Jumping Ship or Jumping on the Bandwagon: When Do Local Politicians Support National Candidates?

Online Appendix

A Proofs and derivations

Proof of the Remark. G monitors one mayor at random, and upon monitoring she receives one of two signals: either the mayor is exerting effort or not. Because we restrict the analysis to "symmetric equilibria", we only consider situations in which all bad-type mayors employ the same decision rule. Under these assumptions, we first derive the posterior probability that a mayor is of good type, conditional on *not* observing effort:

$$\begin{aligned}\Pr(\text{good}|e=0) &= \frac{\Pr(e=0|\text{good})\Pr(\text{good})}{\Pr(e=0|\text{good})\Pr(\text{good})+\Pr(e=0|\text{bad})\Pr(\text{bad})}\\ &= \frac{0\cdot\tau}{0\cdot\tau+(1-\tau)(1-\sigma)}=0,\end{aligned} \tag{2}$$

where σ is the probability that a bad type will exert effort. When G fails to observe effort from a monitored mayor, she can infer with certainty that the mayor is a bad type; if G then decides to promote some other mayor, the probability of selecting a good type is $\tau > 0$.

On the other hand, when G observes effort from a monitored mayor, she updates her beliefs about the mayor's type in the following manner:

$$\begin{aligned}\Pr(\text{good}|e=1) &= \frac{\Pr(e=1|\text{good})\Pr(\text{good})}{\Pr(e=1|\text{good})\Pr(\text{good})+\Pr(e=1|\text{bad})\Pr(\text{bad})}\\ &= \frac{\tau}{\tau+(1-\tau)\sigma}\geq\tau,\end{aligned} \tag{3}$$

with the inequality holding strictly as long as $\sigma < 1$. Thus, upon observing that a mayor is exerting effort, the governor cannot do better than promoting him. ■

Proof of Theorem 1. The argument proceeds in two parts. We begin by showing that the mayor's optimal strategy is to campaign in favor of the copartisan candidate as long as $r \geq r^*$. We then show that, conditional on the mayor choosing e=1, the governor will also campaign on behalf of the copartisan candidate if $k \leq k^*$. The first part follows from comparing expected utilities $\mathsf{E}(U|e=1, c=m=0)$ and $\mathsf{E}(U|e=0, c=m=0)$ (we use U for the mayor's utility, V for the governor's utility). The only complication arises from the fact that these utilities comprise probabilities for different events, namely, the number of incumbent party mayors that exert effort because they are of good type. In a scenario where we need to guess the types of the other two candidates in the district, the probabilities that 0, 1, or 2 other mayors are of good type are, respectively, $(1-\tau)^2$, $2\tau(1-\tau)$, and τ^2, and the expected reward is $r/3$. With these quantities in hand, we can derive the Proof of Theorem 1 for a scenario with three copartisan mayors:

$$\begin{aligned}
\mathsf{E}(U|e=1, c=m=0) \geq \mathsf{E}(U|e=0, c=m=0) \\
\underbrace{\left[3\frac{1-s}{3}\psi - v\psi + \frac{1}{2}\right]}_{\text{Pr(Win}|e=1)}\frac{r}{3} - f \geq \underbrace{\left[(1-\tau)^2\left(\frac{1}{2} - v\psi\right)\right]}_{\text{Pr(Win|3 bad types)}}\frac{r}{3} + \\
+\underbrace{\left[2\tau(1-\tau)\left(\frac{1}{2} + \frac{1-s}{3}\psi - v\psi\right)\right]}_{\text{Pr(Win|2 bad types)}}\frac{r}{3} + \\
+\underbrace{\left[\tau^2\left(\frac{1}{2} + 2\frac{1-s}{3}\psi - v\psi\right)\right]}_{\text{Pr(Win|1 bad type)}}\frac{r}{3} \\
r_3^* \geq \frac{9f}{\psi(1-s)(3-2\tau)}.
\end{aligned} \tag{4}$$

The calculus for scenarios with one and two copartisan mayors is similar. Note that when there are two mayors, a bad-type mayor knows the probability that the other mayor is also a bad-type to be $(1-\tau)$, and he also needs to factor in the probability that opposition mayors are of good type, which explains the additional terms that depend on τ in the probabilities of victory. The calculus of expected utilities for the case with two copartisan mayors follows:

$$\begin{aligned}
\mathsf{E}(U|e=1, c=m=0) \geq \mathsf{E}(U|e=0, c=m=0) \\
\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right]\frac{r}{2} - f \geq \tau\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right]\frac{r}{2} \\
+(1-\tau)\left[\psi\left(-\tau\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right]\frac{r}{2} \\
r_2^* \geq \frac{6f}{\psi(1-s)(2-\tau)}.
\end{aligned} \tag{5}$$

A bad-type mayor in a 1-copartisan-mayor scenario campaigns if $\mathsf{E}(U|e=1,c=m=0) \geq \mathsf{E}(U|e=0,c=m=0)$, i.e., if

$$\left(\psi\left(\frac{1-s}{3}\right) - \psi\left[v+2\tau^2\left(\frac{1-s}{3}\right)+2\tau(1-\tau)\left(\frac{1-s}{3}\right)\right]+\frac{1}{2}\right)r - f \geq \left(\frac{1}{2}-\psi\left[v+2\tau^2\left(\frac{1-s}{3}\right)+2\tau(1-\tau)\left(\frac{1-s}{3}\right)\right]\right)r, \quad (6)$$

which obtains when $r_1^* \geq \dfrac{3f}{\psi(1-s)}$. Note that $r_1^* < r_2^* < r_3^*$.

The second part of the proof is as follows: First, if $r > r^*$, the reward is large enough that bad mayors always have an incentive to exert effort. In consequence, G does not have an incentive to monitor, which would at best reveal that the mayors are exerting effort. However, G has an incentive to campaign if the cost is low enough. This follows from comparison of expected utilities $\mathsf{E}(V|e=1,c=1)$ and $\mathsf{E}(V|e=1,c=0)$. In the scenario with three copartisan mayors, the comparison is as follows:

$$\begin{gathered}\mathsf{E}(V|e=1,c=1) \geq \mathsf{E}(V|e=1,c=0)\\ \underbrace{\left[s\psi+3\left(\frac{1-s}{3}\right)\psi - v\psi+\frac{1}{2}\right]}_{\Pr(\text{Win}|e=1,c=1)}\tau - k \geq \underbrace{\left[3\left(\frac{1-s}{3}\right)\psi - v\psi+\frac{1}{2}\right]}_{\Pr(\text{Win}|e=1,c=0)}\tau \\ k_3^* \leq \psi s\tau.\end{gathered} \quad (7)$$

Note that τ enters these expected utilities as the probability that the mayor that the governor taps for promotion is actually of good type. The calculus for the other two scenarios is similar. When there are two copartisan mayors, G campaigns if

$$\begin{gathered}\left[s\psi+(2-\tau)\left(\frac{1-s}{3}\right)\psi - v\psi+\frac{1}{2}\right]\tau - k \geq \left[(2-\tau)\left(\frac{1-s}{3}\right)\psi - v\psi+\frac{1}{2}\right]\tau \\ k_2^* \leq \psi s\tau,\end{gathered} \quad (8)$$

and, with a single copartisan mayor, G campaigns if

$$\begin{gathered}\left[s\psi+(1-2\tau)\left(\frac{1-s}{3}\right)\psi - v\psi+\frac{1}{2}\right]\tau - k \geq \left[(1-2\tau)\left(\frac{1-s}{3}\right)\psi - v\psi+\frac{1}{2}\right]\tau \\ k_1^* \leq \psi s\tau.\end{gathered} \quad (9)$$

Note that $k_1^* = k_2^* = k_3^* = k^*$, which means that effort-inducing cost k^* does not depend on the number of copartisan mayors in the district.

We finally show that bad-type mayors have no incentive to deviate from campaigning when the governor also campaigns. The mayor's calculus in a scenario with three copartisan mayors is as follows:

$$\begin{aligned}
&\mathsf{E}(U|e=1,c=1) \geq \mathsf{E}(U|e=0,c=1)\\
&\left[s\psi+3\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right]\frac{r}{3}-f \geq \left[(1-\tau)^2\left(\frac{1}{2}+s\psi-v\psi\right)\right]\frac{r}{3}+\\
&\qquad+\left[2\tau(1-\tau)\left(\frac{1}{2}+s\psi+\left(\frac{1-s}{3}\right)-v\psi\right)\right]\frac{r}{3}+\\
&\qquad+\left[\tau^2\left(\frac{1}{2}+s\psi-2\left(\frac{1-s}{3}\right)\psi-v\psi\right)\right]\frac{r}{3}\\
&r_3^{\dagger} \geq \frac{9f}{\psi(1-s)(3-2\tau)} \equiv r_3^{*}.
\end{aligned} \tag{10}$$

Inspection of the relevant utilities in the other two scenarios confirms that $r_2^{\dagger}=r_2^{*}$ and $r_1^{\dagger}=r_1^{*}$. With two copartisan mayors:

$$\begin{aligned}
&\left[\psi\left(s+\left(2-\tau\right)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{r}{2}-f \geq \tau\left[\psi\left(s+\left(1-\tau\right)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{r}{2}\\
&\qquad+(1-\tau)\left[\psi\left(s-\tau\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{r}{2}\\
&r_2^{\dagger} \geq \frac{6f}{\psi(1-s)(2-\tau)} \equiv r_2^{*};
\end{aligned} \tag{11}$$

and with one copartisan mayor:

$$\begin{aligned}
&\left[\psi\left(s+\left(1-2\tau\right)\left(\frac{1-s}{3}\right)-\nu\right)+\frac{1}{2}\right]r-f \geq \left[\psi\left(s-2\tau\left(\frac{1-s}{3}\right)-\nu\right)+\frac{1}{2}\right]r\\
&r_1^{\dagger} \geq \frac{3f}{\psi(1-s)} \equiv r_1^{*}.
\end{aligned} \tag{12}$$

We confirm that threshold r^* alone determines whether bad-type mayors will have an incentive to mobilize in pure-strategy equilibria. ■

Proof of Theorem 2. We first prove that the governor will have no incentives to campaign (i.e., she will choose $c=0$) whenever $k \geq k^*$. We will then show that G might still have an incentive to monitor under these circumstances.

Consider first the scenario with three copartisan mayors:[30]

$$\mathsf{E}(V|e=0,c=0) \geq \mathsf{E}(V|e=0,c=1)$$
$$3\tau(1-\tau)^2\left[\left(\frac{1-s}{3}\right)\psi - v\psi + \frac{1}{2}\right]\frac{1}{3} + 3\tau^2(1-\tau)\left[2\left(\frac{1-s}{3}\right)\psi - v\psi + \frac{1}{2}\right]\frac{2}{3} +$$
$$\tau^3\left[3\left(\frac{1-s}{3}\right)\psi - v\psi + \frac{1}{2}\right] \quad \geq$$
$$3\tau(1-\tau)^2\left[\psi\left(s + \frac{1-s}{3} - v\right) + \frac{1}{2}\right]\frac{1}{3} + 3\tau^2(1-\tau)\left[\psi\left(s + 2\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right]\frac{2}{3} +$$
$$\tau^3\left[\left(s + 3\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right] - k$$
$$k \geq \psi s \tau \equiv k^*. \tag{13}$$

Similarly, a comparison of utilities in the 2- and 1-copartisan-mayor scenarios reveals that G will not campaign if $k \geq k^*$. For the 2-copartisan-mayor scenario:

$$\mathsf{E}(V|e=0,c=0) \geq \mathsf{E}(V|e=0,c=1)$$
$$\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right] + 2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right]\frac{1}{2} \geq$$
$$\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right) + s - v\right) + \frac{1}{2}\right] + 2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right) + s - v\right) + \frac{1}{2}\right]\frac{1}{2} - k$$
$$k \geq \psi s \tau \equiv k^*; \tag{14}$$

while in the 1-copartisan mayor scenario:

$$\mathsf{E}(V|e=0,c=0) \geq \mathsf{E}(V|e=0,c=1)$$
$$\psi\left(\frac{1}{2} - v - 2\tau\left(\frac{1-s}{3}\right)\right)\tau \geq \psi\left(\frac{1}{2} - v - 2\tau\left(\frac{1-s}{3}\right) + s\right)\tau - k$$
$$k \geq \psi s \tau \equiv k^*. \tag{15}$$

We have shown that G never has an incentive to campaign if $k > k^*$. We now consider whether there exists an incentive to monitor under different values of k. We consider two

[30]Here, $3\tau(1-\tau)^2$ captures the probability that exactly 1 copartisan mayor will be of good type, therefore exerting effort. The probabilities of exactly 2 and exactly 3 good types are similarly defined as $3\tau^2(1-\tau)$ and τ^3, respectively.

cases, depending on whether k is larger or smaller than k^*.

Case 1: $k > k^*$. We will define v_3^* as the value of v that makes G (weakly) prefer monitoring in a 3-copartisan mayor scenario:[31]

$$\begin{aligned}
&\mathsf{E}(V|m=1,e=0) \geq \mathsf{E}(V|c=m=0,e=0)\\
&\quad (1-\tau)^3\left(-v\psi+\frac{1}{2}\right)\cdot 0+3\tau(1-\tau)^2\left(\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right)\frac{2}{3}+\\
&\qquad +3\tau^2(1-\tau)\left(2\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right)\cdot 1+\tau^3\left(3\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right)\cdot 1-g \geq\\
&\quad (1-\tau)^3\left(-v\psi+\frac{1}{2}\right)\cdot 0+3\tau(1-\tau)^2\left(\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right)\frac{1}{3}+\\
&\qquad 3\tau^2(1-\tau)\left(2\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right)\frac{2}{3}+\tau^3\left(3\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right)\cdot 1\\
&\qquad v_3^* \leq \frac{1}{2\psi}-\frac{g}{\tau(1-\tau)\psi}+\left(\frac{1-s}{3}\right)(1+\tau).
\end{aligned} \tag{16}$$

The value of v_2^* corresponding to a scenario with two mayors follows:

$$\begin{aligned}
&\mathsf{E}(V|m=1,e=0) \geq \mathsf{E}(V|c=m=0,e=0)\\
&\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\cdot 1+2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\cdot 1-g \geq\\
&\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\cdot 1+2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{2}\\
&\qquad v_2^* \leq \frac{1}{2\psi}-\frac{g}{\tau(1-\tau)\psi}+\left(\frac{1-s}{3}\right)(1-\tau).
\end{aligned} \tag{17}$$

In a 1-copartisan-mayor scenario, G never has an incentive to monitor. The comparison of relevant utilities reveals that the governor would monitor only if this action were costless — there is no value of v that provides G with an incentive to monitor a potential bad-type for

[31] The structure of these expected utilities looks daunting, but each term contains the product of three elements: (a) the probability that there are exactly n good-type copartisan mayors; (b) the probability of winning given that there are exactly n good-type copartisan mayors; and (c) the posterior probability of promoting a good-type mayor conditional on G choosing to monitor or not.

whom there are no substitutes because there are no other copartisan mayors to promote:

$$\begin{gathered}
\mathsf{E}(V|m=1,e=0) \geq \mathsf{E}(V|c=m=0,e=0) \\
\tau\left[\left(\frac{1-s}{3}\right)\psi - v\psi - 2\tau\left(\frac{1-s}{3}\right)+\frac{1}{2}\right]-g \geq \tau\left[\left(\frac{1-s}{3}\right)\psi - v\psi - 2\tau\left(\frac{1-s}{3}\right)+\frac{1}{2}\right] \\
0 \geq g.
\end{gathered} \tag{18}$$

The last statement contradicts the assumption that $g>0$. Values of $v \leq v^*$ lead to equilibria in which G might be better off choosing $m=1$ rather than choosing $c=m=0$. The *headhunter monitoring* equilibrium further requires a very low reward, which removes any motivation that bad types may have to exert effort in the presence of monitoring. The threshold r' under which mayors will not exert effort follows from comparing utilities, first for the scenario with three mayors:

$$\begin{gathered}
\mathsf{E}(U|e=1,m=1)=\mathsf{E}(U|e=0,m=1) \\
\left(3\left(\frac{1-s}{3}\right)\psi - v\psi + \frac{1}{2}\right)\frac{r}{3} - f \geq 2\tau(1-\tau)\left(\left(\frac{1-s}{3}\right)\psi - v\psi + \frac{1}{2}\right)\frac{r}{6} + (1-\tau)^2\left(\frac{1}{2}-\psi v\right)\frac{r}{3} \\
r'_3 = \frac{3f}{\psi\left(\frac{1-s}{3}\right)[3-\tau+\tau^2]+\tau\left(\frac{1}{2}-v\psi\right)}.
\end{gathered} \tag{19}$$

The relevant comparison of utilities for the scenario with two mayors follows:

$$\begin{gathered}
\mathsf{E}(U|e=1,m=1)=\mathsf{E}(U|e=0,m=1) \\
\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{r}{2} - f \geq (1-\tau)\left[\psi\left(-\tau\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{r}{2} \\
r'_2 = \frac{2f}{\psi\left(\frac{1-s}{3}\right)(2-\tau^2)+\tau\left(\frac{1}{2}-v\psi\right)}.
\end{gathered} \tag{20}$$

Because parameters s and τ range between 0 and 1, and because of the restrictions on v and

ψ (which imply that $1/2-v\psi$ is positive), it is easy to verify that $r_2' \leq r_2^*$ and $r_3' \leq r_3^*$:

$$\begin{aligned}
r_3^* &\geq r_3' \\
\frac{9f}{\psi(1-s)(3-2\tau)} &\geq \frac{3f}{\psi\left(\frac{1-s}{3}\right)[3-\tau+\tau^2]+\tau\left(\frac{1}{2}-v\psi\right)} \\
\frac{3}{\psi(1-s)(3-2\tau)} &\geq \frac{1}{\frac{1}{3}\psi(1-s)[3-\tau+\tau^2]+\tau\left(\frac{1}{2}-v\psi\right)} \\
3\cdot\frac{1}{3}\psi(1-s)[3-\tau+\tau^2]+3\tau\left(\frac{1}{2}-v\psi\right) &\geq \psi(1-s)(3-2\tau) \\
\psi(1-s)(3-\tau+\tau^2)-\psi(1-s)(3-2\tau) &\geq -3\tau\left(\frac{1}{2}-v\psi\right) \\
\psi(1-s)(\tau+\tau^2) &\geq -3\tau\left(\frac{1}{2}-v\psi\right) \\
\psi(1-s)\tau(1+\tau) &\geq -3\tau\left(\frac{1}{2}-v\psi\right) \\
\underbrace{\psi(1-s)(1+\tau)}_{>0} &\geq \underbrace{-3}_{<0}\underbrace{\left(\frac{1}{2}-v\psi\right)}_{>0}.
\end{aligned} \tag{21}$$

The proof that $r_2^* \geq r_2'$ is analogous. Finally, there are values of r under which a bad-type would exert effort in a scenario where he is the single copartisan mayor and G monitors. However, since G never has an incentive to monitor in these circumstances, effort by a bad-type mayor cannot be sustained in equilibrium.

Case 2: $k<k^*$. We established that G will prefer $c=1$ over $c=m=0$, but we have not considered the possibility that she will prefer $m=1$ over $c=1$, which could happen when k is low. With three copartisan mayors, this occurs whenever $v>v_3'$:

$$\mathsf{E}(V|m=1,e=0) \geq \mathsf{E}(V|c=1,e=0)$$

$$\begin{aligned}
&3\tau(1-\tau)^2\left[\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right]\frac{2}{3}+3\tau^2(1-\tau)\left[2\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right]+ \\
&\qquad \tau^3\left[3\left(\frac{1-s}{3}\right)\psi-v\psi+\frac{1}{2}\right]-g \geq \\
&3\tau(1-\tau)^2\left[\psi\left(s+\frac{1-s}{3}-v\right)+\frac{1}{2}\right]\frac{1}{3}+3\tau^2(1-\tau)\left[\psi\left(s+2\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{2}{3}+ \\
&\qquad \tau^3\left[\psi\left(s+3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]-k
\end{aligned}$$

$$v_3' \leq \left(\frac{1-s}{3}\right)(1+\tau) - \frac{s}{1-\tau} + \frac{1}{2\psi} + \frac{k-g}{\tau(1-\tau)\psi}. \tag{22}$$

With two copartisan mayors, G prefers $m=1$ to $c=1$ when:

$$\begin{aligned}
&\mathsf{E}(V|m=1,e=0) \geq \mathsf{E}(V|c=1,e=0)\\
&\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]+2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]-g \geq\\
&\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)+s-v\right)+\frac{1}{2}\right]+2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right)+s-v\right)+\frac{1}{2}\right]\frac{1}{2}-k\\
&v_2' \leq \left(\frac{1-s}{3}\right)(1-\tau) - \frac{s}{1-\tau} + \frac{1}{2\psi} + \frac{k-g}{\tau(1-\tau)\psi}.
\end{aligned} \tag{23}$$

Note that $v_3' > v_2'$. Finally, G has no incentive to choose monitoring over campaigning in scenarios with one mayor. To see this, consider G's utility calculus:

$$\begin{aligned}
&\mathsf{E}(V|m=1,e=0) \geq \mathsf{E}(V|c=1,e=0)\\
&\tau\left(\psi\left(\left(\frac{1-s}{3}\right)-v-2\tau^2\left(\frac{1-s}{3}\right)-2\tau(1-\tau)\left(\frac{1-s}{3}\right)\right)+\frac{1}{2}\right)-g \geq\\
&\tau\left(\psi\left(s+\left(\frac{1-s}{3}\right)-2\tau^2\left(\frac{1-s}{3}\right)-2\tau(1-\tau)\left(\frac{1-s}{3}\right)\right)+\frac{1}{2}\right)-k\\
&k-\psi s\tau \geq g.
\end{aligned} \tag{24}$$

Because we are inspecting a situation where $k<k^*$, we know that $k-\psi s\tau<0$, which means that the statement $k-\psi s\tau \geq g$ cannot be true since g is strictly positive. ■

Proof of Theorem 3. From Theorem 1, we know that exerting effort is not a dominant strategy for mayors when $r<r^*$. We will now assume that $v \geq \max\{v_3', v_3^*\}$, i.e., that the opposition's normal vote in the district is relatively high (which need not mean that $v>0$). From Theorem 2 we know that the governor has no incentive to monitor when opposition support is high, and therefore mayors have no incentive to exert effort. The governor may still have an incentive to campaign, though, and this will happen under the following circumstances:

$$\begin{aligned}
&\mathsf{E}(V|c=1,e=0) \geq \mathsf{E}(V|c=0,e=0)\\
&\tau^3\left[\psi\left(s+3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]+3\tau^2(1-\tau)\left[\psi\left(s+2\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{2}{3}+
\end{aligned}$$

$$3\tau(1-\tau)^2\left[\psi\left(s+\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{3}-k\geq$$
$$\tau^3\left[\psi\left(3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]+3\tau^2(1-\tau)\left[\psi\left(2\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{2}{3}+$$
$$3\tau(1-\tau)^2\left[\psi\left(\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{3}$$
$$k_3''\leq s\psi\tau=k^*. \tag{25}$$

For a scenario with two mayors:

$$\mathsf{E}(V|c=1,e=0)\geq\mathsf{E}(V|c=0,e=0)$$
$$\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)+s-v\right)+\frac{1}{2}\right]+2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right)+s-v\right)+\frac{1}{2}\right]\frac{1}{2}-k\geq$$
$$\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]+2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{2}$$
$$k_2''\leq s\psi\tau=k^*. \tag{26}$$

And in the one-mayor scenario:

$$\mathsf{E}(V|c=1,e=0)\geq\mathsf{E}(V|c=0,e=0)$$
$$\tau\left[\psi\left(s+\left(\frac{1-s}{3}\right)-2\tau^2\left(\frac{1-s}{3}\right)-2\tau(1-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]-k\geq$$
$$\tau\left[\psi\left(\left(\frac{1-s}{3}\right)-2\tau^2\left(\frac{1-s}{3}\right)-2\tau(1-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]$$
$$k_1''\leq s\psi\tau=k^*. \tag{27}$$

The next step is to check that bad-type mayors still lack an incentive to exert effort when the governor campaigns. This follows from comparison of utilities:

$$\begin{aligned}\mathsf{E}(U|e=1,c=1) &\geq \mathsf{E}(U|e=0,c=1)\\ \left[\psi\left(s+3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{r}{3}-f &\geq \tau^2\left[\psi\left(s+2\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{r}{3}+\\ &\quad 2\tau(1-\tau)\left[\psi\left(s+\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{r}{3}\\ &\quad +(1-\tau)^2\left[\psi(s-v)+\frac{1}{2}\right]\frac{r}{3}\\ r &\geq \frac{9f}{\psi(1-s)(3-2\tau)}=r_3^*.\end{aligned}$$

Bad types still lack an incentive to exert effort in a scenario with two copartisan mayors:

$$\begin{aligned}\mathsf{E}(U|e=1,c=1) &\geq \mathsf{E}(U|e=0,c=1)\\ \left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)+s-v\right)+\frac{1}{2}\right]\frac{r}{2}-f &\geq \tau\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right)+s-v\right)+\frac{1}{2}\right]\frac{r}{2}\\ &\quad +(1-\tau)\left[\psi\left(-\tau\left(\frac{1-s}{3}\right)+s-v\right)+\frac{1}{2}\right]\frac{r}{2}\\ r &\geq \frac{6f}{\psi(1-s)(2-\tau)}=r_2^*,\end{aligned} \tag{28}$$

and the same is true in the one-mayor scenario:

$$\begin{aligned}\mathsf{E}(U|e=1,c=1) &\geq \mathsf{E}(U|e=0,c=1)\\ &\left[\psi\left(s+(1-2\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]r-f \geq\\ &\left[\psi\left(s-2\tau\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]r\\ r &\geq \frac{3f}{\psi(1-s)}=r_1^*.\end{aligned} \tag{29}$$

These derivations confirm that the bad-type mayor will mobilize only if $r \geq r^*$, but this contradicts the condition $r < r^*$ under which the *selfless governor* equilibrium exists.

Finally, we assumed that the *selfless governor* equilibrium can obtain when the normal vote for the opposition is sufficiently low, i.e., when $v < \max\{v', v^*\}$. We now show that $k < k^* \Rightarrow v' < v^*$ and $k > k^* \Rightarrow v' > v^*$. We do this by showing that $v_3' = v_3^* \Leftrightarrow k = k_3^*$ (the proof that $v_2' = v_2^* \Leftrightarrow k = k_2^*$ is basically identical):

$$\begin{gathered} v_3^* = v_3' \\ \frac{1}{2\psi} - \frac{g}{\tau(1-\tau)\psi} + \left(\frac{1-s}{3}\right)(1+\tau) = \frac{1}{2\psi} + \frac{k-g}{\tau(1-\tau)\psi} + \left(\frac{1-s}{3}\right)(1+\tau) - \frac{s}{1-\tau} \\ \frac{k}{\tau(1-\tau)\psi} = \frac{s}{1-\tau} \\ k = \psi s \tau. \end{gathered} \tag{30}$$

■

Proof of Theorem 4. The first step is to show that no combination of pure strategies constitutes an equilibrium, and thus the game can *only* have equilibria in mixed strategies. We derive this step for the scenario with three mayors, but the proof for the scenario with two mayors follows along similar lines. We proceed in three parts: (i) Theorem 2 implies that if $v < v^*$ or $v < v'$ and bad mayors play $e = 0$, the governor always prefers monitoring to campaigning. This rules out all equilibria in which the governor always plays $c = 1$, as well as any equilibrium in which $e = 0$ and $m = 0$. (ii) From Theorem 1 we know that an equilibrium in which bad mayors always play $e = 1$ while the governor always plays $m = 0$ is only possible if $r > r_3^*$; therefore, the pair of strategies $(e = 1, m = 0)$ cannot be an equilibrium either. (iii) Assume that bad mayors always play $e = 1$. In this case, the governor will respond with $m = 1$ iff

$$\begin{gathered} E(V|m=1, e=1) > E(V|m=0, e=1) \\ \left[\psi\left(3\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right]\tau - g > \left[\psi\left(3\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right]\tau \\ 0 > g, \end{gathered} \tag{31}$$

which contradicts the assumption that $g > 0$, and thus rules out the pair $(e = 1, m = 1)$ as an equilibrium choice. (iv) Finally, for $(e = 0, m = 1)$ to be an equilibrium, it must be the case that $E(U|m=1, e=0) > E(U|m=1, e=1)$; but from the Proof of Theorem 2, we know that this can only be the case if $r < r_3'$. Thus, no combination of pure strategies can be sustained in equilibrium when $r_3'<r<r_3^*$ *and* $v<v_3^*$ or $v<v_3'$ A similar proof holds for a scenario with two mayors. In the case of a single mayor, Theorem 2 and 3 imply that the governor never has an incentive to monitor, which means that $m=1$ cannot be part of a mixed strategy equilibrium; in the one-copartisan scenario, no equilibrium in mixed strategies is thus possible. This completes the proof that under the set of assumptions established for Theorem 4, there are no equilibria in pure strategies.

The second step requires showing that there is *at least one* equilibrium in mixed strategies under the conditions specified by Theorem 4. McCarty and Meirowitz (2007) show that Bayesian normal form games are special cases of normal form games, and thus they have at least one (Bayesian) Nash equilibrium in mixed strategies (see Proposition 6.1, pp. 168-9). Specifically, the requirement for such an equilibrium to exist is that (i) the set of players, (ii) the set of feasible strategies and (iii) the set of players' types are all finite. Under the parameter restrictions specified by the assumptions of Theorem 4, the game satisfies these conditions, and thus it must have *at least one* equilibrium in mixed strategies.

The third step requires showing that there is *at most one* equilibrium in mixed strategies. To do so, we first show that the probabilities with which G and the bad type mix depend on r, f and g. Let π and $(1-\pi)$ be the probabilities with which G mixes between monitoring ($m=1$, $c=0$) and not monitoring ($m=0$, $c=0$), respectively. Let π^* be the value of π that makes bad types indifferent between exerting and not exerting effort, i.e., $E(U|e=1,m=\pi)=E(U|e=0,m=\pi)$. With three copartisan mayors, the equality takes the following form:

$$\begin{aligned}&\left[\psi\left(3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{r}{3}-f=\\ &\qquad 2\tau(1-\tau)\left(\psi\left(\frac{1-s}{3}-v\right)+\frac{1}{2}\right)\frac{\pi r}{6}+(1-\tau)^2\left(-\psi v+\frac{1}{2}\right)\frac{\pi r}{3}+\\ &+\left[\tau^2\left(\psi\left(2\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right)+2\tau(1-\tau)\left(\psi\left(\frac{1-s}{3}-v\right)+\frac{1}{2}\right)+(1-\tau)^2\left(-\psi v+\frac{1}{2}\right)\right](1-\pi)\frac{r}{3}\end{aligned} \tag{32}$$

After algebraic manipulation, the equation reduces to the following expression:

$$\pi_3^* = \frac{\frac{3f}{\psi r} - (3-2\tau)\frac{1-s}{3}}{\left[\frac{1-s}{3}(1+\tau) + \frac{1}{2\psi} - v\right]\tau} \tag{33}$$

Since the denominator is always positive, π_3^* will be in the unit-range if

$$\left[\left(\frac{1-s}{3}\right)(1+\tau) + \frac{1}{2\psi} - v\right]\tau > \frac{3f}{\psi r} - (3-2\tau)\left(\frac{1-s}{3}\right) > 0. \tag{34}$$

Similarly, in the 2-mayor case the equilibrium value of π is given by

$$\begin{aligned} &\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right]\frac{r}{2} - f = \\ &\quad \tau\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right](1-\pi)\frac{r}{2} + (1-\tau)\left[\psi\left(-\tau\left(\frac{1-s}{3}\right) - v\right) + \frac{1}{2}\right]\frac{r}{2} \\ &\quad\quad \pi\tau\left[(1-\tau)\left(\frac{1-s}{3}\right) + \frac{1}{2\psi} - v\right] = \frac{2f}{r\psi} - (2-\tau)\left(\frac{1-s}{3}\right), \end{aligned} \tag{35}$$

which reduces to:

$$\pi_2^* = \frac{\frac{2f}{r\psi} - (2-\tau)\frac{1-s}{3}}{\left[(1-\tau)\frac{1-s}{3} + \frac{1}{2\psi} - v\right]\tau}. \tag{36}$$

Since the denominator is always positive, π_2^* will be in the unit range as long as

$$\left[\left(\frac{1-s}{3}\right)(1-\tau) + \frac{1}{2\psi} - v\right]\tau > \frac{2f}{r\psi} - (2-\tau)\left(\frac{1-s}{3}\right) > 0. \tag{37}$$

On the other hand, let ρ and $(1-\rho)$ be the probabilities with which the mayor mixes between exerting and not exerting effort ($e=1$ and $e=0$, respectively) in order to make G indifferent between monitoring and not monitoring, i.e., in order to yield $\mathsf{E}(V|m=1, e=\rho) = \mathsf{E}(V|m=0, e=\rho)$. With three copartisan mayors:[32]

[32]We consider only symmetric equilibria (see fn. 12 in the main text), which means that when one bad type campaigns, all bad types campaign. Yet, inside the first curly bracket in Expression 38 we have terms for effort exerted by two bad types (ρ^2), one bad type ($2\rho(1-\rho)$), or none ($(1-\rho)^2$). This is because, in keeping with the analysis of symmetric equilibria, all bad types are playing a mixed strategy in which each of them exerts effort, independently, with probability ρ.

$$3\tau(1-\tau)^2\left\{\left[\psi\left(3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{3}\rho^2+\left[\psi\left(2\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{2}2\rho(1-\rho)\right.$$
$$\left.+\left[\psi\left(\frac{1-s}{3}-v\right)+\frac{1}{2}\right]\frac{2}{3}(1-\rho)^2\right\}$$
$$+3\tau^2(1-\tau)\left\{\left[\psi\left(3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{2}{3}\rho+\left[\psi\left(2\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right](1-\rho)\right\}$$
$$+\tau^3\left\{\psi\left(3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right\}-g=$$
$$3\tau(1-\tau)^2\left\{\left[\psi\left(3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{3}\rho^2+\left[\psi\left(2\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{3}2\rho(1-\rho)\right.$$
$$\left.+\left[\psi\left(\frac{1-s}{3}\right)+\frac{1}{2}\right]\frac{1}{3}(1-\rho)^2\right\}$$
$$+3\tau^2(1-\tau)\left\{\left[\psi\left(3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{2}{3}\rho+\left[\psi\left(2\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{2}{3}(1-\rho)\right\}$$
$$+\tau^3\left\{\psi\left(3\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right\} \quad (38)$$

After rearranging terms, this expression reduces to:

$$\rho^2\underbrace{(-1)(1-\tau)\left(\frac{1-s}{3}\right)}_{a<0}+\rho\underbrace{(-1)\left[\frac{1}{2\psi}-v+2\left(\frac{1-s}{3}\right)\tau\right]}_{b<0}+\underbrace{\frac{1}{2\psi}-v+\left(\frac{1-s}{3}\right)(1+\tau)-\frac{g}{\tau(1-\tau)\psi}}_{c}=0,$$

and thus the equilibrium value of ρ will be given by

$$\rho_3^*=\frac{\overbrace{-b}^{(+)}\pm\overbrace{\sqrt{b^2-4ac}}^{(+)}}{\underbrace{2a}_{(-)}}, \quad (39)$$

which can *at most* take one value within the $(0,1)$ interval. On the other hand, in the 2-mayor case the equilibrium value of ρ is given by

$$\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]+2\tau(1-\tau)\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{2}\rho$$
$$+2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right](1-\rho)-g=$$
$$\tau^2\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]+2\tau(1-\tau)\left[\psi\left((2-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{2}\rho$$
$$+2\tau(1-\tau)\left[\psi\left((1-\tau)\left(\frac{1-s}{3}\right)-v\right)+\frac{1}{2}\right]\frac{1}{2}(1-\rho),$$

$$\tag{40}$$

which reduces to

$$\rho\left[v-\frac{1}{2\psi}-(1-\tau)\left(\frac{1-s}{3}\right)\right]=\frac{g}{\tau(1-\tau)\psi}+v-\frac{1}{2\psi}-(1-\tau)\left(\frac{1-s}{3}\right)$$

$$\rho_2^*=\frac{g}{\tau(1-\tau)\psi\underbrace{\left[v-\frac{1}{2\psi}-(1-\tau)\left(\frac{1-s}{3}\right)\right]}_{<0}}+1, \tag{41}$$

which takes a unique value. ■

B A potential identification strategy

We propose a causal identification strategy based on Folke's (2014) pioneering idea of employing the number of seats that a party barely wins or loses in a high-magnitude proportional representation district as an instrument for the total number of seats that a party captures in a district. Adapting this insight to Mexico's political system, we use the proportion of municipalities that a party *barely wins or loses* in a district as an instrument for the proportion of copartisan mayors that the party controls in that district.[33] The rationale behind this approach is that as the distance between the winner and first loser of a mayoral race goes to zero, the partisan status of the winning mayor is assigned *as if* randomly (this is the same principle exploited by regression discontinuity designs). In any given district, a varying number of municipalities may be closely won or lost, yielding a measure of the number of municipalities whose partisan status could be seen as exogenously determined. While a party's electoral performance in a district cannot be considered exogenous, a given party will win more close municipal races in some districts than in others, introducing variation in the proportion of copartisan mayors it controls in similar districts. That is, our identification strategy is based on comparing districts that differ in the proportion of municipalities controlled by copartisans solely because that party won (or lost) a relatively large number of close municipal races. Instrumenting for the proportion of copartisan mayors in this way satisfies the exclusion restriction as long as we control for the proportion of close races in a district, which cannot be characterized as exogenous.

Consider the following example of two districts, A and B, each with four municipalities numbered 1 through 4. Table A1 shows hypothetical vote shares for mayoral candidates in the eight municipalities. In District A, the PRI has three mayors and the PAN has one, whereas in District B one municipality belongs to the PRI, one to the PAN, and two to the PRD. We want to estimate the effect of the number of copartisan mayors on the vote share for party j's candidate at the district level, that is: $\text{Vote Share}_j = f(\#\text{Mayors}_j)$.

[33] We alternatively define close elections as those decided by less than 5 or 2.5 percent points.

Table A1: Vote shares for mayoral candidates in eight municipalities

	District A				District B		
	PRI	PAN	PRD		PRI	PAN	PRD
A1	37.0	36.5	26.5	B1	50.0	25.0	25.0
A2	33.2	33.4	33.3	B2	49.0	1.7	49.3
A3	60.0	20.0	20.0	B3	22.0	42.0	36.0
A4	50.2	49.7	0.1	B4	30.0	0.0	70.0
# captured municipalities	3	1	0	# captured municipalities	1	1	2

To instrument #Mayors, we consider the number of copartisan mayors — i.e., "captured municipalities" — in the district that won their elections by a margin of less than five percentage points. Parties that barely lost a mayor (by a margin of 5pp or less) get a $-\frac{1}{2}$ and parties that barely won get $\frac{1}{2}$.[34] The instrument at the district level is the sum of barely lost/barely won mayors, *but the summands are weighted by the municipality's share of the district population*. Table A2 shows the values of the instrumental variable for the three parties in the two districts (w_d is the municipality's share of the total district population, so $\Sigma_d w_d = 1$):

Table A2: Construction of instrumental variable in eight municipalities

	District A				District B		
	PRI	PAN	PRD		PRI	PAN	PRD
A1	$\frac{1}{2}w_1$	$-\frac{1}{2}w_1$	0	B1	0	0	0
A2	$-\frac{1}{2}w_2$	$\frac{1}{2}w_2$	$-\frac{1}{2}w_2$	B2	$-\frac{1}{2}w_2$	0	$\frac{1}{2}w_2$
A3	0	0	0	B3	0	0	0
A4	$\frac{1}{2}w_4$	$-\frac{1}{2}w_4$	0	B4	0	0	0
Instrument	$\frac{w_1 - w_2 + w_4}{2}$	$\frac{-w_1 + w_2 - w_4}{2}$	$\frac{-w_2}{2}$	Instrument	$\frac{-w_2}{2}$	0	$\frac{w_2}{2}$

We use the district-level instrument in a 2SLS regression setup. The first-stage is

$$\#\text{Mayors}_{d,t} = \alpha + \lambda \sum c_{m,t-1} \cdot w_{m,d} + \gamma \sum v_{m,t-1} \cdot w_{m,d} + g\Big(\sum \mathbf{V}_{m,d,t-1}\Big) + \delta_t + \varepsilon_{d,t},$$

[34]We follow Folke (2014) in writing (minus) $\frac{1}{2}$ rather than (minus) 1 because we include an additional control variable that takes the value of $\frac{1}{2}$ whenever there was a close race in a district. Thus, we have $\frac{1}{2} + \frac{1}{2} = 1$ if there was a close municipal race that party j won, and $\frac{1}{2} - \frac{1}{2} = 0$ if there was a close race that party j lost.

where d indexes districts, m indexes municipalities (nested within districts), t indexes election years, c takes the value of $\frac{1}{2}$ if the immediately preceding election (at $t-1$) in municipality m was close, v takes the value of (minus) $\frac{1}{2}$ if the party (lost) won the close election in municipality m at $t-1$, $w_{m,d}$ indicates municipality m's share of district d's population, $g(\sum \mathbf{V}_{m,d,t-1})$ is a fourth-order polynomial of the party's *normal vote share*, and δ_t is a year fixed effect. To continue with the example above, assume that $w_d = 0.25\ \forall\, d$ (i.e., all four municipalities within both districts have equal population). Table A3 shows the population-weighted number of mayors elected *as if* randomly (the instrument that we use), the population-weighted number of close elections for the party in the district, and the population-weighted number of actual mayors (the endogenous variable that we instrument):

Table A3: *Instrument*, *close elections*, and *observed mayors* in eight municipalities

	District A			District B		
Party	Instrument	Close election	Observed	Instrument	Close election	Observed
PRI	1/8	3/8	3/4	–1/8	1/8	1/4
PAN	–1/8	3/8	1/4	0	0	1/4
PRD	–1/8	1/8	0	1/8	1/8	2/4

The second stage regression is

$$\text{winner}_{d,t} = \mu + \rho \sum c_{m,t-1} \cdot w_{m,d} + \beta\, \widehat{\#\text{Mayors}}_{d,t} + g\Big(\sum \mathbf{V}_{m,d,t-1}\Big) + \zeta_t + \nu_{d,t},$$

where $\widehat{\#\text{Mayors}}_{d,t}$ is the instrumented treatment and $winner_{d,t}$ is an indicator of whether the copartisan congressional candidate carried the district.

To consider the number of mayors that a party barely won/lost in a district a valid instrument of the number of mayors that a party holds, we need to assess the verisimilitude of a number of assumptions (Sovey and Green 2011). The instrument is *relevant* in that its correlation with *weighted number of mayors* is high (around 0.55–0.60). The instrument is *strong*, as suggested in Table A4 by F-statistics for excluded instruments that amply surpass the F=10 cut-off point commonly used as a rule-of-thumb for weak instruments (Staiger and Stock 1997). The instrument complies with the *monotonicity* assumption, which would be violated if, for example, lower values of mayoralties barely won/lost corresponded to much higher numbers of actual municipalities held by a party because the party manages to subvert electoral results in court or by force (contested mayoral elections exist but are a tiny fraction of all mayoral elections).

The most important assumption, the *exclusion restriction*, is violated by the instrument if the number of close races is not controlled for. Consider an alternative mechanism through which the number of mayors barely won/lost may correlate with the probability of winning the district. First, the number of mayors barely won/lost increases with the number of close races in the district. Second, the number of close races in the district may affect future electoral outcomes directly, for example if parties campaign harder and spend more money in the district in anticipation that elections will be close. This potential violation of the exclusion restriction is in fact noted by Folke (2014). Yet as he argues, this alternative pathway is blocked — and validity of the exclusion restriction restored — through inclusion of the proportion of *close elections* in the district as an additional control. Thus, while including the instrument as a regressor *on its own* would violate the exclusion restriction, the instrument satisfies the exclusion restriction once we control for the proportion of close elections in a district. Note as well that we include controls for district vote share, and in some specifications also for fixed state- and election-effects. These are necessary if the number of close elections is somewhat driven by vote share (parties with larger vote shares are more competitive, and therefore likely to be close to the discontinuity more often), by state characteristics (the voters in a particular state may be evenly divided among parties, therefore generating a larger number of close elections), or by electoral race (federal elections in a particular year may have led to a larger number of even contests).

In keeping with an the analysis of PRI-controlled municipalities to avoid lack of independence across observations with copartisan and opposition governors, Table A4 displays two-stage least squares estimates of the effect of the proportion of copartisan PRI mayors in a district on the probability that the PRI congressional candidate will win the district based on two sets of observations: for candidates with PRI governors and without PRI governors). In both cases, the instrument is the (population-weighted) proportion of municipalities in a district that a party won or lost by a margin of five percentage points or less. In the first-stage specification we regress the (population-weighted) *observed* number of copartisan mayors on the instrument, the (population-weighted) proportion of close mayoral races in the district, a fourth-order polynomial of the party's *normal vote* share in the district, the *margin of victory* in the previous election, the municipality's *poverty* level, and a set of election-year dummies. Models 2 and 4 add state fixed effects. Across all models the instrument is a statistically significant predictor of the (population-weighted) proportion of copartisan mayors in the municipality. Furthermore, F-statistics for the excluded instrument amply surpass the cut-off point suggested by Staiger and Stock (1997), alleviating concerns that the instrument may be weak.

In the second-stage model we estimate the effect of the *instrumented* proportion of copartisan mayors on the probability that the party will win the district, based on a linear probability model.[35] In line with expectations, comparing the intercept estimates in Models 1–2 against those in Models 3–4 confirms that, when there are no copartisan mayors in the district, the probability that a congressional candidate will carry a district in a state with a copartisan governor is larger than the probability that she will win the election if the state is led by a non-copartisan governor. We also find that the (instrumented) proportion of copartisan mayors in the district has a stronger effect on the probability of winning the district when there is an opposition governor. Table A5 shows that the results are essentially identical when we do *not* weight the proportion of copartisan mayors by the municipality's population, and Table A6 shows that the results remain in place — though the estimates become substantially noisier — when using a 2.5 pp. instead of a 5 pp. margin to define close elections.

[35]We did not use a probit/logit specification for the second stage for lack of a "canned" routine to estimate standard errors that account for estimation uncertainty in the first stage.

Table A4: 2SLS estimates: effect of *proportion of copartisan mayors* on a congressional candidate's *probability of victory* in Mexico, 2000-2012 (PRI only)

	Copartisan governor		*Opposition governor*	
	(1)	(2)	(3)	(4)
(a) Second stage regression (outcome: *copartisan victory*)				
Proportion of copartisan mayors	-0.11	-0.03	0.63	0.52
(instrumented)	(0.14)	(0.15)	(0.13)*	(0.12)*
Margin of victory (lagged)	0.32	-0.11	0.08	0.15
	(0.19)	(0.24)	(0.46)	(0.51)
Poverty	0.11	0.19	0.13	0.05
	(0.03)*	(0.04)*	(0.03)*	(0.04)
Proportion of close elections	-0.05	-0.01	0.22	0.27
	(0.18)	(0.17)	(0.17)	(0.19)
Intercept	0.76	0.47	0.08	0.08
	(0.12)*	(0.15)*	(0.06)	(0.08)
RMSE	0.37	0.36	0.34	0.33
(b) First stage regression (outcome: *proportion of copartisan mayors*)				
Proportion barely won/lost	0.96	0.96	1.05	0.97
	(0.04)*	(0.05)*	(0.05)*	(0.13)*
Margin of victory (lagged)	-0.29	-0.46	-0.37	-0.43
	(0.11)*	(0.10)*	(0.30)	(0.16)*
Poverty	-0.04	-0.04	0.01	-0.05
	(0.01)*	(0.03)	(0.02)	(0.08)
Proportion of close elections	-0.47	-0.39	0.36	0.31
	(0.12)*	(0.14)*	(0.15)*	(0.11)*
Intercept	0.77	1.10	0.30	0.32
	(0.03)*	(0.05)*	(0.05)*	(0.05)*
RMSE	0.20	0.18	0.26	0.21
F-statistic (excl. instrument)	548	436	510	59
Previous vote share	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
State effects	No	Yes	No	Yes
num. obs.	336	336	146	146

* p<0.05; † p<0.10; Instrumental variable estimations of models in Table 2b.

2SLS estimates, standard errors clustered by state in parentheses.

IV is the population-weighted *proportion barely won/lost*, based on a 5 percentage point margin.

Table A5: 2SLS estimates: effect of *proportion of copartisan mayors* on a congressional candidate's *probability of victory* in Mexico, 2000-2012 (unweighted averages, PRI only)

	Copartisan governor		*Opposition governor*	
	(1)	(2)	(3)	(4)
(a) Second stage regression (outcome: *copartisan victory*)				
Proportion of copartisan mayors	-0.03	0.04	0.70	0.56
(instrumented)	(0.14)	(0.15)	(0.15)*	(0.18)*
Margin of victory (lagged)	0.33	-0.09	0.09	0.16
	(0.19)†	(0.22)	(0.47)	(0.49)
Poverty	0.12	0.19	0.12	0.05
	(0.03)*	(0.04)*	(0.03)*	(0.05)
Proportion of close elections	-0.04	-0.01	0.15	0.24
	(0.19)	(0.18)	(0.13)	(0.17)
Intercept	0.70	0.39	0.07	0.06
	(0.11)*	(0.15)*	(0.07)	(0.10)
RMSE	0.37	0.36	0.34	0.33
(b) First stage regression (outcome: *proportion of copartisan mayors*)				
Proportion barely won/lost	0.99	0.99	1.08	0.97
	(0.05)*	(0.05)*	(0.06)*	(0.14)*
Margin of victory (lagged)	-0.40	-0.55	-0.41	-0.45
	(0.13)*	(0.11)*	(0.31)	(0.19)*
Poverty	-0.04	-0.05	0.02	-0.04
	(0.01)*	(0.03)	(0.02)	(0.08)
Proportion of close elections	-0.49	-0.41	0.29	0.22
	(0.12)*	(0.14)*	(0.15)†	(0.10)*
Intercept	0.78	1.11	0.30	0.32
	(0.03)*	(0.05)*	(0.05)*	(0.06)*
RMSE	0.21	0.19	0.27	0.22
F-statistic (excl. instrument)	346	369	366	50
Previous vote share	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
State effects	No	Yes	No	Yes
num. obs.	336	336	146	146

* $p<0.05$; † $p<0.10$; Instrumental variable estimations of models in Table 2b.

2SLS estimates, standard errors clustered by state in parentheses.

IV is the (unweighted) *proportion barely won/lost*, based on a 5 percentage point margin.

Table A6: 2SLS estimates: effect of *proportion of copartisan mayors* on a congressional candidate's *probability of victory* in Mexico, 2000-2012 (2.5pp. margin, PRI only)

	Copartisan governor		*Opposition governor*	
	(1)	(2)	(3)	(4)
(a) Second stage regression (outcome: *copartisan victory*)				
Proportion of copartisan mayors	-0.34	-0.26	0.72	0.52
(instrumented)	(0.12)*	(0.13)*	(0.24)*	(0.26)*
Margin of victory (lagged)	0.27	-0.20	0.13	0.15
	(0.22)	(0.27)	(0.48)	(0.52)
Poverty	0.10	0.17	0.13	0.06
	(0.03)*	(0.04)*	(0.03)*	(0.05)
Proportion of close elections	-0.09	0.06	0.27	0.29
	(0.20)	(0.19)	(0.14)*	(0.15)†
Intercept	0.92	0.72	0.06	0.11
	(0.10)*	(0.17)*	(0.09)	(0.13)
RMSE	0.39	0.37	0.35	0.33
(b) First stage regression (outcome: *proportion of copartisan mayors*)				
Proportion barely won/lost	1.05	1.03	1.15	1.06
	(0.05)*	(0.06)*	(0.07)*	(0.17)*
Margin of victory (lagged)	-0.32	-0.43	-0.38	-0.68
	(0.18)†	(0.22)†	(0.32)	(0.22)*
Poverty	-0.05	-0.05	0.01	-0.05
	(0.01)*	(0.03)†	(0.03)	(0.08)
Proportion of close elections	-0.42	-0.38	0.24	0.11
	(0.11)*	(0.13)*	(0.17)	(0.10)
Intercept	0.76	1.16	0.33	0.43
	(0.02)*	(0.05)*	(0.05)*	(0.05)*
RMSE	0.22	0.21	0.29	0.24
F-statistic (excl. instrument)	380	276	268	37
Previous vote share	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
State effects	No	Yes	No	Yes
num. obs.	336	336	146	146

* $p<0.05$; † $p<0.10$; Instrumental variable estimations of models in Table 2b.

2SLS estimates, standard errors clustered by state in parentheses.

IV is the (populatin-weighted) *proportion barely won/lost*, based on a 2.5 percentage point margin.

C Robustness

Table A7: Probit estimates: *proportion of copartisan mayors* on a congressional candidate's *probability of victory* in Mexico, 2000-2012

	Copartisan governor		*Opposition governor*	
	(1)	(2)	(3)	(4)
(a) Full Sample				
Proportion of copartisan mayors	0.81	0.50	0.89	1.03
	(0.31)*	(0.32)	(0.41)*	(0.44)*
Margin of victory (lagged)	0.36	-0.85	1.92	2.16
	(0.70)	(1.08)	(0.94)*	(0.82)*
Poverty	0.33	0.68	-0.08	-0.38
	(0.13)*	(0.22)*	(0.12)	(0.10)*
Intercept	0.12	0.57	-0.78	-0.89
	(0.19)	(0.25)*	(0.18)*	(0.19)*
Null deviance	531.8	531.8	552.1	552.1
Residual deviance	424.3	357.3	424.5	378.7
AIC	448.3	417.3	448.5	438.7
num. obs.	453	453	461	461
(b) PRI-only Sample				
Proportion of copartisan mayors	0.04	0.01	2.82	2.99
	(0.31)	(0.32)	(0.41)*	(0.44)*
Margin of victory (lagged)	1.59	-0.56	2.57	6.08
	(0.70)*	(1.08)	(0.94)*	(0.82)*
Poverty	0.52	0.96	0.69	0.37
	(0.13)*	(0.22)*	(0.12)*	(0.10)*
Intercept	0.57	0.22	-2.62	-3.26
	(0.19)*	(0.25)	(0.18)*	(0.19)*
Null deviance	377.9	377.9	200.2	200.2
Residual deviance	268.9	228.7	86.0	67.5
AIC	292.9	284.7	110.0	111.5
num. obs.	336	336	146	146
Previous vote share	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
State effects	No	Yes	No	Yes

* $p<0.05$; † $p<0.10$; Probit replications of models in Table 2.

ML estimates, standard errors clustered by state in parentheses.

Table A8: OLS estimates: The outcome is the congressional candidate's *vote share*.

	Copartisan governor		*Opposition governor*	
	(1)	(2)	(3)	(4)
(a) Full Sample				
Proportion of copartisan mayors	0.03	0.02	0.01	0.03
	(0.02)	(0.02)	(0.02)	(0.02)
Margin of victory (lagged)	-0.14	-0.13	-0.07	-0.00
	(0.05)*	(0.05)*	(0.06)	(0.06)
Poverty	0.02	0.05	0.01	-0.01
	(0.01)*	(0.01)*	(0.01)	(0.01)
Intercept	0.40	0.41	0.32	0.33
	(0.02)*	(0.02)*	(0.01)*	(0.01)*
RMSE	0.08	0.07	0.08	0.08
num. obs.	453	453	461	461
(b) PRI-only Sample				
Proportion of copartisan mayors	-0.02	-0.01	0.03	-0.00
	(0.02)	(0.02)	(0.02)	(0.01)
Margin of victory (lagged)	0.03	-0.02	-0.18	-0.08
	(0.04)	(0.03)	(0.07)*	(0.06)
Poverty	0.02	0.05	0.05	0.03
	(0.01)*	(0.01)*	(0.01)*	(0.01)*
Intercept	0.44	0.41	0.33	0.37
	(0.02)*	(0.03)*	(0.01)*	(0.01)*
RMSE	0.07	0.06	0.07	0.06
num. obs.	336	336	146	146
Previous vote share	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
State effects	No	Yes	No	Yes

* $p<0.05$. Specifications replicate the models in Table 2, but employing *vote share* as the outcome. OLS estimates (standard errors clustered by state in parentheses).

Table A9: OLS estimates: The outcome is the congressional candidate's *margin of victory.*

	Copartisan governor		*Opposition governor*	
	(1)	(2)	(3)	(4)
(a) Full Sample				
Proportion of copartisan mayors	0.09	0.05	0.08	0.06
	(0.03)*	(0.02)*	(0.04)*	(0.03)†
Margin of victory (lagged)	0.05	-0.04	0.11	0.04
	(0.06)	(0.09)	(0.09)	(0.09)
Poverty	0.03	0.07	-0.01	-0.04
	(0.01)*	(0.02)*	(0.02)	(0.02)*
Intercept	0.01	0.05	-0.10	-0.10
	(0.02)	(0.02)*	(0.02)*	(0.02)*
RMSE	0.14	0.12	0.14	0.13
num. obs.	453	453	461	461
(b) PRI-only Sample				
Proportion of copartisan mayors	0.01	0.00	0.12	0.03
	(0.03)	(0.02)	(0.04)*	(0.04)
Margin of victory (lagged)	0.29	0.13	-0.08	-0.09
	(0.05)*	(0.07)†	(0.12)	(0.15)
Poverty	0.03	0.07	0.06	0.01
	(0.01)*	(0.02)*	(0.01)*	(0.01)
Intercept	0.05	0.08	-0.09	-0.06
	(0.02)*	(0.05)†	(0.02)*	(0.02)*
RMSE	0.12	0.11	0.12	0.11
num. obs.	336	336	146	146
Previous vote share	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
State effects	No	Yes	No	Yes

* $p<0.05$. Specifications replicate the models in Table 2, but employing *margin of victory* as the outcome. OLS estimates (standard errors clustered by state in parentheses).

Table A10: OLS estimates: controlling for the effective number of municipalities

	Copartisan governor		*Opposition governor*	
	(1)	(2)	(3)	(4)
(a) Full Sample				
Proportion of copartisan mayors	0.25	0.17	0.28	0.29
	(0.10)*	(0.09)†	(0.13)*	(0.13)*
Effective number of municipalities	0.01	-0.00	-0.01	-0.00
	(0.00)*	(0.01)	(0.00)*	(0.00)
Margin of victory (lagged)	0.03	-0.25	0.54	0.43
	(0.19)	(0.26)	(0.23)*	(0.20)*
Poverty	0.06	0.17	0.01	-0.09
	(0.03)*	(0.04)*	(0.03)	(0.03)*
Intercept	0.49	0.65	0.29	0.24
	(0.07)*	(0.07)*	(0.06)*	(0.05)*
RMSE	0.40	0.38	0.39	0.38
num. obs.	453	453	461	461
(b) PRI-only Sample				
Proportion of copartisan mayors	0.03	0.05	0.52	0.42
	(0.08)	(0.08)	(0.10)*	(0.13)*
Effective number of municipalities	0.01	0.01	0.01	0.00
	(0.01)	(0.00)	(0.01)	(0.01)
Margin of victory (lagged)	0.29	-0.13	-0.04	0.10
	(0.18)	(0.23)	(0.41)	(0.53)
Poverty	0.10	0.17	0.11	0.04
	(0.03)*	(0.04)*	(0.05)*	(0.07)
Intercept	0.62	0.34	0.12	0.14
	(0.05)*	(0.11)*	(0.07)	(0.08)†
RMSE	0.37	0.36	0.34	0.33
num. obs.	336	336	146	146
Previous vote share	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes
State effects	No	Yes	No	Yes

* $p<0.05$. Specifications replicate the models in Table 2.

OLS estimates (standard errors clustered by state in parentheses).

Table A11: RD estimates of *municipal incumbency* on the municipal-level performance of Mexican congressional candidates, 2000-2012

	LATE	SE	bwd.	N^{-}	N^{+}
Local linear regression (outcome: *copartisan victory*)					
Copartisan governor	-0.09	0.03*	0.21	1355	1649
Non-copartisan governor	-0.11	0.05*	0.11	996	742
Local linear regression (outcome: *turnout*)					
Copartisan governor	-0.01	0.01	0.17	1225	1472
Non-copartisan governor	0.01	0.01	0.15	1322	900

* $p<0.05$. The outcome variables correspond to *federal* elections held at $t+1$, but are measured at the level of the municipality. The running variable is the *margin of victory* in the municipal election held at t. Bias-corrected estimates are based on a local linear regression fitted separately at both sides of the threshold and employing a triangular kernel. The bandwidth is calculated according to the automatic selection procedure proposed by Calonico, Cattaneo and Titiunik (2014).