP performance and promotion*: There is no reason to think that better reported or actual performance on OCP should affect probability of promotion.

   The principal, by assumption in the set up, does not care what the ability of the promoted mayor is. So, if agents are observed to be promoted, it must be on a dimension other than ability. Of course, this dimension may be positively correlated with ability. Thus, since we observe better OCP performance increasing the probability of promotion, if we are in this world, it must be that the principal is promoting based on a dimension that is positively correlated with ability (where higher ability also improves OCP performance).

2. *Noisiness predictions*: If we think that ability is positively correlated with the dimension on which the principal is promoting, then, in this world, this implies that better OCP performance has a *larger* effect (increase) on probability of promotion in environments that are *more noisy.*

   This is because, both in the "all lie" and the "all true" equilibrium:

$$\begin{aligned}
\frac{\partial e_i^{lie}}{\partial a_i} &= \frac{1}{a_i}\left[1 - \frac{1}{2 + \frac{1}{\lambda} + \log[(1-p)\delta\alpha a_i]}\right] \\
\frac{\partial e_i^{true}}{\partial a_i} &= \frac{1}{a_i}\left[1 - \frac{1}{2 + \frac{1}{\lambda} + \log[\alpha a_i]}\right]
\end{aligned}$$

   But note that both of these expressions are *larger* when $\frac{1}{\lambda}$, the variance of the error, is *larger*. That is, higher ability has a larger positive impact on effort and thus OCP performance when the environment is noisier.

   This contradicts our observation that better OCP performance has a *larger* effect (increase) on probability of promotion in environments that are *less noisy.*

3. *Competition predictions*: In this world, there should be no difference in the effect of decreasing the birth rate on the probability of promotion in more versus less competitive environments, since e_i and the piece rate α depend only on *own* ability, and not the abilities of any other mayor in your region, or on the number of mayors in your region.

This contradicts our empirical observation that the impact of better OCP performance on the probability of promotion does depend on competitiveness.

A.4 Proof of Proposition 3

Now, suppose ω is small enough—the principal cares sufficiently about the ability of the agent she promotes, and she no longer directly observes ability—so that the principal prefers to use the tournament mechanism (the lower ω is, that is, the more that the principal values output as a screening device to identify the highest ability mayor, the more likely the principal is to prefer the tournament mechanism).

Then, by introspection, there are potentially three types of pure strategy equilibria:

1. "Pure lie": all the mayors misreport
2. "Pure truth": all the mayors report truthfully
3. "Partial truth": some mayors misreport, and some mayors report truthfully

Again, we focus on the first equilibrium ("pure lie"), since we observe that everyone misreports in the data. (See Appendix A.6 for a discussion of the other equilibrium possibilites, and the sustainability of the "pure lie" equilibrium we focus on.)

Recall that we have already characterized how a given mayor a_i responds optimally depending on the effort exerted by the other $(N-1)$ mayors (in the analysis of Case 1), which we need to characterize the conditions under which each of these types is supported as an equilibrium (if ever).

$$\begin{aligned} e_i(i\ lies, others\ lie) &= \frac{(N-1)}{N(1-\lambda)-1}\log(a_i) - \frac{\lambda}{N(1-\lambda)-1}\sum_{j=1}^{N}\log(a_j) \\ &\quad + \log[(1-p)B\lambda] + \log E(p,N) \end{aligned}$$

What incentives B does the principal set in this pure lie equilibrium? Suppose it costs the principal $c(B)$ to offer a bonus B, where $c'(B) > 0$.

The principal solves:

$$\max_B \omega \left[\sum_{j=1}^{N} \left(\begin{array}{c} \frac{(N-1)}{N(1-\lambda)-1}\log(a_i) - \frac{\lambda}{N(1-\lambda)-1}\sum_{j=1}^{N}\log(a_j) + \log[(1-p)B\lambda] \\ + \log E(p,N) \end{array} \right) \right]$$
$$+(1-\omega)E[a_i|i\ is\ promoted] - c(B)$$

And:

$$FOC_B : \omega\frac{N}{(1-p)B\lambda}(1-p)\lambda - c'(B) = 0,$$

since $E[a_i|i\ is\ promoted]$ does not depend on B, since $\Pr(a_i|i\ is\ promoted)$ does not depend on B.

Let $c(B) = B$. Then the bonus is: $B^* = \omega N$. Then, given the bonus B, what is the probability that a_i is promoted in this "pure lie" equilibrium, given $e_i(i\ lies, others\ lie)$?

$$\Pr(a_i\ is\ promoted, cheating) = \frac{a_i^{\frac{N-1}{N(1-\lambda)-1}-1} E(p,N)}{\left(\sum_{j=1}^{N} \log a_j\right)^{\frac{\lambda}{N(1-\lambda)-1}}} + p^{N-1}(1-p)$$

Compare this to the probability that a_i is promoted in the model where output is contractible and there is thus no possibility of cheating.

$$\Pr(a_i\ is\ promoted, no\ cheating) = \frac{a_i^{\frac{\lambda}{(1-\lambda)}} \exp\left(-\sum_{j=1}^{N-1} \frac{1}{j}\right)}{\left(\sum_{j=1}^{N} \log a_j\right)^{\frac{\lambda}{N(1-\lambda)}}}$$

(See Appendix A.5 for the analysis of this case when output is contractible and there is thus no cheating.)

Note that the important differences are:

(1) there is an extra element in the "pure lie" probability a_i gets promoted which does not depend on ability at all: $p^{N-1}(1-p)$

(2) there is subsequently less weight in the "pure lie" scenario on the ability term. This is captured by the expected maximal order statistic (because our lying mayor a_i is not always competing against all the other mayors, because some of them will get caught): $E(p, \lambda, N) < \left[e^{-\lambda \bar{\varepsilon}_{N-1}}\right]$.

The main takeaway is that, when output is not contractible and we are in the "pure lie" cheating equilibrium, the probability that any mayor a_i is promoted depends less on her ability a_i than in the "no cheating" scenario.

This is driven home when we look at:

$$E[a_i | i\ is\ promoted] = \sum_{j=1}^{N} \Pr(a_i | i\ is\ promoted) a_i.$$

Then:

$$E^{cheat}[a_i | i\ is\ promoted] = \frac{\sum_{j=1}^{N} a_j^{\frac{N-1}{N(1-\lambda)-1}-1} E(p,N)}{N\left(\sum_{j=1}^{N} \log a_j\right)^{\frac{\lambda}{N(1-\lambda)-1}}} + p^{N-1}(1-p)\sum_{j=1}^{N} a_j$$

$$E^{no\ cheat}[a_i | i\ is\ promoted] = \frac{\sum_{j=1}^{N} a_j^{\frac{\lambda}{(1-\lambda)}} \exp\left(-\sum_{j=1}^{N-1} \frac{1}{j}\right)}{N\left(\sum_{j=1}^{N} \log a_j\right)^{\frac{\lambda}{N(1-\lambda)}}}$$

This just emphasizes *the key difference in effectiveness of the promotion mechanism at identifying the highest ability mayor*: in the cheating equilibrium, the expected ability of the promoted mayor includes an extra term which is not present in the benchmark model: $p^{N-1}(1-p)\sum_{j=1}^{N} a_j$. This is the average ability in expectation, which is proportional to the population average:

$$p^{N-1}(1-p)\sum_{j=1}^{N} a_j \propto \frac{1}{N}\sum_{j=1}^{N} a_j.$$

Note that

$$p^{N-1}(1-p) < \frac{1}{N},\ p \in [0,1].$$

The LHS is maximized at $p = \frac{N-1}{N}$ (first order condition is necessary and sufficient since LHS is concave in p as long as $\frac{N-2}{N} < p$, which holds since $p^* = \frac{N-1}{N}$). Then

$$\left(\frac{N-1}{N}\right)^{N-1}\frac{1}{N} < \frac{1}{N},$$

which holds since $\frac{N-1}{N} < 1$.

That is, *the promotion mechanism in the cheating scenario is closer to simply choosing a mayor at random.* Moreover, note that the promotion mechanism performs the worst (is the most random) for intermediate audit probabilities p: if $p = 0$, so people's lies are completely undetectable as long as they stay within δ of their true output, then even though everyone is lying, the highest ability guys are still exerting the most effort and so their reported lie will still be reasonably likely to be the highest. If $p = 1$, then people are detected for sure if they lie and we are in the truthful equilibrium. *It is when p is low and intermediate that individuals distort the most.*

We find that the following important comparative statics are reflected in our data:

1. Increasing effort increases a_i's probability of promotion:

$$\begin{aligned}\frac{\partial \Pr(a_i \text{ is promoted, cheating})}{\partial e_i} &= \frac{\frac{\partial \Pr(a_i \text{ is promoted,cheating})}{\partial a_i}}{\frac{\partial e_i}{\partial a_i}} \\ &> 0\end{aligned}$$

2. Increasing effort has a larger positive impact on a_i's probability of promotion in less noisy environments (larger λ):

$$\frac{\partial \frac{\partial \Pr(a_i \text{ is promoted,cheating})}{\partial e_i}}{\partial \lambda} > 0$$

3. Increasing effort has a smaller positive impact on a_i's probability of promotion in more competitive environments (larger N):

$$\frac{\partial \frac{\partial \Pr(a_i \text{ is promoted,cheating})}{\partial e_i}}{\partial N} < 0$$

Comparative statics - the details

1. Increasing effort increases a_i's probability of promotion.

Denote:

$$A(N) \equiv \left[\left(\frac{N-1}{N(1-\lambda)-1} - 1 \right) \sum_{j=1}^{N} \log(a_j) - \frac{\lambda}{N(1-\lambda)-1} \right]$$

$$\begin{aligned}
\frac{\partial \Pr(a_i \ \textit{is promoted, cheating})}{\partial e_i} &= \frac{\frac{\partial \Pr(a_i \ \textit{is promoted,cheating})}{\partial a_i}}{\frac{\partial e_i}{\partial a_i}} \\
&= \frac{\left(\sum_{j=1}^{N} \log a_j\right)^{\frac{\lambda}{N(1-\lambda)-1}-1} a_j^{\frac{N-1}{N(1-\lambda)-1}-2} E(p,N)A(N)}{\left(\sum_{j=1}^{N} \log a_j\right)^{\frac{2\lambda}{N(1-\lambda)-1}}} \\
&= \frac{a_j^{\frac{N-1}{N(1-\lambda)-1}-2} E(p,N)A(N)}{\left(\sum_{j=1}^{N} \log a_j\right)^{1+\frac{\lambda}{N(1-\lambda)-1}}} \\
&> \frac{a_j^{\frac{N-1}{N(1-\lambda)-1}-2} E(p,N)\left[\frac{(N-1)\lambda}{N(1-\lambda)-1}\right]}{\left(\sum_{j=1}^{N} \log a_j\right)^{1+\frac{\lambda}{N(1-\lambda)-1}}} \\
&> 0
\end{aligned}$$

where the last inequality holds since $A(N) > \frac{(N-1)\lambda}{N(1-\lambda)-1} > 0$. (Recall that $a_i > 1$ for all i, so that $\log(a_i) > 0$ for all i.)

2. Increasing effort has a larger positive impact on a_i's probability of promotion in less noisy environments (larger λ):

$$\begin{aligned}
\frac{\partial \frac{\partial \Pr(a_i \ \textit{is promoted,cheating})}{\partial e_i}}{\partial \lambda} &= \frac{a_j^{\frac{N-1}{N(1-\lambda)-1}-2} \log(a_j) \frac{N(N-1)}{[N(1-\lambda)-1]^2} E(p,N)A(N)}{\left(\sum_{j=1}^{N} \log a_j\right)^{\frac{2\lambda}{N(1-\lambda)-1}}} + \\
&+ \frac{a_j^{\frac{N-1}{N(1-\lambda)-1}-2} E(p,N)\left[\frac{N(N-1)}{[N(1-\lambda)-1]^2} \sum_{j=1}^{N} \log a_j - \frac{(N-1)}{[N(1-\lambda)-1]^2}\right]}{\left(\sum_{j=1}^{N} \log a_j\right)^{\frac{2\lambda}{N(1-\lambda)-1}}}
\end{aligned}$$

Then this expression is positive since each term is positive:

(a) We know that $A(N) > 0$ from the lower bound established in (1). And, $a_i > 1$ for all i, so that $\log(a_i) > 0$ for all i. Thus, the first term is positive.

(b) $\frac{N(N-1)}{[N(1-\lambda)-1]^2} \sum_{j=1}^{N} \log a_j - \frac{(N-1)}{[N(1-\lambda)-1]^2} = \frac{(N-1)}{[N(1-\lambda)-1]^2} \left(N \sum_{j=1}^{N} \log a_j - 1\right) > 0$, so the second term is positive.

3. Increasing effort has a smaller positive impact on a_i's probability of promotion in more competitive environments (larger N).

There is a question of what ability to assume that the additional $(N+1)st$ mayor has. We characterize an upper bound (below) of the effect of increasing effort on a_i's probability of promotion when competition increases by supposing that $\sum_{j=1}^{N}\log a_j = \sum_{j=1}^{N+1}\log a_j$ (this is an upper bound because this is the weakest possible way in which competition can increase–the additional mayor is of the lowest ability), and we show that this upper bound is negative.

Denote $\log \bar{a} = \sum \log a_j$, $\bar{a} = \sum a_j$.

$$\begin{aligned}
\frac{\partial \frac{\partial \Pr(a_i \ is \ promoted, cheating)}{\partial e_i}}{\partial N} &= \frac{-\log \bar{a}^{1+\frac{\lambda}{N(1-\lambda)-1}} a_j^{\frac{N-1}{N(1-\lambda)-1}-2} E(p,N)A(N)}{\log \bar{a}^{2\left(1+\frac{\lambda}{N(1-\lambda)-1}\right)}} \\
&\times \left[\left[\log(a_j) - \log(\log(a_j))\right] \frac{\lambda}{\left[N(1-\lambda)-1\right]^2} \right] \\
&+ \frac{\frac{\partial E(p,N)}{\partial N} a_j^{\frac{N-1}{N(1-\lambda)-1}-2} E(p,N)A(N)}{\log \bar{a}^{2\left(1+\frac{\lambda}{N(1-\lambda)-1}\right)}} \\
&- \frac{\left(\frac{\lambda}{[N(1-\lambda)-1]^2}\right)(\bar{a}-1)\, a_j^{\frac{N-1}{N(1-\lambda)-1}-2} E(p,N)}{\log \bar{a}^{2\left(1+\frac{\lambda}{N(1-\lambda)-1}\right)}} \\
&< 0
\end{aligned}$$

where the negative relationship holds because each term is negative:

(a) The first term is negative because $\log(a_j) > \log(\log(a_j))$.

(b) The second term is negative because $\frac{\partial E(p,N)}{\partial N} < 0$: the maximal order statistic for the error is larger in larger samples, so the probability that a given a_i's error draw is weakly larger than the maximal order statistic is smaller in larger samples.

(c) $\bar{a} > 1 \Leftrightarrow \sum a_j > 1$ since $a_i > 1$ for all i.

A.5 Model with Contractible Output (no cheating)

The optimization program is:

$$\begin{aligned}
&\max_B \omega E\left[\sum_{i=1}^{N} y_i\right] + (1-\omega)E[a_i | i \ is \ promoted] - c(B) \quad s.t. \\
e_i \ &\in \ \arg\max_{\tilde{e}_i} B \Pr(y_i \ is \ \max | \tilde{e}_i, e_{-i}) - \frac{1}{a_i} e^{\tilde{e}_i} \quad \forall i \in \{1, ..., N\}
\end{aligned}$$

We approximate $\Pr(y_i \ is \ \max | e_1, ..., e_N)$ as in Appendix A.1. Thus, this probability is:

$$\begin{aligned}
\Pr(y_i \ is \ \max | e_1, ..., e_N) &\approx 1 - F_{\exp(\lambda)}\left(\frac{1}{N}\sum_{i=1}^{N} e_i - e_i + \left(1 + \frac{1}{2} + ... + \frac{1}{N}\right)\frac{1}{\lambda}\right) \\
&= e^{-\lambda\left[\frac{1}{N}\sum_{i=1}^{N} e_i - e_i + \left(1+\frac{1}{2}+...+\frac{1}{N}\right)\frac{1}{\lambda}\right]}
\end{aligned}$$

We use this to characterize the optimal effort of each agent i with ability a_i. Recall $(IC)_i$, the incentive compatibility constraint of agent i:

$$\begin{aligned} e_i &\in B\Pr(y_i \ is \ \max|e_1,...,e_N) - \frac{1}{a_i}e^{e_i} \\ FOC_{e_i} &: B\frac{\partial \Pr(y_i \ is \ \max|e_1,...,e_N)}{\partial e_i} - \frac{1}{a_i}e^{e_i} = 0. \end{aligned}$$

The FOC is necessary and sufficient because $c(e_i;a_i)$ is convex in e_i and $\frac{\partial \Pr(y_i \ is \ \max|e_1,...,e_N)}{\partial e_i}$ is concave in e_i (as will shortly be seen), so strict concavity of the agent's objective function is assured.

Hence, we can characterize the first-order conditions which characterize optimal effort choice by each agent:

$$FOC_{e_i} : \frac{1}{a_i}e^{e_i} = B\lambda\left(1-\frac{1}{N}\right)e^{-\lambda\left[\frac{1}{N}\sum_{i=1}^{N}e_i - e_i + \left(1+\frac{1}{2}+...+\frac{1}{N}\right)\frac{1}{\lambda}\right]}.$$

Taking logs and rearranging yields:

$$FOC_{e_i} : (1-\lambda)\,e_i + \lambda\frac{1}{N}\sum_{i=1}^{N}e_i = \log(a_i) + \log(B\lambda) + \log\left(1-\frac{1}{N}\right) - \left(1+\frac{1}{2}+...+\frac{1}{N}\right).$$

Sum over all the FOCs:

$$\sum_{i=1}^{N}e_i^* = \sum_{i=1}^{N}\log(a_i) + N\log(B\lambda) + N\log\left(1-\frac{1}{N}\right) - (1+...+N).$$

Then use this characterization of total effort to solve for individual effort, using the individual FOC_{e_i}:

$$e_i^* = \frac{1}{(1-\lambda)}\log(a_i) - \frac{\lambda}{(1-\lambda)}\frac{1}{N}\sum_{i=1}^{N}\log(a_i) + \log(B\lambda) + \log\left(1-\frac{1}{N}\right) - \left(1+\frac{1}{2}+...+\frac{1}{N}\right).$$

Then we can use this to solve for the optimal bonus B^* set by the principal:

$$\max_B \omega E\left[\sum_{i=1}^{N}y_i\right] + (1-\omega)E[a_i|i \ is \ promoted] - c(B).$$

But note that $E[a_i|i \ is \ promoted]$ does not depend on B: that is, the bonus that the principal sets does not influence the screening quality of the "promote the mayor with the highest output" rule. In other words, a higher bonus B is not *differentially* better or worse than a lower B at screening for ability. The quality of screening (how close $E[a_i|y_i \ is \ \max]$ is to a_1, which is the ideal case) depends only on the exogenously-given distribution of the abilities in the economy, $\{a_1,...,a_N\}$, and how noisy output is given effort. That is, if variance is high (lots of noise), the quality of screening will be lower (because less weight will be placed on a_1 given that y_1 is the maximum output vs. when the principal can be very sure that high effort corresponds to high output (low noise)).This quality depends only on parameters:

$$\begin{aligned} E[a_i | i \textit{ is promoted}] &= \sum_{i=1}^{N} a_i \Pr(i \text{ has max } y_i) \\ &= \sum_{i=1}^{N} a_i e^{-\lambda\left[\frac{1}{(1-\lambda)}\frac{1}{N}\sum_{i=1}^{N}\log(a_i) - \frac{1}{1-\lambda}\log(a_i) + \frac{1}{\lambda}(1+...+\frac{1}{N})\right]}. \end{aligned}$$

Thus, the principal's problem can be re-expressed as:

$$\max_B \omega E\left[\sum_{i=1}^{N}\log(a_i) + N\log(B\lambda) + N\log\left(1-\frac{1}{N}\right) - (1+...+N)\right] + \omega\frac{N}{\lambda} - c(B) \ .$$

The first-order condition for the principal is:

$$FOC_B : \omega\frac{N}{B} = c'(B).$$

If $c(B) = B$, for example, then:

$$B^*(N,\omega) = \omega N.$$

The equilibrium effort exerted by an agent with ability a_i is:

$$e_i^* = \frac{1}{(1-\lambda)}\log(a_i) - \frac{\lambda}{(1-\lambda)}\frac{1}{N}\sum_{i=1}^{N}\log(a_i) + \log(\omega N) + \log(\lambda) + \log\left(1-\frac{1}{N}\right) - \left(1+\frac{1}{2}+...+\frac{1}{N}\right).$$

Important comparative statics are:

1. The partial derivative of equilibrium effort in ability:

$$\frac{\partial e_i^*}{\partial a_i} = \frac{1}{1-\lambda}\frac{1}{a_i}\left(1-\frac{\lambda}{N}\right) > 0$$

   Essentially, equilibrium effort monotonically increases in ability as long as $\lambda \in (0,1)$ (looking at the expression for e_i^*, one can see that for $\lambda > 1$, the expression for equilibrium effort becomes negative). But recall that we have imposed the parametric assumption $\lambda \in (0,1)$.

2. The partial derivative of equilibrium effort in N (assume that $\frac{1}{N}\sum_{i=1}^{N}\log(a_i) = \frac{1}{N+1}\sum_{i=1}^{N+1}\log(a_i) \equiv \bar{a}$, so that in expectation the $(N+1)st$ agent has average ability–this is the most logical way to analyze the effect of increased competition):

$$\frac{\partial e_i^*}{\partial N} = \frac{1}{2N} + \frac{1}{N(N-1)} + \frac{1}{N^2} > 0$$

   Individual effort increases as competition increases.

   Note that if we are considering the addition of an $(N+1)st$ agent with a *specific* ability–say, a very high ability agent–than the lowest ability agents may decrease their effort in equilibrium, because the average ability rises: $\frac{1}{N+1}\sum_{i=1}^{N+1}\log(a_i) > \frac{1}{N}\sum_{i=1}^{N}\log(a_i)$, and this increase depresses the effort of the lowest ability agents by the most. It is possible that the added agent has such high ability that all of the original N agents decrease their effort levels. Similarly, if the $(N+1)st$ added agent is known to have very low ability, so low that he lowers the average ability of the new population, and all of the original N agents increase their effort.

3. Individual effort is:

$$e_i^* = \frac{1}{(1-\lambda)}\log(a_i) - \frac{\lambda}{(1-\lambda)}\frac{1}{N}\sum_{i=1}^{N}\log(a_i) + \log(\omega N) + \log(\lambda) + \log\left(1-\frac{1}{N}\right) - \left(1+\frac{1}{2}+...+\frac{1}{N}\right)$$

Then:

$$\begin{aligned}\frac{\partial e_i^*}{\partial \lambda} &= \frac{1}{(1-\lambda)^2}\log(a_i) - \frac{1}{(1-\lambda)^2}\frac{1}{N}\sum_{i=1}^{N}\log(a_i) + \frac{1}{\lambda} \\ &= \frac{1}{(1-\lambda)^2}\left[\log(a_i) - \frac{1}{N}\sum_{i=1}^{N}\log(a_i)\right] + \frac{1}{\lambda}\end{aligned}$$

Higher λ implies lower variance ($V(\varepsilon) = \frac{1}{\lambda^2}$). Thus, for the high ability agents, that is, those agents who have above average ability, the less variance there is, the more effort they exert. On the other hand, for the low ability agents, that is, those agents who have substantially below average ability, the less variance there is, the less effort they exert.

4. Average effort is:

$$\begin{aligned}\bar{e} &= \frac{1}{N}\sum_{i=1}^{N} e_i^* \\ &= \frac{1}{N}\sum_{i=1}^{N}\log(a_i) + \log(\omega N) + \log(\lambda) + \log\left(1-\frac{1}{N}\right) - \left(1+\frac{1}{2}+...+\frac{1}{N}\right)\end{aligned}$$

Then:

$$\begin{aligned}\frac{\partial \bar{e}}{\partial \lambda} &= \frac{1}{\lambda} \\ &> 0\end{aligned}$$

Thus, the lower the variance of the noise factor, the higher the average effort exerted in equilibrium.

5. The equilibrium bonus is:

$$B = \omega N$$

It's straightforward to observe that B is increasing in ω and in N.

6. The probability that agent i with ability a_i is promoted is:

$$\Pr(i\ is\ promoted) = e^{-\lambda\left[\frac{1}{N(1-\lambda)}\sum_{i=1}^{N}\log(a_i)-\frac{1}{(1-\lambda)}\log(a_i)+\left(1+\frac{1}{2}+...+\frac{1}{N}\right)\frac{1}{\lambda}\right]}$$

Then:

$$\begin{aligned}\frac{\partial \Pr(i\ is\ promoted)}{\partial a_i} &= \frac{\lambda}{1-\lambda}\frac{1}{a_i}\left(1-\frac{1}{N}\right)e^{-\lambda\left[\frac{1}{N(1-\lambda)}\sum_{i=1}^{N}\log(a_i)-\frac{1}{(1-\lambda)}\log(a_i)+\left(1+\frac{1}{2}+...+\frac{1}{N}\right)\frac{1}{\lambda}\right]} \\ &> 0\end{aligned}$$

since $N > 1$.

7. The marginal impact of increased effort from agent a_i on the probability that a_i gets promoted is smaller when output is a noisier signal of effort and ability, that is, the variance of the noise factor is high. (The same is true of the average of the marginal impacts.)

Then:

$$\begin{aligned}\frac{\partial\frac{\partial \Pr(y_i\ \max|e_i,e_{-i})}{\partial e_i}}{\partial\lambda} &= \frac{\partial\lambda\left(1-\frac{1}{N}\right)e^{-\lambda\left[\frac{1}{1-\lambda}\frac{1}{N}\sum_{i=1}^{N}\log a_i-\frac{1}{1-\lambda}\log a_i+\left(1+\frac{1}{2}+...+\frac{1}{N}\right)\frac{1}{\lambda}\right]}}{\partial\lambda} \\ &= \left[\left(1-\frac{1}{N}\right)+\lambda\left(1-\frac{1}{N}\right)\left(\frac{1}{(1-\lambda)^2}\frac{1}{N}\sum_{i=1}^{N}\log a_i+\frac{1}{(1-\lambda)^2}\log a_i\right)\right]e^{-\lambda[...]} \\ &> 0\end{aligned}$$

Recall that higher λ means *less noise*, since $Var(\varepsilon) = \frac{1}{\lambda}$. Hence, as output becomes a more and more precise signal of effort/ability, the marginal impact of increasing effort on prob. of promotion increases.

In other words, we should observe a gap between actual and target OCP being more predictive of promotion in low variance vs. high variance places.

Clearly, this property also holds for the average marginal impact of increasing effort on probability of promotion:

$$\begin{aligned}\frac{\partial\frac{1}{N}\sum\frac{\partial \Pr(y_i\ \max|e_i,e_{-i})}{\partial e_i}}{\partial\lambda} &> 0 \\ \text{since } \frac{\partial \Pr(y_i\ \max|e_i,e_{-i})}{\partial e_i} &> 0 \text{ for each } i\end{aligned}$$

8. The marginal impact of increased effort from agent a_i on the probability that a_i gets promoted is smaller when there is more competition. (The same is true of the average of the marginal impacts.)

Assume that there are N mayors who are candidates for promotion, and we add an $(N+1)^{th}$ mayor. We ask: is the marginal impact of increasing effort on the probability of promotion less for each of the i mayors when there is more competition?

Assume that the $(N+1)^{th}$ mayor exerts average effort, which is the rational assumption to make ex ante when the ability of the $(N+1)^{th}$ mayor is not known. Thus, $\frac{1}{N}\sum e_i = \frac{1}{N+1}\sum e_{i+1} \equiv \bar{e}$.

$$\begin{aligned}
\frac{\partial \Pr(y_i \ \max|e_i,e_{-i})}{\partial e_i}(N+1) - \frac{\partial \Pr(y_i \ \max|e_i,e_{-i})}{\partial e_i}(N) &= \lambda e^{-\lambda\left[\bar{e}-e_i+\left(1+\frac{1}{2}+\ldots+\frac{1}{N}\right)\frac{1}{\lambda}\right]}\left(e^{-\frac{1}{N+1}}-1\right) \\
&\quad -\frac{\lambda}{N+1}e^{-\lambda\left[\bar{e}-e_i+\left(1+\frac{1}{2}+\ldots+\frac{1}{N}+\frac{1}{N+1}\right)\frac{1}{\lambda}\right]} \\
&\quad +\frac{\lambda}{N}e^{-\lambda\left[\bar{e}-e_i+\left(1+\frac{1}{2}+\ldots+\frac{1}{N}\right)\frac{1}{\lambda}\right]} \\
&< e^{-\lambda\left[\bar{e}-e_i+\left(1+\frac{1}{2}+\ldots+\frac{1}{N}+\frac{1}{N+1}\right)\frac{1}{\lambda}\right]}\lambda\left(\left(e^{-\frac{1}{N+1}}-1\right)-\frac{1}{N+1}+\frac{1}{N}\right)
\end{aligned}$$

where the upper bound follows from $e^{-\lambda\left[\bar{e}-e_i+\left(1+\frac{1}{2}+\ldots+\frac{1}{N}+\frac{1}{N+1}\right)\frac{1}{\lambda}\right]} < e^{-\lambda\left[\bar{e}-e_i+\left(1+\frac{1}{2}+\ldots+\frac{1}{N}\right)\frac{1}{\lambda}\right]}$.

But:

$$\begin{aligned}
&e^{-\lambda\left[\bar{e}-e_i+\left(1+\frac{1}{2}+\ldots+\frac{1}{N}+\frac{1}{N+1}\right)\frac{1}{\lambda}\right]}\lambda\left(\left(e^{-\frac{1}{N+1}}-1\right)-\frac{1}{N+1}+\frac{1}{N}\right) = \\
&e^{-\lambda\left[\bar{e}-e_i+\left(1+\frac{1}{2}+\ldots+\frac{1}{N}+\frac{1}{N+1}\right)\frac{1}{\lambda}\right]}\lambda\left(\frac{1}{N(N+1)}+e^{-\frac{1}{N+1}}-1\right) < 0
\end{aligned}$$

since

$$\begin{aligned}
\frac{1}{6}+e^{-\frac{1}{3}}-1 &< 0 : \ (N=2) \\
\frac{\partial\left[\frac{1}{N(N+1)}+e^{-\frac{1}{N+1}}-1\right]}{\partial N} &> 0 \\
\lim_{N\to\infty}\frac{1}{N(N+1)}+e^{-\frac{1}{N+1}}-1 &= 0
\end{aligned}$$

Hence, the more competitive the region, the less predictive better performance in output production should be of promotion.

The same property holds for the average, since it holds for each individual marginal impact.

A.6 Equilibrium Possibilities Other Than "All Lie"

Suppose that a_i anticipates that all the other mayors will lie. What is her expected utility from exerting some effort e_i and reporting truthfully?

In this case, a_i solves:

$$\max_{e_i}\exp\left(-\lambda\left[\frac{1}{N-1}\sum_{j\neq i}\delta e_j - e_i\right]\right)E(p,\lambda,N)B + p^{N-1}B - \frac{1}{a_i}e^{e_i}$$

Then the first-order condition is:

$$FOC_{e_i} : (1-\lambda)e_i = \log(a_i) + \log(B\lambda) + \log E(p,\lambda,N) - \frac{\lambda}{(N-1)}\sum_{j\neq i}\delta e_j$$

Thus, the equilibrium effort exerted by a_i when she anticipates that the other agents will all exert $\{e_j\}_{j\neq i}$and misreport, but she tells the truth, is:

$$e_{i\ truth}^{others\ lie} = \frac{1}{1-\lambda}\left[\log(a_i\lambda) + \log(B) + \log(E(p,N)) - \frac{\lambda}{N-1}\sum_{j\neq i}\delta e_j\right]$$

Recall that the effort she exerts when she also chooses to misreport is:

$$e_{i\ lie}^{others\ lie} = \frac{1}{1-\lambda}\left[\log(a_i\lambda(1-p)) + \log(B) + \log(E(p,N)) - \frac{\lambda}{N-1}\sum_{j\neq i}\delta e_j\right]$$

Note that $e_{i\ lie}^{others\ lie} < e_{i\ truth}^{others\ lie}$, since the only difference is the $(1-p) < 1$ in the first term of the expression characterizing $a_i's$ effort when she also misreports.

Then, her expected utility when she chooses to be truthful and her expected utility when she also chooses to misreport are described by:

$$EU_{i\ truth}^{others\ lie} = e^{-\frac{\lambda}{1-\lambda}\frac{1}{N-1}\sum_{j\neq i}\delta e_j} a_i^{\frac{\lambda}{1-\lambda}} \lambda^{\frac{\lambda}{1-\lambda}} E(p,N)^{\frac{1}{1-\lambda}} B^{\frac{1}{1-\lambda}}(1-\lambda) + p^{N-1}B$$

$$EU_{i\ lie}^{others\ lie} = e^{-\frac{\lambda}{1-\lambda}\frac{1}{N-1}\sum_{j\neq i}e_j} a_i^{\frac{\lambda}{1-\lambda}} \lambda^{\frac{\lambda}{1-\lambda}} E(p,N)^{\frac{1}{1-\lambda}} B^{\frac{1}{1-\lambda}}(1-\lambda) + (1-p)p^{N-1}B + pF$$

An "all lie" equilibrium is therefore maintained when $EU_{i\ lie}^{others\ lie} > EU_{i\ truth}^{others\ lie}$ for every i:

$$p^N B - pF < \left[(1-p)^{\frac{1}{1-\lambda}} e^{-\frac{\lambda}{1-\lambda}\frac{1}{N-1}\sum_{j\neq i}e_j} - e^{-\frac{\lambda}{1-\lambda}\frac{1}{N-1}\sum_{j\neq i}\delta e_j}\right] a_i^{\frac{\lambda}{1-\lambda}} \lambda^{\frac{\lambda}{1-\lambda}} E(p,N)^{\frac{1}{1-\lambda}} B^{\frac{1}{1-\lambda}}(1-\lambda)$$

The discussion of comparative statics and the conditions under which the "all lie" equilibrium will be sustained is in the text.

B Data Appendix

B.1 Promotion data

To construct a comprehensive database on the promotion of mayors, we have gone through extensive searches for records of Chinese officials at and above the prefecture level. We first match the name list of mayors with the name lists of the potential positions they could be promoted to. We collected the following complete lists of officials in office during 1985-2000:

1. List of prefecture party secretaries: from *the History of Party Organizations* published by each provincial party office;

2. List of provincial governor or vice-governor, Party secretary or vice-secretary, Party committee member, chairman or vice-chairman of the People's Political Consultative, chairman or vice-chairman of the People's Congress: from *the History of Party Organizations* by province and *Who is Who in China*.

3. List of minister or vice-minister of central ministries: from *Who is Who in China*.

The matching algorithm is straightforward. There is no instance where two mayors have the same Chinese name. We use the unique Chinese name and the term year to match. If a mayor's name is matched with the name of an official in any of these higher ranked positions listed above, after his/her term as mayor, he/she is promoted.

We have also searched for resumes of mayors. We double checked the completeness of our matching from their working experiences. If one is not promoted, we learn from the resume where they move to next. Data on the demographic characteristics of mayors are also compiled from their resumes, such as age, gender, education, province and prefecture of birth, etc.

B.2 Measuring birth rate from census data

First, in the 2000 Census, migrants who moved in 1995-2000 reported the prefecture they moved from. We use the information on out-migration and in-migration by prefecture and year in constructing birth rate measures in 1995-2000. We find that ignoring migration leads to underestimation of birth rate in 1995-2000. The average birth rate from 1995-2000 accounting for migration is 4.8 (per 1000 population), while it is 4.3 without considering migration. Second, in 1985-1994 when we do not observe migration by prefecture and year, the migration rate was below 4%. The potential underestimation without accounting for precise migration information would be much lower. Nevertheless, we include a set of controls on migration in these earlier years in our estimation. Specifically, we control for interactions of average migration measures in 1990-1994 and 1985-1990 and time fixed effects. See the details of the measurement below.

Birth rates in 1995-2000 from the 2000 Census

Migration rate in 1995-2000 is relatively high in the period 1985-2000, rising from 4% in 1995 to 11% in 2000. For these five years, we observe the prefecture-by-year migration information in the

2000 Census that we use to construct birth rate. We use the following formula to compute birth rate:

$$Brate_{cpt} = \frac{Births^1_{cpt} + Births^2_{cpt}}{Population_{cpt+1} - (Births^1_{cpt+1} + Births^2_{cpt+1}) + Outmigrants_{cpt+1} - Inmigrants_{cpt+1}}$$

Where c denotes the prefecture, p the province, and t the year. $Brate_{cpt}$ is the ratio of the number of births in prefecture c of province p in year t to the end-of-year population in the same prefecture. In the numerator, $Births^1_{pct}$ is the number of births born in prefecture c in year t who are in prefecture c in 2000, and $Births^2_{pct}$ is number of birth born in prefecture c in year t who moved out of prefecture c. The sum of $Births^1_{pct}$ and $Births^2_{pct}$ is the total number of births born in prefecture c in year t. In the denominator, the end-of-year population in prefecture c in year t is computed by the population in year $t+1$ subtract the number of births in $t+1$, plus migrants who moved out of prefecture c in year $t+1$, and subtract migrants who moved into prefecture c in year $t+1$. $Population_{cpt+1}$ is computed retrospectively. Starting with the number of population in 2000, we compute population in 1999 based on the denominator, and then population in 1998, and so on.

Birth rates in 1990-1994 from the 2000 Census and in 1985-1989 from the 1990 Census

Migration rate before 1995 is below 4%. For 1985-1994, migration information in the census is limited. The 2000 Census did not have the prefecture-by-year migration information for migrants who moved to the current prefecture before 1995, and it is the same for migrants in the 1990 Census. In 1985-1994, we are not able to use the exact migration information in computing birth rate. See the formula for 1985-1994 below:

$$Brate_{cpt} = \frac{Births^1_{cpt}}{Population_{cpt+1} - Births^1_{cpt+1}}$$

Nevertheless, we use available information from census to control for migration in our estimation. In the 2000 Census, migrants who moved to the current prefecture before 1995 reported the province they moved from. In the 1990 Census, migrants reported the province they moved from since 1985. We construct two sets of aggregate migration measures at the prefecture level:

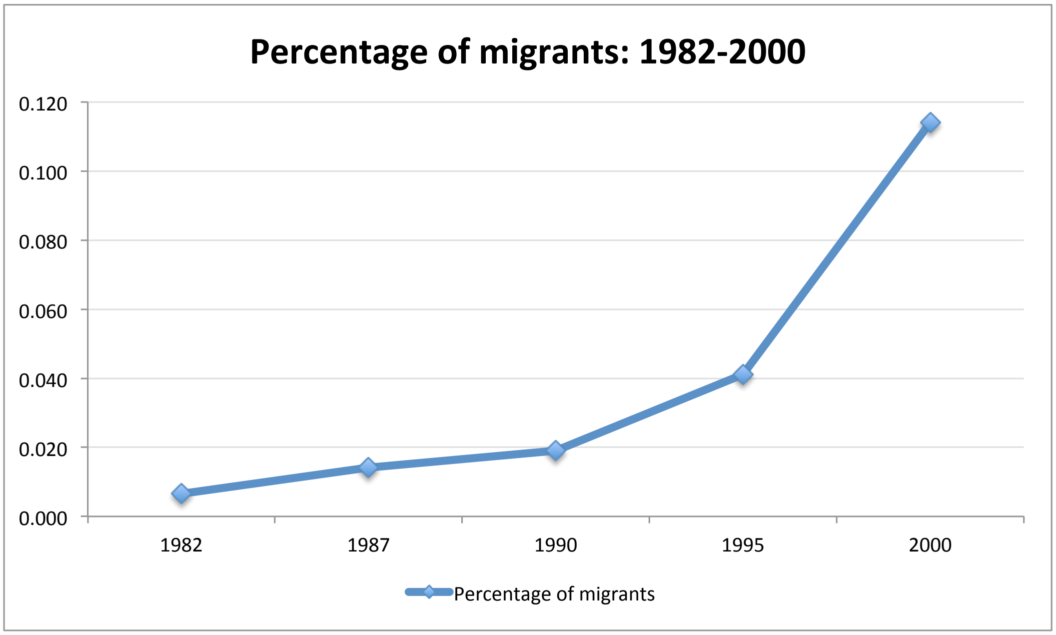
1) The number of out-migrants by the province they moved out in 1990-1994 and in 1985-1989, respectively

2) The number of in-migrants by the current province in 1990-1994 and 1985-1989, respectively

To control for migration in these years, we include interactions of each aggregate measure at the prefecture or province level interacted with time dummy, for example, the number of out-migrants in province p in 1990-1994 interacted with a dummy indicating the time period 1990-1994.

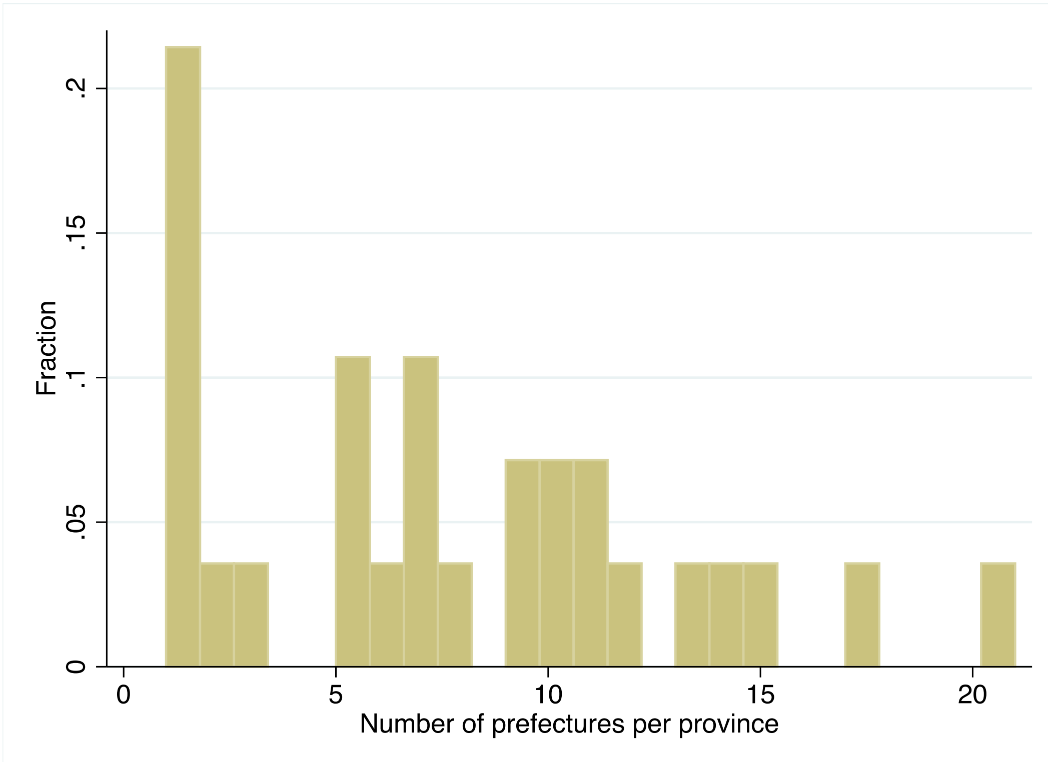
C Additional Figures

Figure A.1: Percentage of Migrants in 1982-2000



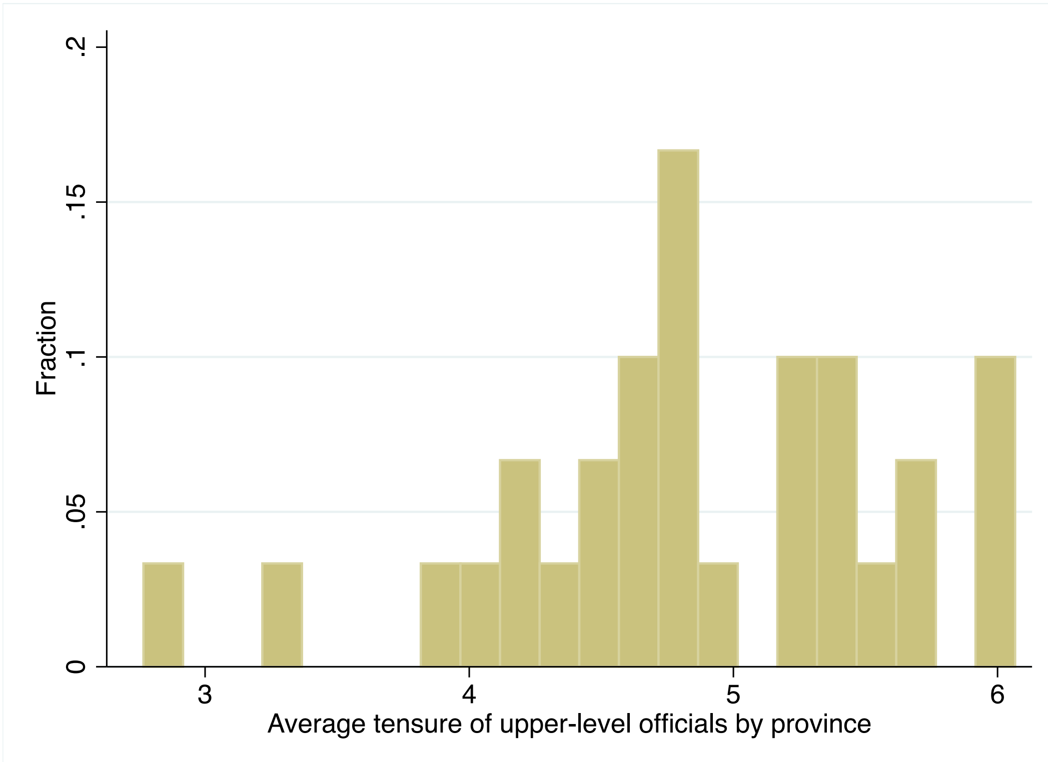
Notes: Data are from population census 1990 and 2000.

Figure A.2: Number of Prefectures Per Province



Notes: Provinces in our mayor data that have 1 prefectures include province-level prefectures (Beijing, Shanghai, Tianjin and Chongqing), and Hainan and Xinjiang province.

Figure A.3: Competitiveness Measure: Average Tenure of Upper-Level Officials



Notes: Data are from digitized term information of all province-level officials ranked higher than mayors. The competitiveness measure is across provinces.

D Additional Tables

Table A.1: Number of Prefectures with Mayor Data and with Birth Rate Data

Year	Number of prefectures	
	(1) Mayor data	(2) Birth rate data
1986	182	151
1987	182	148
1988	191	0
1989	200	160
1990	203	162
1991	204	163
1992	209	164
1993	208	161
1994	214	168
1995	218	173
1996	219	170
1997	221	170
1998	220	171
1999	226	199
2000	192	170

Table A.2: OCP Performance Reported and Promotion: Full Sample

	Promotion=1			
	(1)	(2)	(3)	(4)
OCP performance reported	0.0004	-0.0008	0.0003	0.0006
	[0.002]	[0.004]	[0.004]	[0.004]
Log (GDP)	0.0025	0.0495	0.0912	0.1155*
	[0.008]	[0.040]	[0.056]	[0.064]
Observations	2,160	2,160	2,160	2,160
R-squared	0.050	0.497	0.504	0.506
Year FE	X	X	X	X
Person FE		X	X	X
City FE			X	X
Prefecture-year controls				X

Notes: OCP performance reported is birth rate target minus reported birth rate. Tenure fixed effects are controlled for in all regressions. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table A.3: Test of IV Identifying Assumptions

Panel A	**Stay in position after the change in target=1**	
	(1)	(2)
	Increase in birth rate target	Decrease in birth rate target
OCP performance reported	0.013	0.003
	[0.011]	[0.010]
Log(GDP)	0.031	0.078***
	[0.032]	[0.029]
Observations	356	414
R-squared	0.051	0.065

Panel B	**Promotion=1**
Next change in birth rate target	-0.016
	[0.010]
Log(GDP)	0.120
	[0.102]
Observations	989
R-squared	0.587

Panel C	**Birth rate from census - birth rate reported**
Change in birth rate target	0.006
	[0.045]
Log(GDP)	0.049
	[0.421]
Observations	1,260
R-squared	0.668

Notes: OCP performance reported is birth rate target minus reported birth rate. OCP performance from census is birth rate target minus birth rate from census data. In Panel A, column (1) uses the subsample in provinces and years with an increase in birth rate target, and column (2) uses the subsample in provinces and years with an decrease in birth rate target. In Panel B, we use the sample of 1986-1995, during which the next change in birth rate target is observed in the data. Tenure FE are controlled for in all regressions. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table A.4: 2SLS Estimates: Heterogeneity by Noisiness

	Promotion=1		
2SLS	(1)	(2)	(3)
OCP performance reported	0.022*	0.135***	0.132***
	[0.012]	[0.050]	[0.049]
OCP performance reported*SD of actual birth rate		-0.022**	-0.022**
		[0.009]	[0.009]
Log(GDP)			0.122**
			[0.062]
Kleibergen-Paap rk Wald F statistic	71	20	20
Observations	1490	1490	1,490
Year FE	X	X	X
Person FE	X	X	X
City FE	X	X	X
Prefecture-year controls	X	X	X

Notes: OCP performance reported is birth rate target minus reported birth rate. In column 2 and 3, OCP performance and OCP performance*SD of actual birth rate are instrumented by change in birth rate target and change in target*SD of actual birth rate. Tenure fixed effect are controlled for in all regressions. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table A.5: 2SLS Estimates: Heterogeneity by Competitiveness

	Promotion=1		
2SLS	(1)	(2)	(3)
OCP performance reported	0.022*		
	[0.012]		
Low competition*OCP performance reported		0.016	0.015
		[0.012]	[0.012]
High competition*OCP performance reported		0.035	0.037*
		[0.021]	[0.021]
Log(GDP)			0.145**
			[0.062]
Kleibergen-Paap rk Wald F statistic	71	14	14
Observations	1,490	1,490	1,490
Year FE	X	X	X
Person FE	X	X	X
City FE	X	X	X
Prefecture-year controls	X	X	X

Notes: OCP performance reported is birth rate target minus reported birth rate. The instrument for OCP performance is change in birth rate target. Instruments for low competitiveness*OCP performance and high competitiveness*OCP performance are low competitiveness*change in birth rate target and high competitiveness*change in birth rate target. Tenure fixed effect are controlled for in all regressions. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table A.6: When migration is not fully considered in OCP performance and promotion

	Promotion=1					
	OCP performance reported>0			Full sample		
	(1)	(2)	(3)	(4)	(5)	(6)
OCP performance from Census (not accounting for migration in 1995-2000)	0.007 [0.007]	0.007 [0.007]		-0.0004 [0.005]	-0.0004 [0.005]	-0.001 [0.005]
OCP performance from Census (accounting for migration in 1995-2000)			0.006 [0.007]			
Log (GDP)	0.126* [0.076]	0.126* [0.076]	0.125* [0.076]	0.121* [0.063]	0.121* [0.063]	0.122* [0.063]
Observations	1,545	1,545	1,545	1920	1920	1920
R-squared	0.570	0.570	0.570	0.527	0.527	0.527
Migration in 1985-94*Time FE		X			X	
Year FE	X	X	X	X	X	X
Person FE	X	X	X	X	X	X
City FE	X	X	X	X	X	X
Prefecture-year controls	X	X	X	X	X	X

Notes: OCP performance from census is birth rate target minus birth rate from census data. Tenure fixed effect are controlled for in all regressions. Migration controls are included in all regressions. Prefecture-year controls include log(population), percentage of urban population and log(government investment). Standard errors are clustered at province-year level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.