Fraud-proofing Beyond Election Monitors: An Institutional Design Approach

Abstract

Electoral fraud happens frequently. An emerging literature focuses on election monitors to counter it. We propose an unexplored complementary approach: institutional design. We build on Rundlett and Svolik's (2016) insight that fraud requires coordination among agents who participate only when they are optimistic about the incumbent's chances. We identify an electoral design that eliminates election fraud and simultaneously preserves the majoritarian outcome, so that the incumbent wins exactly when a majority supports her. In this design, the electorate is divided into near-identical districts; the incumbent wins the election if she wins a super-majority of districts; and she wins a district if she receives the majority of the district's votes. Requiring the incumbent to win a super-majority of districts amplifies her agents' coordination problem by inducing mutual fear that others will abandon the incumbent. We highlight multiple directions for future research on fraud-proof institutional design.

Keywords: Election Fraud, Fraud-proofing, Institutional Design, Global Games, Mechanism Design

Word Count: 4,000

There are increasing concerns about the integrity of elections across the world (World Bank 2016, pp. 171-2, 2017, pp. 226-7). The literature has focused on monitoring schemes to mitigate the problem (Hyde 2011; Luo and Rozenas 2018; Garbiras-Díaz and Montenegro 2022; Brancati and Penn 2023). But institutional design, which is deeply rooted in Western political thought, is another approach to guaranteeing a well-functioning polity: for instance, some electoral rules better aggregate information (Austen-Smith and Banks 1996), and separation of powers (Persson et al. 1997) and checks and balances (Acemoglu et al. 2013) improve accountability. Taking an institutional-design approach, we show how an appropriately designed electoral system can prevent the coordination necessary to carry out a wide range of electoral fraud.

We build on Rundlett and Svolik's (2016) observations that (1) various forms of electoral manipulation require costly efforts by many agents, (2) the agents' efforts are not directly observable to the incumbent, individual or party, who tries to incentivize the agents, and (3) promised rewards are contingent on the incumbent's victory. In our model, there are many districts, each with many voters and an agent. Agents can engage in degrees of fraud (e.g., degrees of ballot-stuffing or vote-buying) and the incumbent optimally designs reward schemes to solve a collective agency problem involving coordination and screening.

We demonstrate the power of institutional design by focusing on the following class of electoral systems: the incumbent wins the election if she wins T' percent of districts, and she wins a district if she wins T percent of the district's votes. Selecting the risk-dominant equilibrium in the coordination game among the agents, we show that (T, T') with $T = 1/2$ and T' above a threshold $\widehat{T}'$ eliminates electoral fraud *and* preserves the outcome of majority rule absent fraud. The threshold $\widehat{T}'$ is increasing in the incumbent's office-rent and $(T, T') \approx (1/2, 1)$ delivers this outcome for almost all values of the incumbent's rent. Viewed as proof-of-concept, the results hint at a more general relationship between electoral rules and fraud-prevention.[1] In the Conclusion, we discuss the limitations of this result and directions for future research.

[1]In a concurrent paper, Egorov and Sonin (2023) argue that an electoral college system mitigates fraud because (non-strategic) electoral commissions in party strongholds are more likely to allow fraud, and an electoral college system shifts electoral competition to contested districts.

In the Appendix, we identify an error in Rundlett and Svolik's (2016) analysis and highlight its implications for institutional design.

1 Model

There is an incumbent seeking re-election and a continuum of agents indexed by $i \in [0,1]$, each operating in an electoral district i. If the incumbent wins, she receives a payoff $b > 0$. If she loses, her payoff is 0. The incumbent wins an electoral district if and only if she receives at least $T \in (0,1)$ percent of the district's votes. The incumbent wins the election if and only if she wins at least $T' \in (0,1)$ percent of electoral districts. Agent i can take a costly action $a_i \geq 0$ to raise the incumbent's votes in district i. Taking action a_i costs the agent $\alpha_i a_i$, where $\alpha_i > 0$. The incumbent's popularity in district i is $x_i \in \mathbb{R}$. Given a level of popularity x_i and agent i's action a_i, the incumbent's vote share in district i is $L(t_i) \in (0,1)$, where $t_i = a_i + x_i$, and $L(\cdot) : \mathbb{R} \to [0,1]$ is a strictly increasing, differentiable function. We normalize $L(0)$ and make the following assumption to map popularity levels to vote shares.

Assumption 1 $\lim_{x\to\infty} L(x) = 1$, $\lim_{x\to-\infty} L(x) = 0$, *and* $L(0) = 1/2$.

The incumbent's popularity levels x_i across districts are correlated. In particular, $x_i = \theta + \sigma\epsilon_i$, so that θ captures the incumbent's aggregate popularity across districts (national popularity). We assume $\theta \sim G$, and $\epsilon_i \sim F$ are iid and independent of θ; F and G admit smooth log-concave densities f and g on their support; f has bounded support; g has a sufficiently large support (to ensure dominance regions). An agent i observes x_i, but not θ or any ϵ_j. The incumbent observes her total vote share in each district i, $L(t_i)$, but she does not observe θ, x_i, or ϵ_i.

Before the election, the incumbent can commit to a reward scheme to motivate her agents to take costly actions that increase her vote. The incumbent's promised rewards accrue to the agents, and she pays for them, if and only if the incumbent wins the election. Generally, the incumbent could condition each agent's reward on the entire vector of vote shares $(L(t_i))_{i\in I}$. We assume that agent i's reward can depend only on the incumbent's vote share in district i, $L(t_i)$. Since $L(\cdot)$ is strictly increasing, the incumbent's reward scheme takes the form $B_i(t_i) \in [0, b]$ for each i. We require $B_i(t)$ to be weakly increasing in t.[2]

The game proceeds as follows. The incumbent chooses a reward scheme $B = (B_i(\cdot))_{i\in[0,1]}$. Nature determines θ and ϵ_i for all i. Each agent i observes x_i and B, and decides his action a_i. The election outcome is determined, payoffs are received, and the game ends.

2 Analysis

Fix the incumbent's reward scheme $B_i(t_i)$. An agent i's strategy is a mapping from her signal x_i to an action, $a_i(x_i) : \mathbb{R} \to \mathbb{R}_{\geq 0}$. Without loss of generality, we focus on bounded strategies. We focus further on strategies such that $a_i(x_i) + x_i$ is non-decreasing in x_i. The incumbent wins a district i if and only if $a_i(x_i) + x_i \geq T_d \equiv L^{-1}(T) \in \mathbb{R}$.

A single agent's action does not influence the outcome, because there is a continuum of districts. Moreover, the only source of informational heterogeneity among agents is their signals. Thus, an agent i's belief that the incumbent wins depends on his characteristics only through his signal x_i. Let $p(x_i)$ be i's belief that the incumbent wins. Agent i's best response $a_i^*(x_i)$ is

$$a_i^*(x_i) \in \arg\max_{a_i \geq 0} p(x_i) B_i(a_i + x_i) - \alpha_i a_i. \tag{1}$$

Because the incumbent sees t_i, but not x_i, she faces a screening problem intertwined with coordination between agents. The incumbent's problem is

$$\max_{B(\cdot)} E_\theta \left[\mathbb{1}_{\{\text{incumbent wins}\}} \left(b - \int_i \int_{x_i} B(a_i^*(x_i) + x_i)\ pdf(x_i|\theta) dx_i di \right) \right].$$

We first characterize the incumbent's reward scheme. Proofs are in an Online Appendix.

[2] If agents could lower the incumbent's votes, a non-increasing B_i would invite "reverse fraud", leaving the incumbent worse off.

Proposition 1 *Any optimal reward scheme is outcome-equivalent to* $B_i(t_i) = k_i \cdot \mathbb{1}_{\{t_i \geq T_d\}}$*, for some* $k_i \geq 0$*. That is, it is optimal for the incumbent to reward an agent* i *if and only if the incumbent wins district* i*.*

It is wasteful to pay more for $t_i > T_d$ than for $t_i = T_d$. The incumbent could cap the rewards at the T_d level, but this could discourage some types from reaching the winning threshold. If the incumbent raises rewards at the threshold to compensate, this could encourage too much fraud because now lower types are more motivated to contribute. Proposition 1 shows that a step-function reward scheme suffices for optimally balancing these countervailing effects.

Given this reward scheme, agent i's problem in (1) implies $a_i^*(x_i) \in \{0, T_d - x_i\}$. In particular, $a_i^*(x_i) = T_d - x_i$ if and only if both $T_d - x_i \geq 0$ and $p(x_i)\ k_i - \alpha_i(T_d - x_i) \geq 0$, where we assume that agent i takes the highest action when indifferent. If $p(x_i)$ is increasing, the left-hand side is increasing in x_i. Then, there exists a $x_i^* \leq T_d$ such that the above inequality holds for $x_i \geq x_i^*$, where

$$p(x_i^*)\ k_i/\alpha_i + x_i^* = T_d. \tag{2}$$

Moreover,

$$a_i^*(x_i) = \begin{cases} 0 & ; x_i < x_i^* \\ T_d - x_i & ; x_i^* \leq x_i \leq T_d \\ 0 & ; T_d < x_i, \end{cases} \quad \text{so that} \quad a_i^*(x_i) + x_i = \begin{cases} x_i & ; x_i < x_i^* \\ T_d & ; x_i^* \leq x_i \leq T_d \\ x_i & ; T_d < x_i. \end{cases} \tag{3}$$

Consequently, $a_i^*(x_i) + x_i \geq T_d$ if and only if $x_i \geq x_i^*$.

We now verify that $p(x_i)$ is increasing. Given a strategy profile $(a_i(x_i))_{i\in[0,1]}$ and θ, let $m(\theta)$ be the fraction of districts that the incumbent wins:

$$m(\theta) = \int_0^1 Pr(a_i(x_i) + x_i \geq T_d|\theta)di.$$

Because $a_i(x_i) + x_i$ is increasing in x_i, and x_i is FOSD-increasing in θ, $m(\theta)$ is also increasing in θ. Moreover, when θ is sufficiently large, the vote share in all districts is above T, and the incumbent wins. When θ is sufficiently small, because the rewards are bounded, the incumbent will lose. Thus, there exists a unique threshold $\theta^* \in \mathbb{R}$ such that

$$m(\theta^*) = \int_0^1 Pr(a_i(x_i) + x_i \geq T_d | \theta = \theta^*) di = T' \tag{4}$$

and the incumbent wins if and only if $\theta \geq \theta^*$. Thus, belief consistency implies $p(x_i) = Pr(\theta \geq \theta^* | x_i)$, which is increasing in x_i. Thus, any $((x_i^*)_{i \in [0,1]}, \theta^*)$ that satisfies the individual rationality (2) and belief consistency (4) constitutes an equilibrium. We focus on symmetric settings where $\alpha_i = \alpha$ and $k_i = k$, so that $x_i^* = x^*$ for all $i \in [0, 1]$. Using the global games approach to equilibrium selection, we have

Proposition 2 *Suppose $\alpha_i = \alpha$ and $k_i = k$ and focus on the class of strategies with monotone $a_i(x_i) + x_i$. Any equilibrium is characterized by a pair of thresholds (x^*, θ^*). In equilibrium, an agent i engages in election fraud if and only if his signal $x_i \in [x^*(k), T_d)$, and the incumbent wins if and only if $\theta \geq \theta^*$. Moreover, in the limit as $\sigma \to 0$, there is a unique equilibrium with*

$$\lim_{\sigma \to 0} \theta^*(\sigma; k) = \lim_{\sigma \to 0} x^*(\sigma; k) = T_d - (1 - T')k/\alpha$$

where we recall that $T_d = L^{-1}(T)$.

An agent with the threshold signal x^* believes that the fraction of players with signals above x^* is approximately uniformly distributed on $[0, 1]$:

$$Pr(\ Pr(x_j \geq x^* | \theta) \leq p \mid x_i = x^*\ ) \ = \ p.$$

From (3), this fraction is the fraction of districts in which the incumbent wins. Thus, the agent with threshold signal believes that the fraction of districts that the incumbent wins is uniformly distributed on $[0, 1]$. Then this agent believes that the probability that the incumbent wins at least T' percent of districts—and hence the election—is $1 - T'$: $p(x^*) = 1 - T'$. Combined with (2) this yields the limit values of x^* and θ^*. This reveals how the term $1 - T'$ stems from the strategic considerations that arise in the agents' coordination game.

From Proposition 2, in the limit $\theta^*(k) \approx x^*(k)$, so that $t_i \geq T_d$ in almost all the districts whenever $\theta > \theta^*(k)$. Thus, the incumbent's problem becomes

$$\max_{k \in [0,b]} (1 - G(\theta^*(k)))\ (b - k).$$

Proposition 3 characterizes the incumbent's optimal reward.

Proposition 3 *Suppose $a_i(x_i) + x_i$ is monotone, $\alpha_i = \alpha$, and the incumbent uses symmetric rewards. In equilibrium, as $\sigma \to 0$, the incumbent's choice of k converges to $k^* = \max\{0, \hat{k}\}$, where $\hat{k}$ is the unique solution to*

$$\frac{g(\theta^*(k))}{1 - G(\theta^*(k))}\ (1 - T') = \frac{1}{b - k}, \quad k < b.$$

Moreover, $k^* > 0$ if and only if $b > \underline{b} = \frac{1}{1-T'} \frac{1-G(T_d)}{g(T_d)}$.

Example 1. Suppose $\theta \sim U[-l, h]$, with $l, h > 0$ large enough to have dominance regions, and $\alpha = 1$. Then, the FOC becomes $(1 - T')(b - k^*) = h - \theta^*(k) = h - T_d + (1 - T')k^*$, so that

$$k^* = \max\left\{0, \frac{1}{2}\ \left(b - \frac{h - T_d}{1 - T'}\right)\right\}.$$

Substituting from k^* into $\theta^*(k)$ in Proposition 2 yields

$$\lim_{\sigma \to 0} \theta^*(k^*) = \min\left\{\frac{T_d - b(1 - T') + h}{2}, T_d\right\},$$

where the min ensures that $\theta^*(k^*) \leq T_d$.

In the example, setting a higher bar for the incumbent *within each district* increases rewards for fraud. In contrast, increasing *the fraction of districts* that the incumbent must win reduce them: $\partial k^*/\partial T' < 0 < \partial k^*/\partial T_d$, when $k^* > 0$. The intuition relies on coordination. A higher T' hinders coordination, because it induces the marginal agent with the threshold signal x^* into believing that the incumbent is less likely to win, thereby reducing the agent's responsiveness to rewards. This, in turn, reduces the incumbent's marginal gains from raising rewards. Our approach to fraud prevention builds on this observation.

2.1 Fraud-proofing

We now investigate what electoral design, i.e., what (T, T'), minimizes election fraud and whether there are trade-offs. We consider two notions of fraud: the probability that fraud changes election results and the measure of fraudulent votes. The first measure follows from Proposition 2 by recognizing that $k = 0$ captures settings without fraud. Computing the second is more elaborate given the non-monotone equilibrium strategies. Nevertheless, combining (3) with Proposition 2 allows us to calculate the expected measure of fraudulent votes.

Corollary 1 *In the limit as $\sigma \to 0$, the equilibrium probability that fraud changes the election outcome, denoted by PF, is*

$$PF = G(T_d) - G(T_d - (1 - T')k^*/\alpha).$$

The ex-ante measure of fraudulent votes, denoted by MF, is

$$MF = PF \cdot (T_d - \mathbb{E}[\theta | T_d - (1 - T')k^*/\alpha \leq \theta \leq T_d]).$$

Moreover, fraud does not arise in equilibrium according to either measure, i.e., $PF = MF = 0$, if and only if $(1 - T')k^* = 0$.

One way to prevent fraud is to bar the incumbent from running again. A less extreme way is to require a super-majority for the incumbent. But the incumbent may have sufficient popular support to win even without fraud. Thus, we must consider whether and when a policy aimed at preventing fraud also prevents a genuinely more popular incumbent from re-election.

Suppose absent the possibility of fraud, the optimal electoral rule is majoritarian. We first characterize when the incumbent wins under majority rule.

Proposition 4 *Suppose there is no fraud. In the limit as $\sigma \to 0$, the incumbent wins the majority of votes if $\theta > 0$, and loses the majority of votes if $\theta < 0$.*

Can we eliminate or minimize fraud (Corollary 1) while requiring that the incumbent wins whenever she would win absent fraud (Proposition 4)?

Proposition 5 *Fix the incumbent's payoff b from winning the election. There is a threshold $\widehat{T}'(b) \in (0,1)$ such that, in the limit as $\sigma \to 0$, any rule (T, T') with $T = 1/2$ and $T' \geq \widehat{T}'(b)$ eliminates electoral fraud and preserves the outcome of the majority rule absent fraud. Moreover, $\widehat{T}'(b)$ is increasing in b with $\lim_{b\to\infty} \widehat{T}'(b) = 1$.*

The intuition hinges on the coordination incentives of agents. As discussed following Proposition 2, the marginal agent with signal x^* believes that the share of districts that the incumbent will win is uniformly distributed on $[0, 1]$. When the incumbent needs to win more districts to win the election, the marginal agent becomes more pessimistic about the incumbent's chances and hence about his own chances of receiving the promised rewards of fraud. Because it is more expensive to motivate a more pessimistic agent, the incumbent will have less incentives to motivate fraud. In particular, when $T' \approx 1$, there is no fraud for almost all values of b. Moreover, if $T = 1/2$ (i.e., $T_d = 0$), the incumbent wins whenever $\theta > 0$ for sufficiently small σ.

The following example demonstrates what would happen if a different electoral rule is used.

Example 2. Suppose $\alpha = 1$ and $\theta \sim U[l, h]$, with $h > T_d$ and $l < T_d - b$, $b \geq 0$, so that there are dominance regions. From Example 1, there is fraud if and only if

$$k^* > 0 \quad \Leftrightarrow \quad T_d > h - b + bT'.$$

Thus, setting $T_d \leq h - b + bT'$ implies $k^* = 0$. Moreover, $k^* = 0$ implies $\theta^* = T_d$. Thus, the set of (T_d, T') under which there is no fraud *and* the incumbent wins if $\theta > 0$ and loses if $\theta < 0$ becomes $\{(T_d, T')\ s.t.\ T_d = 0,\ T' \geq 1 - h/b\}$. When b is small, setting $T_d = 0$ (corresponding to majority rule within districts) works well: in the extreme case when $b = 0$, an incumbent has no incentive to induce fraud. When b is large, however, in addition to $T_d = 0$, we must also require $T' \geq \widehat{T}'(b) = 1 - h/b$. Importantly $\lim_{b\to\infty} \widehat{T}'(b) = 1$, so that $T'T \approx 0.5$.

A numerical specification further illustrates. Let $L(x) = (Tanh(x) + 1)/2$, $h = 1$, $T' = 0.8$, and $T = 0.625$, so that the incumbent needs at least a share $T'T = 0.5$ of the votes to win. Then, $T_d = L^{-1}(T) \approx 0.26 < h$. In the limit as $\sigma \to 0$, from Example 1,

$$k^* = \max\left\{0, \frac{1}{2}\ \left(b - \frac{1 - 0.26}{0.2}\right)\right\} = \max\{0, 0.5b - 1.85\}.$$

Thus, $k^* > 0$ if and only if $b > 3.7$. And,

$$\theta^*(k^*) = \min\left\{\frac{0.26 - 0.2b + 1}{2}, 0.26\right\} = \min\left\{0.63 - 0.1b, 0.26\right\}.$$

When $b \leq 3.7$, there is no fraud and the incumbent wins whenever $\theta > 0.26$. But then, for $\theta \in (0, 0.26)$, the incumbent loses when she would win with majority rule absent fraud. When $b > 3.7$, the incumbent wins whenever $\theta > 0.63 - 0.1b$. Thus, when $b > 6.3$, for $\theta \in (0.63 - 0.1b, 0)$, the incumbent wins when she would lose under majority rule absent fraud. The ex-ante probability of this event, $\frac{0.1b - 0.63}{1-l}$, is increasing in b.

3 Conclusion

Our analysis demonstrates the value of mechanism design and coordination games in providing policy recommendations for institutional designs to mitigate electoral fraud. We made several simplifying assumptions; the relaxation of each provides a direction for future research. We list them here. (1) We abstracted from heterogeneity among districts in terms of both complete information heterogeneity (e.g., districts can have high and low α_i) and incomplete information heterogeneity (e.g., σ need not be small). (2) We focused on a limited class of mechanisms, among other things, by restricting the incumbent to condition an agent's rewards only on the outcome of that agent's district, and not on the distribution of outcomes across districts. (3) Our analysis suggests that having many districts is helpful for fraud prevention. But this requires dividing a large population into many smaller groups. This seems in tension with Madisonian view, in Federalist 10, about the advantages of a larger electorate. If so, what is the optimal balance? (4) Any reasonably small district is likely too large for a single agent. Within each district, agents' actions are perfect substitutes (as in Rundlett and Svolik (2016)). A more general model could account for coordination both within and between districts.

A Appendix: Revisiting Rundlett-Svolik Model

We now revisit Rundlett and Svolik's (2016) model, correct an error in their first Proposition, and show: when the incumbent's payoff from winning exceeds a threshold, more "heterogeneity" among districts *reduces* the probability that fraud changes the true winner. This suggests that, all else equal, an institutional designer should consider creating electoral districts that are different from each other—this policy may not be feasible or desirable, e.g., due to creating enclaves.

A.1 Model

There is an incumbent and a continuum of agents indexed by $i \in [0,1]$, each operating in an electoral district i. Agents simultaneously decide whether to engage in election fraud. An agent i's action is denoted by $a_i \in \{0,1\}$, where $a_i = 1$ if i engages in fraud, and $a_i = 0$ otherwise. The fraction of votes in district i is $x_i + a_i F$, for some exogenous $F \in (0, 1/2)$. The incumbent's vote shares across the districts are correlated. In particular, $x_i = \theta + \sigma\epsilon_i$, where $\epsilon_i \sim U[-1,1]$, $\theta \sim G$, and θ and ϵ_is are independent from each other. Let $\phi \in [0,1]$ be the fraction of agents who engage in fraud, so that $\phi = \int_0^1 a_i di$. For a given θ, the incumbent's vote share, R, is

$$R(\theta) = \int_0^1 (x_i + a_i F) di = \theta + \phi F. \tag{A1}$$

The incumbent wins if and only if her vote share exceeds 1/2, that is, $\phi \geq \frac{1/2-\theta}{F}$. If an agent engages in fraud, his payoff is $w(x_i + F)$ if the incumbent wins and $-cF$ if the incumbent loses, where $c > 0$. If an agent does not engage in fraud, his payoff is wx_i if the incumbent wins and 0 if the incumbent loses. The left panel of Figure 1 depicts the payoffs. This payoff structure is strategically equivalent to the right panel.

The incumbent ex-ante chooses $w \geq 0$ to maximize her expected payoff. If the incumbent wins, she receives a payoff $b > 0$. If she loses, she receives 0. The incumbent pays wR regardless of the election outcome.

		$\phi \geq \frac{1/2-\theta}{F}$	$\phi < \frac{1/2-\theta}{F}$
agent i	fraud	$wF + wx_i$	$-cF$
	no fraud	wx_i	0

$\phi \geq \frac{1/2-\theta}{F}$	$\phi < \frac{1/2-\theta}{F}$
wF	$-cF$
0	0

Figure 1: Rundlett and Svolik's Fraud Game

The timing of the game is as follows. The incumbent chooses w. Nature draws θ and ϵ_i, $i \in [0,1]$. Each agent i observes w and his signal $x_i = \theta + \sigma\epsilon_i$. Agents simultaneously decide whether to engage in fraud. The election outcome is determined, payoffs are received, and the game ends.

Rundlett and Svolik (2016) assume $G = U[0,1]$, putting aside that the incumbent's vote shares, x_i or $x_i + F$, may fall outside $[0,1]$.

A.2 Analysis

Consider first the game played by the agents given a value of w. Because the game is a standard global game (Morris and Shin 2003), we delegate details to the Online Appendix. The global games approach selects a unique equilibrium when noise is sufficiently small under general smooth prior and noise distributions. Given the uniform distributions of prior and noise, the following assumption suffices.

Assumption A1 $\sigma < \frac{1}{4} - \frac{F}{2}$.

We focus on symmetric monotone equilibria, where each agent i engages in fraud if and only if $x_i \geq x^*$ for some x^*.

Proposition A1 *The agents' coordination game has a unique monotone equilibrium characterized by a pair of thresholds (x^*, θ^*). In equilibrium, an agent i with signal x_i engages in fraud if and only if his signal $x_i \geq x^*$, and the incumbent wins if and only if $\theta \geq \theta^*$. Moreover,*

$$\theta^* = \frac{1}{2} - \frac{w}{w+c}F \quad \textit{and} \quad x^* = \theta^* + \sigma\frac{c-w}{c+w}.$$

Thus, the incumbent's problem is

$$\max_{w\geq 0}(1-G(\theta^*(w)))b - w\int_\theta R(\theta;x^*(w))dG(\theta)$$

where $R(\theta;x^*)=\theta+\phi(\theta;x^*)F$, from (A1), is the total vote share across all districts for a given θ.

Proposition A2 *The optimal* w^* *is*

$$w^* = \max\left\{0, \sqrt{\frac{cF[2cF+2(b+2c\sigma)]}{2F^2+(1+2\sigma)F+1}} - c\right\}.$$

Moreover, $\lim_{\sigma\to 0} w^*(\sigma) > 0$ *if and only if* $b > b^* \equiv \frac{c}{2F}(1+F)$.

Heterogeneity and Fraud Proposition A2 shows that the optimal wage reported in Rundlett and Svolik (2016, p. 186), which we denote by w^*_{RS}, is incorrect. They view σ "as 'small' and interpret it as a measure of heterogeneity in the incumbent's support across precincts" (p. 184). With this interpretation, their Proposition 1 implies that if $w^*_{RS} > 0$, then $\frac{\partial w^*_{RS}(\sigma)}{\partial\sigma} > 0$. Thus, their measure of election fraud, $1/2-\theta^*$, is increasing in σ (Supplementary Appendix A.4). In fact, $\lim_{\sigma\to 0} w^*_{RS}(\sigma) = 0$: lower heterogeneity in the incumbent's support across districts reduces the likelihood that fraud changes the true winner; and when there is little heterogeneity, there will be almost no fraud. Proposition A3 shows that when b is large, the *opposite* holds.

Proposition A3 *Suppose the incumbent's office-rent is sufficiently large that she offers rewards for fraud:* $b > b^*$. *The rewards and the extent of fraud is decreasing in* σ *if and only if the incumbent's office-rent is sufficiently large. Formally, there exists* $\hat{b} > b^*$ *such that* $\frac{\partial w^*}{\partial\sigma}$, $\frac{\partial(1/2-\theta^*)}{\partial\sigma} < 0$ *if* $b > \hat{b}$, *but* $\frac{\partial w^*}{\partial\sigma}$, $\frac{\partial(1/2-\theta^*)}{\partial\sigma} > 0$ *if* $b \in (b^*, \hat{b})$, *where* $\hat{b} = c(1+F+F^2)/F$.

References

Acemoglu, Daron, James Robinson, and Ragnar Torvik. 2013. "Why Do Voters Dismantle Checks and Balances?" *Review of Economic Studies* 80(3): 845-75.

Austen-Smith, David, and Jeffrey Banks. 1996. "Information Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review* 90(1): 34-45.

Brancati, Dawn, and Maggie Penn. 2023. "Stealing an Election: Violence or Fraud?" *Journal of Conflict Resolution* 67(5): 858-92.

Egorov, Georgy, and Sonin, Konstantin. 2023. "Electoral College and Election Fraud." Available: `https://ssrn.com/abstract=4507012`.

Garbiras-Díaz, Natalia, and Mateo Montenegro. 2022. "All Eyes on Them: A Field Experiment on Citizen Oversight and Electoral Integrity." *American Economic Review* 112(8): 2631-68.

Hyde, Susan. 2011. *The Pseudo-Democrat's Dilemma: Why Election Observation Became an International Norm.* Ithaca, NY: Cornell University Press.

Luo, Zhaotian, and Arturas Rozenas. 2018. "The Election Monitor's Curse." *American Journal of Political Science* 62(1): 148-60.

Morris, Stephen, and Hyun Song Shin. 2003. "Global Games: Theory and Applications." In *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress, Volume I,* edited by Dewatripont, Hansen, and Turnovsky. New York: Cambridge University Press.

Persson, Torsten, Gérard Roland, and Guido Tabellini. 1997. "Separation of Powers and Political Accountability." *Quarterly Journal of Economics* 112(4): 1163-1202.

Rundlett, Ashlea, and Milan Svolik. 2016. "Deliver the Vote! Micromotives and Macrobehavior in Electoral Fraud." *American Political Science Review* 110(1): 180-97.

World Bank. 2016. *World Development Report: Digital Dividends.* Washington, DC: World Bank.

World Bank. 2017. *World Development Report: Governance and the Law.* Washington, DC: World Bank.

Online Appendix: Proofs

We first prove a lemma, which will be used in the proof of Proposition 1.

Lemma 1 *Let $(a_j(x_j))_{j\in[0,1]}$ be any strategy profile such that $t_j(x_j)$ is weakly increasing for all $j \in [0,1]$. Then:*

(i) $(a_j(x_j))_{j\in[0,1]}$ induces a winning probability $p(x) := P(\text{incumbent wins}|x_i = x)$ that is continuous; weakly increasing; and strictly increasing wherever $p(x) \in (0,1)$.

(ii) $(a_j(x_j))_{j\in[0,1]}$, paired with any weakly increasing reward scheme $B_i(t_i)$, induces a best response $a_i^*(x_i)$ by agent i such that $t_i^*(x_i) = x_i + a_i^*(x_i)$ is weakly increasing.

Proof of Lemma 1: (i) By assumption, $t_j(x_j) := x_j + a_j(x_j)$ is weakly increasing in x_j for all j. Moreover, $x_j = \theta + \sigma\epsilon_j$ is FOSD-increasing in θ for all j. Thus $t_j = t_j(x_j)$ is weakly FOSD-increasing in θ. In particular, $P(t_j \geq T_d|\theta)$ is weakly increasing in θ for all j. Moreover, $P(t_j \geq T_d|\theta)$ is a continuous function of θ because $P(x_j \geq x|\theta) = 1 - F\left(\frac{x-\theta}{\sigma}\right)$ is continuous in θ for any x. Then the incumbent's share of districts,

$$m(\theta; (a_j(\cdot))_{j\in[0,1]}) := \int_0^1 P(x_j + a_j(x_j) \geq T_d|\theta)dj,$$

is weakly increasing and continuous in θ. Then there is θ^* such that the incumbent wins if and only if $\theta \geq \theta^*$, where θ^* is such that $\int_0^1 P(t_j \geq T_d|\theta = \theta^*)dj = T'$. Then

$$p(x) = P(\theta \geq \theta^*|x_i = x) = \frac{\int_{\theta^*}^{\infty} g(\theta)f\left(\frac{x-\theta}{\sigma}\right)\frac{1}{\sigma}d\theta}{\int_{-\infty}^{\infty} g(\theta)f\left(\frac{x-\theta}{\sigma}\right)\frac{1}{\sigma}d\theta},$$

which is continuous in x. Moreover, because f is log-concave, the density $h(\theta|x_i = x) \propto g(\theta)f\left(\frac{x-\theta}{\sigma}\right)$ is MLRP-increasing in x. Hence $\theta|x_i = x$ is FOSD-increasing in x, so $p(x)$ is weakly increasing in x. Furthermore, because $h(\theta|x_i = x)$ must be increasing in a neighborhood of $\theta = x - \sigma\inf(\text{supp}f)$ and decreasing in a neighborhood of $\theta = x - \sigma\sup(\text{supp}f)$, $p(x)$

must be strictly increasing in a neighborhood of x unless $x - \sigma \inf(\text{supp} f)$ and $x - \sigma \sup(\text{supp} f)$ are on the same side of θ^*, i.e., unless $p(x) = 0$ or $p(x) = 1$.

(ii) Suppose $t_i^*(x_i) > t_i^*(x_i')$ for some $x_i < x_i'$. Then

$$
\begin{aligned}
p(x_i)B_i(t_i^*(x_i)) - \alpha_i(t_i^*(x_i) - x_i) &\geq p(x_i)B_i(t_i^*(x_i')) - \alpha_i(t_i^*(x_i') - x_i) \\
p(x_i')B_i(t_i^*(x_i')) - \alpha_i(t_i^*(x_i') - x_i') &\geq p(x_i')B_i(t_i^*(x_i)) - \alpha_i(t_i^*(x_i) - x_i') \\
\implies [p(x_i) - p(x_i')]\,[B_i(t_i^*(x_i)) - B_i(t_i^*(x_i'))] \geq 0.
\end{aligned}
$$

If $p(x_i) = p(x_i') = 0$, there is no fraud in either case, so $t_i^*(x_i) = x_i < x_i' = t_i^*(x_i')$, a contradiction. If $p(x_i) = p(x_i') = 1$, then the two inequalities imply that $t_i^*(x_i)$ and $t_i^*(x_i')$ are both optimal when facing either signal value, x_i or x_i'. Our assumption that the agent chooses the higher action when indifferent then implies that $t_i^*(x_i)$ should be chosen in both cases, contradicting that $t_i^*(x_i') < t_i^*(x_i)$. Finally, if $p(x_i)$ and $p(x_i')$ are not both equal to 0 or 1, part (i) implies that $p(x_i) < p(x_i')$. Since B_i is weakly increasing and $B_i(t_i^*(x_i)) - B_i(t_i^*(x_i')) \leq 0$, we have $B_i(t_i^*(x_i)) = B_i(t_i^*(x_i'))$. But then $t_i^*(x_i')$ would be strictly better for the agent than $t_i^*(x_i)$, a contradiction. □

Proof of Proposition 1: We will show that, for any optimal reward scheme $B = (B_i)_{i \in I}$ (not necessarily with B_i a step function for each i), there is a reward scheme $\tilde{B}$ composed of step functions $\tilde{B}_i = k_i \mathbb{1}_{\{t_i \geq T_d\}}$ which is weakly better for the incumbent than B. We will do this by constructing $\tilde{B}_i$ that induces *equivalent equilibrium behavior by all agents*, while being weakly cheaper for the incumbent.

Let $p(x)$ be the incumbent's winning probability conditional on a signal realization $x_i = x$ and the equilibrium strategies $a_i(x_i)$ that the agents choose in response to the reward scheme B, where we denote $t_i^*(x_i) = x_i + a_i(x_i)$. Since $t_i^*(x_i)$ is weakly increasing in x_i by assumption, there is for each i a unique x_i^* such that $t_i^*(x_i) \geq T_d$ (and hence the incumbent wins district i) iff $x_i \geq x_i^*$. Moreover, p is continuously increasing by Lemma 1.

Suppose now that agent i's reward scheme is changed to a step function $\tilde{B}_i = k_i \mathbb{1}_{\{t_i \geq T_d\}}$, while other agents' schemes and behavior remain unchanged. As shown in the text (equations (2) and

(3)), i's best response $\tilde{t}_i(x_i)$ will be such that there is a threshold $\tilde{x}_i$ such that $\tilde{t}_i(x_i) \geq T_d$ iff $x_i \geq \tilde{x}_i$, where $\tilde{x}_i$ is the unique solution to $p(x)k_i - \alpha_i(T_d - x) = 0$. (Note that a unique x solves this equation because the left-hand side is strictly and continuously increasing in x, as guaranteed by Lemma 1.) Set $k_i = \frac{\alpha_i(T_d - x_i^*)}{p(x_i^*)}$. Then, by construction, $\tilde{x}_i = x_i^*$: i's best response to $\tilde{B}_i$ is such that the incumbent wins district i exactly in the same set of cases under both reward schemes.

Define $\tilde{B} = (\tilde{B}_i)_{i \in [0,1]}$ with $\tilde{B}_i = k_i \mathbb{1}_{\{t_i \geq T_d\}}$ and $k_i = \frac{\alpha_i(T_d - x_i^*)}{p(x_i^*)}$. Our next observation is that, if the incumbent offers the reward scheme $\tilde{B}_i$ (rather than B_i) to *all* agents i, it is an equilibrium for each agent i to play as per (3), with threshold x_i^*. Indeed, if they do, the incumbent wins exactly the same set of districts as under the strategy profile $(a_i(x_i))_{i \in [0,1]}$, so the function p remains unchanged, and hence the agents indeed find it optimal to set $\tilde{x}_i = x_i^*$. In other words, the game induced by $\tilde{B}$ has an equilibrium that leads to equivalent electoral outcomes to those of the original equilibrium strategy profile $(a_i(x_i))_{i \in [0,1]}$.

Finally we check that the incumbent pays weakly less under $\tilde{B}$ than under B. For any $x_i < x_i^*$, the incumbent pays 0 to i under $\tilde{B}_i$ and $B_i(x_i + a_i(x_i)) \geq 0$ under B_i. For any $x_i \geq x_i^*$, the incumbent pays k_i under $\tilde{B}_i$. She pays at least as much under B_i because, under B_i, $x_i^* + a_i(x_i^*) \geq T_d$, so

$$p(x_i^*)B_i(x_i^* + a_i(x_i^*)) - \alpha_i(T_d - x_i^*) \geq p(x_i^*)B_i(x_i^* + a_i(x_i^*)) - \alpha_i a_i(x_i^*) \geq p(x_i^*)B_i(x_i^*) \geq 0$$

whereas k_i satisfies $p(x_i^*)k_i - \alpha_i(T_d - x_i^*) = 0$ by construction. Thus $B_i(x_i^* + a_i(x_i^*)) \geq k_i$ and $B_i(x_i + a_i(x_i)) \geq k_i$ for any $x_i \geq x_i^*$ by the monotonicity of B_i and t_i^*. $\square$

Proof of Proposition 2: Our analysis in the text before Proposition 2 shows: when the incumbent's reward scheme takes the form $B_i(t_i) = k_i \cdot \mathbb{1}_{\{t_i \geq T_d\}}$, the belief consistency condition (equation (4)) becomes

$$T' = \int_0^1 Pr(x_i \geq x_i^* | \theta^*) di$$

and individual rationality condition (equation (2)) becomes

$$T_d = Pr(\theta \geq \theta^* | x_i = x_i^*) \; k_i/\alpha_i + x_i^*.$$

In the symmetric case where $k_i = k$ and $\alpha_i = \alpha$, these conditions become

$$T' = Pr(x_i \geq x_i^* | \theta = \theta^*)$$
$$T_d = Pr(\theta \geq \theta^* | x_i = x_i^*) \; k/\alpha + x_i^*.$$

These, in turn, imply $x_i^* = x^*$. Thus, these equilibrium conditions become

$$T' = Pr(x_i \geq x^* | \theta = \theta^*)$$
$$T_d = Pr(\theta \geq \theta^* | x_i = x^*) \; k/\alpha + x^*,$$

which can be further rewritten as

$$T' = 1 - F\left(\frac{x^* - \theta^*}{\sigma}\right) \tag{A1}$$
$$T_d = \frac{\int_{\theta^*}^{\infty} f\left(\frac{x^* - \theta}{\sigma}\right) g(\theta) d\theta}{\int_{-\infty}^{\infty} f\left(\frac{x^* - \theta}{\sigma}\right) g(\theta) d\theta} \frac{k}{\alpha} + x^*. \tag{A2}$$

Moreover, for small enough $\sigma > 0$, these equations pin down a unique solution $(x^*(k;\sigma), \theta^*(k;\sigma))$, which converges to the limit values $(x^*(k); \theta^*(k))$ given in the Proposition as $\sigma \to 0$. The argument is standard (Morris and Shin 2003). Let $\Delta = \frac{x^* - \theta^*}{\sigma}$, so that $x^* = \theta^* + \sigma\Delta$ and equation (A1) becomes $T' = 1 - F(\Delta)$. Thus, $\Delta = F^{-1}(1 - T') \in \mathbb{R}$ is a constant. A solution of the system (A1)-(A2) then corresponds to a value of θ^* such that

$$\frac{\int_{\theta^*}^{\infty} f\left(\frac{\theta^* - \theta}{\sigma} + \Delta\right) g(\theta) d\theta}{\int_{-\infty}^{\infty} f\left(\frac{\theta^* - \theta}{\sigma} + \Delta\right) g(\theta) d\theta} \frac{k}{\alpha} + \theta^* + \sigma\Delta = T_d.$$

Letting $z = \frac{\theta - \theta^*}{\sigma}$, the above equation becomes

$$\frac{\int_0^\infty f\left(\Delta - z\right) g(\theta^* + \sigma z) dz}{\int_{-\infty}^\infty f\left(\Delta - z\right) g(\theta^* + \sigma z) dz} \frac{k}{\alpha} + \theta^* + \sigma\Delta = T_d. \tag{A3}$$

For the uniqueness, it is enough to show that the left-hand side is strictly increasing in θ^* for $\sigma > 0$ small enough. This follows from the fact that the left-hand side is a smooth function of θ^* and $\sigma \geq 0$, and at $\sigma = 0$, it takes the value $F(\Delta)\frac{k}{\alpha} + \theta^*$, with derivative $1 > 0$ with respect to θ^*.

From the definition of Δ, equation (A3) has a unique solution at $\sigma = 0$

$$\theta^* = T_d - F(\Delta)\ k/\alpha = T_d - (1 - T')k/\alpha.$$

Thus, at $\sigma = 0$, $x^* = \theta^* = T_d - (1 - T')k/\alpha$. That $\theta^*(k;\sigma)$, $x^*(k;\sigma)$ converge to this limit as $\sigma \to 0$ follows because the left-hand side of (A3) is continuous in σ. □

Proof of Proposition 3: First, consider the incumbent's problem directly in the limit case $\sigma = 0$. Letting $\hat{k}$ be the interior optimal reward, $\hat{k}$ must satisfy the FOC: $\frac{g(\theta^*(\hat{k}))}{1 - G(\theta^*(\hat{k}))}(1 - T') = \frac{1}{b - \hat{k}}$. If g is log-concave, then so are G and $1 - G$ (An 1998, Lemma 3). In particular, since $1 - G(x)$ is log-concave in x, $(\log(1 - G(x)))' = \frac{-g(x)}{1 - G(x)}$ is decreasing in x, so $\frac{g(x)}{1 - G(x)}$ is increasing in x. Then the left hand side of the FOC is decreasing in $\hat{k}$, since $\theta^*(k)$ is decreasing in k (Proposition 2). The right hand side of the FOC is increasing in $\hat{k} \in [0, b)$, approaching ∞ as $\hat{k} \to b^-$. An interior solution requires b to be sufficiently large so that $\frac{g(T_d)}{1 - G(T_d)}(1 - T') > \frac{1}{b}$.

For $\sigma > 0$, the incumbent's objective is

$$b(1 - G(\theta^*(k;\sigma))) - k \int_{\theta^*(k;\sigma)}^\infty \left(1 - F\left(\frac{x^*(k;\theta) - \theta}{\sigma}\right)\right) g(\theta) d\theta$$

which by the change of variables $\theta = \theta^* + \sigma\nu$, and writing $x^* = \theta^* + \sigma\Delta$, equals

$$b(1 - G(\theta^*(k;\sigma))) - k \int_{\theta^*(k;\sigma)}^\infty \left(1 - F\left(\Delta - \nu\right)\right) g(\theta^*(k;\sigma) + \sigma\nu)\sigma d\nu. \tag{A4}$$

Our proof of Proposition 2 implies that $\theta^*(k;\sigma)$ is a continuous function of (k,σ); the same is then true of (A4). Then an optimum $k^*(\sigma)$ exists for each σ, and any sequence of optimal $k^*(\sigma)$ as $\sigma \to 0$ must accumulate at the unique optimum k^*. □

Proof of Corollary 1: From Proposition 2, with election fraud, the incumbent wins with probability $1 - G(T_d - (1-T')k^*/\alpha)$; without election fraud, she wins with probability $1 - G(T_d)$. The measure of fraudulent votes can be obtained from equation (3):

$$\begin{aligned}\lim_{\sigma\to 0}\int_{-\infty}^{\infty} a^*(x_i;\sigma) pdf(x_i|\theta,\sigma)dx_i &= \lim_{\sigma\to 0}\int_{x^*(\sigma)}^{T_d}(T_d - x_i)pdf(x_i|\theta,\sigma)dx_i \\ &= \begin{cases} 0 & \text{if } \theta < \lim_{\sigma\to 0}\theta^*(\sigma) \\ T_d - \theta & \text{if } \lim_{\sigma\to 0}\theta^*(\sigma) < \theta < T_d \\ 0 & \text{if } T_d < \theta \end{cases}\end{aligned}$$

This, combined with Proposition 2, implies that the ex-ante measure of fraudulent votes, evaluated at $k = k^*$, is

$$\int_{T_d-(1-T')k^*/\alpha}^{T_d}(T_d-\theta)dG(\theta) = (G(T_d)-G(T_d-(1-T')k^*/\alpha))\,(T_d-E[\theta|T_d-(1-T')k^*/\alpha \le \theta \le T_d])$$

□

Proof of Proposition 4: Absent fraud, the incumbent's total vote share conditional on θ is

$$\lim_{\sigma\to 0}\int_{x_i} L(x_i)pdf(x_i|\theta)dx_i = \lim_{\sigma\to 0}\int_{\epsilon_i} L(\theta+\sigma\epsilon_i)dF(\epsilon_i) = L(\theta)$$

The result follows because $L(x)$ is strictly increasing and, from Assumption 1, $L(0) = 1/2$. □

Proof of Proposition 5: Set $\widehat{T}'(b) = 1 - \frac{1}{b}\frac{1-G(T_d)}{g(T_d)}$. For any $T' \ge \widehat{T}'(b)$, we have $\lim_{\sigma\to 0} k^*(\sigma) = 0$ by Proposition 3. Hence electoral fraud is eliminated in the limit as $\sigma \to 0$. Moreover, Proposition 2 yields that $\lim_{\sigma\to 0}\theta^*(k^*(\sigma);\sigma) = T_d - (1-T')\lim_{\sigma\to 0}\frac{k^*(\sigma)}{\alpha} = T_d$. Since $T = 0.5$, we have $T_d = 0$, so $\theta^*(k^*(\sigma);\sigma) \to 0$ as $\sigma \to 0$, and the majoritarian outcome is preserved.

Finally we note that *any* threshold $\widehat{T}'(b)$ satisfying the given conditions must approach 1 as $b \to \infty$. Suppose otherwise, so there is a sequence $b_n \to \infty$ such that $\widehat{T}'(b_n) \leq 1 - \eta$ for all n and a fixed $\eta > 0$. Then, for n large enough, we obtain $b_n > \underline{b} = \frac{1}{\eta}\frac{1-G(T_d)}{g(T_d)}$ leading to $k^* > 0$ by Proposition 3, and positive fraud by Corollary 1, a contradiction. $\square$

Proof of Proposition A1: Suppose $w > 0$. Because $F \in (0, 1/2)$, there are both upper and lower dominance regions. If $\theta > 1/2$, under complete information, agents have a strictly dominant strategy to engage in fraud. If $\theta < 1/2 - F$, under complete information, agents have a strictly dominant strategy not to engage in fraud.

Assumption A1 ensures that $x^* \in (\sigma, 1-\sigma)$. If $x^* \leq \sigma$, then the upper bound of θ conditional on $x_i = x^*$ is $x^* + \sigma \leq 2\sigma < 1/2 - F$. Thus, for $x_i > x^*$ sufficiently close to x^*, agent i's unique best response is not to engage in fraud, contradicting what the strategy prescribes. Similarly, if $x^* \geq 1-\sigma$, then the lower bound of θ conditional on $x_i = x^*$ is $x^* - \sigma \geq 1 - 2\sigma > 1/2 + F > 1/2$. Thus, for $x_i < x^*$ close to x^*, agent i would strictly prefer to engage in fraud, contradicting what the strategy prescribes.

Given a state θ and a strategy cutoff x^*, the measure of agents who engage in fraud is

$$\phi(\theta) = Pr(x_i \geq x^* | \theta). \tag{A5}$$

This is increasing in θ. Moreover, $\frac{1/2-\theta}{F}$ is strictly decreasing in θ. Thus, there exists a unique $\theta^* \in (0, 1/2)$ such that

$$\phi(\theta^*) = Pr(x_i \geq x^* | \theta = \theta^*) = \frac{1/2 - \theta^*}{F} \quad \text{(belief consistency)} \tag{A6}$$

An agent i's net payoff from engaging in fraud versus not is $Pr(\theta \geq \theta^* | x_i) wF - Pr(\theta < \theta^* | x_i) cF$. This is continuous and increasing in x_i. Moreover, the dominance regions ensure that it is negative when x_i is sufficiently low and positive when x_i is sufficiently high. Thus, there exists a threshold signal x^* that makes the agent with signal $x_i = x^*$ indifferent:

$$Pr(\theta \geq \theta^* | x_i = x^*) = \frac{c}{c+w} \quad \text{(individual rationality)} \tag{A7}$$

Thus, any pair (x^*, θ^*) that satisfies belief consistency (A6) and individual rationality (A7) constitutes an equilibrium.

The following statistical property simplifies the analysis.

Lemma 2 *Suppose $x = \theta + \sigma\epsilon$, with $\theta \sim U[\underline{\theta}, \overline{\theta}]$ and $\epsilon \sim U[-1,1]$, where θ and ϵ are independent, $\overline{\theta} > \underline{\theta}$ and $\sigma > 0$. Fix a pair of thresholds $(\hat{\theta}, \hat{x}) \in [\underline{\theta}, \overline{\theta}] \times (\underline{\theta} + \sigma, \overline{\theta} - \sigma)$. Then,*

$$Pr(x_i \leq \hat{x} | \theta = \hat{\theta}) = Pr(\theta \geq \hat{\theta} | x_i = \hat{x})$$

Combining Lemma 2 with (A6) and (A7) yields

$$\phi(\theta^*) = \frac{1/2 - \theta^*}{F} = \frac{w}{w+c} \quad \Leftrightarrow \quad \theta^* = \frac{1}{2} - \frac{w}{w+c}F \tag{A8}$$

From (A8) and (A6),

$$Pr(x_i \geq x^* | \theta = \theta^*) = \frac{1/2 - \theta^*}{F} = \frac{w}{w+c}$$

Thus, $\frac{\theta^* + \sigma - x^*}{2\sigma} = \frac{w}{w+c}$, i.e.,

$$x^* = \theta^* + \sigma \frac{c-w}{c+w}.$$

The proof also shows that (x^*, θ^*) is unique.

If $w = 0$, agents have a dominant strategy not to engage in fraud unless they believe that the incumbent surely wins, in which case they are indifferent. In any equilibrium, $\theta^* = 1/2$. □

Proof of Proposition A2: Given our distributional assumptions, and the fact that $x^* \in$

$(\sigma, 1-\sigma)$, the incumbent's problem simplifies to

$$\max_{w\geq 0}\,(1-\theta^*(w))b - w\left(\frac{1}{2} + F\int_{x^*(w)-\sigma}^{x^*(w)+\sigma}\frac{\theta+\sigma-x^*(w)}{2\sigma}d\theta + F\int_{x^*(w)+\sigma}^{1}d\theta\right)$$

That is,

$$\max_{w\geq 0}\,(1-\theta^*(w))b - \frac{w}{2} - wF(1-x^*(w)) \tag{A9}$$

Substituting from Proposition A1 into (A9), the objective function becomes

$$\left(\frac{1}{2}+\frac{w}{w+c}F\right)b - \frac{w}{2} - wF\left(\frac{1}{2}+\frac{w}{w+c}F - \sigma\frac{c-w}{c+w}\right).$$

If the optimal w is interior, it must satisfy the first-order condition

$$\begin{aligned}
&0 = bF\frac{c}{(c+w)^2} - \frac{1}{2} - F\left(\frac{1}{2}+\frac{w}{w+c}F-\sigma\frac{c-w}{c+w}\right) - wF\left(\frac{c}{(c+w)^2}F + \sigma\frac{2c}{(c+w)^2}\right)\\
&\Leftrightarrow 0 = bFc - \frac{1+F}{2}(c+w)^2 - wF^2(c+w) + F\sigma(c-w)(c+w) - wF(cF+2\sigma c)\\
&\Leftrightarrow 0 = w^2\left(-\frac{1+F}{2} - F^2 - F\sigma\right) + w\left(-(1+F)c - F^2c - F(cF+2\sigma c)\right) + bFc - \frac{1+F}{2}c^2 + F\sigma c^2\\
&\Leftrightarrow 0 = w^2 + 2cw - \frac{bFc - \frac{1+F}{2}c^2 + F\sigma c^2}{\frac{1+F}{2}+F^2+F\sigma}
\end{aligned}$$

which yields

$$\hat{w} = -c \pm \sqrt{c^2 + \frac{bFc - \frac{1+F}{2}c^2 + F\sigma c^2}{\frac{1+F}{2}+F^2+F\sigma}} = -c \pm \sqrt{\frac{bFc + F^2c^2 + 2F\sigma c^2}{\frac{1+F}{2}+F^2+F\sigma}} \tag{A10}$$

The lower solution is always negative and hence is not feasible. Moreover, the first derivative of the objective function converges to $-\frac{1+F}{2} - F^2 - F\sigma < 0$ as $w \to +\infty$, and the second derivative is $-\frac{2cF(b+c(F+2\sigma))}{(c+w)^3} < 0$ for all $w \geq 0$. Hence the objective function is maximized at a finite w, and any feasible interior solution to the FOC is a global maximum. It follows that the optimum is $\hat{w}_+$ when it is positive, and 0 otherwise. □

Proof of Proposition A3: Differentiating w^* in Proposition A1 with respect to σ yields

$$\frac{\partial w^*}{\partial \sigma} = \frac{\sqrt{2}c^2F^2(-bF + c(1 + F + F^2))(b + c(F + 2\sigma))}{(cF(1 + F + 2F^2 + 2F\sigma)(b + c(F + 2\sigma)))^{3/2}}$$

Thus,

$$\frac{\partial w^*}{\partial \sigma} > 0 \quad \Leftrightarrow \quad b > \hat{b} = \frac{c}{F}(1 + F + F^2) > \frac{c}{2F}(1 + F) = b^*$$

Moreover, from Proposition A1, the probability that fraud changes the outcome of the election is $1/2 - \theta^* = \frac{w}{w+c}F$, which is increasing in w. □

References

An, Mark Yuying. 1998. "Logconcavity versus Logconvexity: A Complete Characterization." *Journal of Economic Theory* 80: 350-69.