Unlikely Democrats: Economic Elite Uncertainty Under Dictatorship and Support for Democratization

Supporting Information

Michael Albertus

University of Chicago

Victor Gay

University of Chicago

A Proof of Proposition 1

We prove part (b) of Proposition 1 first.[1] Suppose $\theta < \theta'$, and denote the associated optimal strategies of those types by $m = m^*(\theta)$ and $m' = m^*(\theta')$. Here, we show that $m < m'$. First, rewrite the autocrat's expected utility as:

$$\int a(m+u)g(u)\mathrm{d}u - C(m,\theta) = \pi(m) - C(m-\theta).$$

Let $\theta < \theta'$, and denote $m = m(\theta)$ and $m' = m(\theta')$. By optimality,

$$\pi(m) - C(m-\theta) \geq \pi(m') - C(m'-\theta) \text{ and } \pi(m') - C(m'-\theta') \geq \pi(m) - C(m-\theta').$$

Therefore, we can write

$$C(m'-\theta) - C(m-\theta) \geq C(m'-\theta') - C(m-\theta').$$

By strict convexity of $C(\cdot)$, $m' > m$, so the optimal messaging strategy $m^*(\theta)$ is increasing in θ. Now, we can show that the elite uses a threshold rule:

$$a^*(s) = \begin{cases} 1 \text{ if } s \geq k^* \\ 0 \text{ if } s < k^* \end{cases}$$

where the equilibrium threshold k^* is such that the elites are indifferent between autocracy and democracy whenever they receive a signal exactly at this threshold, i.e. when $\mathbb{E}[\theta|s = k^*] = 1 - \widetilde{T}$. The equilibrium message $m^*(\theta)$ is increasing in the type θ. Moreover, the distribution of the noise satisfies the monotone likelihood property. That is, for two messages m' and m'', if $m' > m''$ then $g(s-m')/g(s-m'')$ increases in s. This implies

[1]Some of the proofs and derivations follow Cunningham & de Barreda (2015) and Caselli et al. (2014).

that higher signals are "good news" (Milgrom 1981): higher signals imply a higher posterior distribution of types in the sense of first-order stochastic dominance, so that the expected message is strictly increasing in the signal. As a result, the expected type conditional on the signal is increasing in the signal: if $s_1 < s_2$, then $\mathbb{E}[\theta|s_1] < \mathbb{E}[\theta|s_2]$. Therefore, if $\mathbb{E}[\theta|s = k^*] = 1 - \widetilde{T}$ has a solution, it is unique. Hence, the elite follows a threshold rule by which $a(s) = 1$ if and only if $s \geq k^*$.[2]

We can now prove the last part of the proposition. From the strategy of the elite, $a(m + u) = 1 \iff s > k^* \iff u > k^* - m$. Therefore, we can write the autocrat's expected utility:

$$\int_{k^*-m}^{\infty} g(u)\mathrm{d}u - C(m - \theta).$$

The first- and second-order conditions are:

$$g\left(k^* - m^*(\theta)\right) = C'\left(m^*(\theta) - \theta\right) \quad \text{and} \quad -g'\left(k^* - m^*(\theta)\right) - C''\left(m^*(\theta) - \theta\right) < 0.$$

Assumption 1. *inf* $C'' >$ *sup* g'.

By assumption 1, the second order condition is always satisfied as $-\sup(-C'') = \inf C''$, and g' is such that $g(-x) = -g(x)$. Moreover, assumption 1 ensures that the slope of the marginal cost C'' is always larger than the slope of the marginal benefit g'. Intuitively, it implies that the cost function is sufficiently convex so that that the marginal cost and marginal benefit cross only once, ensuring that the solution to the first-order condition is unique. This also implies that there is a unique k^* such that this equation is satisfied. In the case of a quadratic cost function, $C''(x) = c$. Moreover, in the case of a normally distributed noise, $-g'(\cdot)$ attains a maximum at σ_u. To see why, note that $g'(x) = -\left[x/\sigma_u^2\right] g(x)$, and $g''(x) = \left[\left(x^2 - \sigma_u^2\right)/\sigma_u^4\right] g(x)$. So $g'(\cdot)$ attains a minimum at $x = \sigma_u$. Thus, assumption 1 reduces to the following restriction on the cost function: $c > \mathrm{e}^{-1/2}/\sigma_u^2\sqrt{2\pi}$.

[2]See Caselli et al. (2014, 414) for a similar argument.

B Derivation of the distribution of messages

The autocrat's messaging strategy uniquely defines the optimal message $m(\theta, k)$. However, it is defined implicitly. Nevertheless, we can invert this function to get an explicit expression for $\theta(m, k)$ as suggested by Cunningham & de Barreda (2015). This function is the type that would optimally send message m given a threshold k:

$$\theta(m,k) = m - r(g(k-m)) = m - \frac{g(k-m)}{c},$$

where $r(x) = (C')^{-1}(x) = x/c$. From there, we can compute the distribution of messages. For $m \in [\mu - b, \mu + b]$,

$$\begin{aligned}
H(m,k) &= \Pr(\overline{m} \leq m) \\
&= \Pr(\overline{\theta} \leq \theta) \ \text{ as } m(\theta) \text{ is increasing in } \theta \\
&= \Pr(\overline{\theta} \leq m - r(g(k-m))) \\
&= F(m - r(g(k-m))) \\
&= \frac{m - r(g(k-m)) - \mu + b}{2b}.
\end{aligned}$$

The density h is then obtained by differentiating H with respect to m. Therefore, given a threshold $k \in [\mu - b, \mu + b]$, the cumulative distribution function of messages H and the probability density function of messages h are defined by:

$$H(m,k) = \begin{cases} \frac{1}{2} - \frac{\mu}{2b} + \frac{m - r(g(k-m))}{2b} & \text{if } m \in [\mu - b, \mu + b] \\ 0 & \text{if } m < \mu - b \\ 1 & \text{if } m > \mu + b \end{cases}$$

$$h(m,k) = \begin{cases} \frac{1 + r'(g(k-m))g'(k-m)}{2b} & \text{if } m \in [\mu - b, \mu + b] \\ 0 & \text{if } m \notin [\mu - b, \mu + b] \end{cases}$$

With the quadratic cost function, the density of messages becomes:

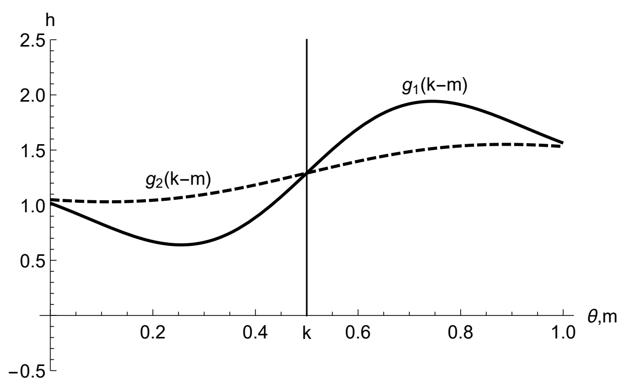
$$h(m,k) = \begin{cases} \dfrac{1 + \frac{g'(k-m)}{c}}{2b} & \text{if } \ m \in [\mu - b, \mu + b] \\ 0 & \text{if } \ m \notin [\mu - b, \mu + b] \end{cases}$$

We display the distribution of the autocrat's messages in Figure B.1a. Clearly, messages are bunching just above the threshold (solid curve). This phenomenon can be explained by Figure B.1b, which plots the autocrat's messaging strategy for ten different types, all equally spread out in the $[0, 1]$ interval. The optimal messages for the various types are represented by the vertical black lines. While the types are equally spaced, the messages are not: the space between the messages is shorter just to the right of the threshold. This is because to the right of the threshold, the marginal cost curves cross the marginal benefit curve on the downward sloping part of the marginal benefit curve. Intuitively, autocrats that are close but to the left of the threshold have strong incentives to exert effort in exerting effort in sending a higher message because the signal received by the elite is more likely to be high enough to cross the threshold.

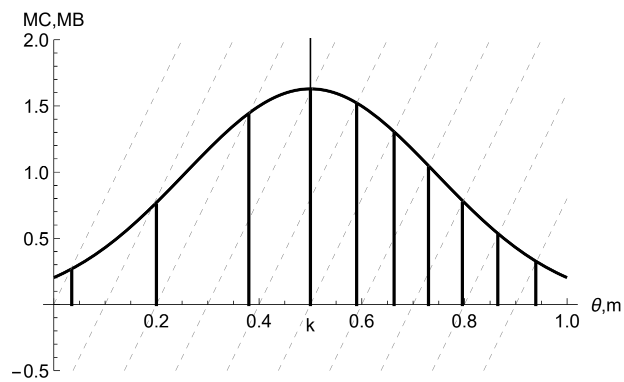
When signals become less precise, there is less bunching of messages above the threshold. This in turn decreases the likelihood that the elites observe signals above the threshold and therefore increases the likelihood of democratization (see the dashed curve on Figure B.1a).

Figure B.1: Messages density and autocrat's messaging strategy

(a)



(b)



Parameters: $\mu = 0.5$, $k = 0.5$, $\sigma^2_{1,u} = 0.06$ (solid curves), $\sigma^2_{2,u} = 0.15$ (dashed curve), $c = 8$.

C Derivation of Equation 2

The probability of democratization given a threshold k is the probability that a signal is below the threshold given the realization of u:

$$\begin{aligned}
\Pr(\text{democracy}; k) &= \Pr(m \leq k - u | k, u) \\
&= H(k-u, k|u) \\
&= \int H(k-u,k)g(u)\mathrm{d}u \\
&= \int \left[\frac{k-u-r(g(u))-\mu+b}{2b}\right] g(u)\mathrm{d}u \\
&= \frac{1}{2} - \frac{\mu}{2b} + \frac{k}{2b} - \frac{1}{2b}\int r(g(u))g(u)\mathrm{d}u \\
&= \frac{1}{2} - \frac{\nu}{2b} + \frac{k-(1-\widetilde{T})}{2b} - \frac{1}{2bc}\int [g(u)]^2\,\mathrm{d}u,
\end{aligned}$$

where we use the facts that $\int ug(u)\mathrm{d}u = 0$ as $g(\cdot)$ is centered around zero and $\nu = \mu - (1-\widetilde{T})$.

D Probability of democratization and noise variance

We formally show that the probability of democratization increases as the variance in the distribution of the noise increases. First, notice that:

$$\begin{aligned}
\int [g(u)]^2 \, \mathrm{d}u &= \int \left[\frac{1}{\sigma_u\sqrt{2\pi}} \mathrm{e}^{-\frac{u^2}{2\sigma_u^2}}\right]^2 \mathrm{d}u \\
&= \frac{1}{2\sigma_u^2\pi} \int \mathrm{e}^{-\frac{u^2}{\sigma_u^2}} \mathrm{d}u \\
&= \frac{1}{2\sigma_u\pi} \int \mathrm{e}^{-x^2} \mathrm{d}x \\
&= \frac{1}{2\sqrt{\pi\sigma_u^2}},
\end{aligned}$$

where we use the fact that u is symmetric, we change variable with $x = u/\sigma_u$, implying $\mathrm{d}u = \sigma_u \mathrm{d}x$, and we use the Gaussian integral defined by $\int_{-\infty}^{+\infty} \mathrm{e}^{-x^2} \mathrm{d}x = \sqrt{\pi}$. Hence,

$$\Pr(\text{democracy}; k) = \frac{1}{2} - \frac{\nu}{2b} + \frac{k - (1 - \widetilde{T})}{2b} - \frac{1}{4bc\sqrt{\pi\sigma_u^2}}.$$

Therefore,

$$\frac{\partial \Pr(\text{democracy}; k)}{\partial \sigma_u^2} = \frac{1}{4bc\sqrt{\pi}\left(\sigma_u^2\right)^{1.5}} > 0.$$

E Derivation of the equilibrium threshold

To derive the optimal threshold, we first show that the expected type conditional on a signal at the threshold can be expressed as:

$$\mathbb{E}\left[\theta|s=k\right] = k + \frac{1}{c}\int ug'(u)g(u)\mathrm{d}u,$$

under the assumption that the distribution of the noise has bounded support on $[-D, D]$, where D is defined as follows with k^0 defined below:

Assumption 2. $D \leq \min\left\{1-k^0, k^0\right\}$.

Using Bayes' rule we can write:

$$\mathbb{E}\left[\theta|s=k\right] = \frac{\int \theta(m,k)g(k-m)h(m,k)\mathrm{d}m}{\int g(k-m)h(m,k)\mathrm{d}m}.$$

Focus on the denominator first. Change the variable of integration m to $u = k - m$. Therefore, $\mathrm{d}m = -\mathrm{d}u$. This gives the following expression for the denominator:

$$\int_0^1 \theta(m,k)g(k-m)h(m,k)\mathrm{d}m = \int_k^{k-1}\left[k-u-r(g(u))\right]g(u)\left[1+r'(g(u))g'(u)\right](-\mathrm{d}u).$$

Now, assumption 2 implies that for all $u \in [-D, D]$, $u \in [k-1, k]$. To see why, it is useful to compute the threshold used by the elites when they perfectly observe the autocrat's type. It is the threshold k^0 such that the probability of democratization is equal to the probability of democratization when there is no information asymmetry. This implies the following value for k^0:

$$k^0 = \frac{1-\widetilde{T}}{2b} + \frac{1}{2bc}\int \left[g(u)\right]^2 \mathrm{d}u.$$

This benchmark threshold is an upper bound for the equilibrium threshold. Intuitively, because the autocrat is able to send higher messages than his type, the elite will necessarily accept autocrats more often under asymmetric information than when they can perfectly observe the autocrat's type. Hence, the optimal threshold k^* must be lower than k^0. Therefore, assumption 2 guarantees that for all $u \in [-D, D]$, $u \in [k-1, k]$. Then, inverting the bounds, we have:

$$\int \theta(m,k) g(k-m) h(m,k) \mathrm{d}m = \int_{-D}^{D} \left[k - u - r(g(u))\right] g(u) \left[1 + r'(g(u)) g'(u)\right] \mathrm{d}u.$$

Moreover, $g(\cdot)$ is symmetric around zero, so $\int u g(u) \mathrm{d}u = 0$. Also, $g(\cdot)$ is an odd function, so its derivative is an even function. Thus, $g(u)g'(u)$ is an odd function. Further, $r(g(u))$ is odd, and $r'(g(u))$ is even. So $g(u)r'(g(u))g'(u)$ is odd, and $g(u)g'(u)r(g(u))r'(g(u))$ is odd. Hence, $\int g(u) r'(g(u)) g'(u) \mathrm{d}u = \int g(u) g'(u) r(g(u)) r'(g(u)) \mathrm{d}u = 0$. Therefore,

$$\int \theta(m,k) g(k-m) h(m,k) \mathrm{d}m = k - \int r(g(u)) g(u) \mathrm{d}u - \int u g(u) r'(g(u)) g'(u) \mathrm{d}u.$$

Using the same argument, we get that the denominator equals one. Therefore,

$$\mathbb{E}\left[\theta | s = k\right] = k - \int r(g(u)) g(u) \mathrm{d}u - \int u g(u) r'(g(u)) g'(u) \mathrm{d}u.$$

Now integrate by parts $\int u g(u) r'(g(u) g'(u) \mathrm{d}u$, with the primitive of $r'(g(u))g'(u)$ being $r(g(u))$:

$$\begin{aligned} \int u g(u) r'(g(u)) g'(u) \mathrm{d}u &= -\int r(g(u)) \left[g(u) + u g'(u)\right] \mathrm{d}u \\ &= -\int r(g(u)) g(u) \mathrm{d}u - \int u g'(u) r(g(u)) \mathrm{d}u, \end{aligned}$$

because $\left[u g(u) r(u))\right]_{-D}^{D} = 0$. With the quadratic cost function, this equation becomes:

$$\frac{1}{c} \int u g'(u) g(u) \mathrm{d}u = -\frac{1}{c} \int \left[g(u)\right]^2 \mathrm{d}u - \frac{1}{c} \int u g'(u) g(u) \mathrm{d}u. \tag{E.1}$$

Therefore,

$$\mathbb{E}\left[\theta | s = k\right] = k + \int u g'(u) r(g(u)) \mathrm{d}u.$$

From there, we get the following expression for the equilibrium threshold:

$$k^* = 1 - \widetilde{T} - \int u g'(u) r\left(g(u)\right) \mathrm{d}u = 1 - \widetilde{T} - \frac{1}{c} \int u g'(u) g(u) \mathrm{d}u.$$

F Setting the equilibrium threshold

We provide here graphical intuition for how the elites set the equilibrium threshold. First, recall that the equilibrium threshold k^* is such that the elites must be indifferent between installing the autocrat in power and pushing for democratization whenever they observe a signal exactly at the threshold (i.e., when $\mathbb{E}[\theta|s = k^*] = 1 - \widetilde{T}$). The x-axis in Figure F.1a represents threshold levels, and the y-axis the expected type given that the signal is exactly at the threshold (i.e., $\mathbb{E}[\theta|s = k]$). We show how the equilibrium threshold changes when the incentives of the autocrat to exert effort in sending higher messages change. In particular, we focus on the case in which the variance of the noise σ_u^2 increases, so that autocrats have incentives to increase their message level – see Figure 2b in the paper. We illustrate how this induces the elites to adjust the threshold downwards because they expect to receive signals that are lower on average.

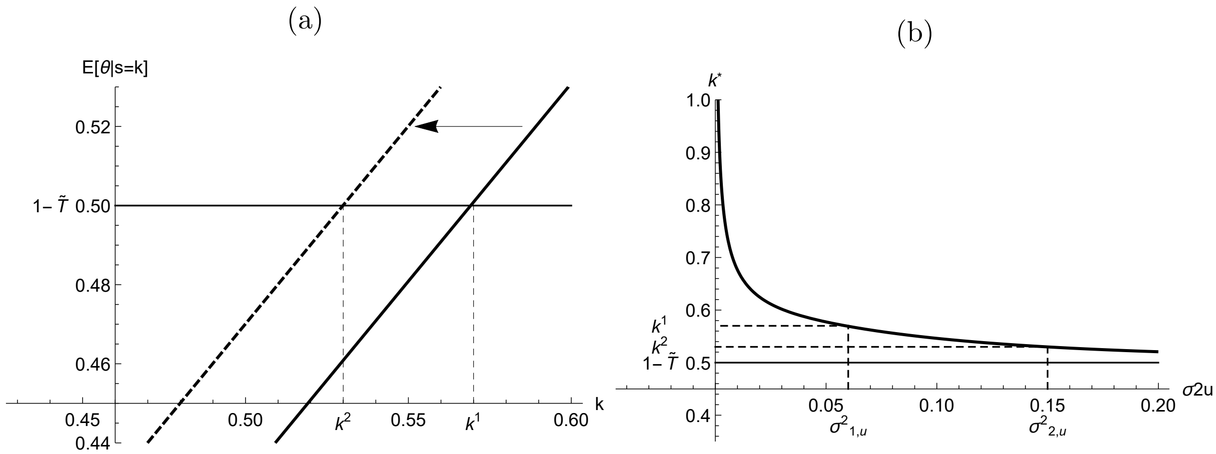
Suppose initially that the variance in the distribution of the noise is low at $\sigma_{1,u}^2 = 0.06$.[3] The expected taxation level under democracy is $\widetilde{T} = 0.5$, so that the expected payoff for the elites under democracy is $1 - \widetilde{T} = 0.5$. Graphically, the equilibrium threshold is determined by the intersection of the curve representing $\mathbb{E}[\theta|s = k]$ and the horizontal line of level $1-\widetilde{T}$. With the parameter values at hand, the equilibrium threshold is $k^1 = 0.57$ (solid line). As explained in the paper after Equation 4, $k^1 > 0.5 = 1 - \widetilde{T}$.

Now, suppose that the variance in the distribution of the noise increases to $\sigma_{2,u}^2 = 0.15$. As illustrated in Figure 2b, this reduces the incentives for the potential autocrats to exert effort in sending higher messages, which reduces the levels of signals, leading to less types being selected and thus to a lower likelihood of democratization. However, the elites anticipate this change in the behavior of potential autocrats. Therefore, they adjust the threshold downwards. With the parameter values at hand, the equilibrium threshold decreases to $k^2 = 0.53$ (dashed line). Overall, higher variance in the distribution of the noise decreases the

[3]For ease of comparison, we use the same parameters as in Figure 2b in the paper.

equilibrium threshold. As we can see on Figure F.1b, the equilibrium threshold approaches the payoff under democracy $1 - \widetilde{T}$ from above as the variance in the distribution of the noise increases.

Figure F.1: Equilibrium threshold



Parameters: $\sigma^2_{1,u} = 0.06$ (solid line), $\sigma^2_{2,u} = 0.15$ (dashed line), $c = 8$, $\widetilde{T} = 0.6$.

G Proof of Proposition 3

Using the expression of the probability of democratization in Proposition 2 along with Equation E.1, we can write the probability of democratization as

$$\Pr(\text{democracy}) = \frac{1}{2} - \frac{\nu}{2b} + \frac{1}{2bc}\int ug'(u)g(u)\mathrm{d}u.$$

We start by proving the first comparative statics, $\partial \Pr(\text{democracy})/\partial\sigma_\theta^2 > 0$:

$$\frac{\partial \Pr(\text{democracy})}{\partial\sigma_\theta^2} = \frac{\partial \Pr(\text{democracy})}{\partial b}\frac{\partial b}{\partial\sigma_\theta^2} = \left[\frac{\nu}{2b^2} - \frac{1}{2b^2c}\int ug'(u)g(u)\mathrm{d}u\right]\frac{\partial b}{\partial\sigma_\theta^2}.$$

$\partial b/\partial\sigma_\theta^2 = 3/2\,(3\sigma_\theta^2)^{1/2} > 0$ and $\int ug'(u)g(u)\mathrm{d}u < 0$ so $\partial \Pr(\text{democracy})/\partial\sigma_\theta^2$ is strictly positive.

We next prove the second comparative statics, $\partial \Pr(\text{democracy})/\partial\sigma_u^2 > 0$. The distribution of the noise has bounded support on $[-D, D]$. Therefore, the distribution $g(\cdot)$ can be expressed as:

$$g(u) = \frac{\frac{1}{\sigma_u}\phi\left(\frac{u}{\sigma_u}\right)}{\Phi\left(\frac{D}{\sigma_u}\right) - \Phi\left(-\frac{D}{\sigma_u}\right)} = \frac{\mathrm{e}^{-\frac{u^2}{2\sigma_u^2}}}{\sigma_u\sqrt{2\pi}\,\mathrm{erf}\left(\frac{D}{\sqrt{2\sigma_u^2}}\right)},$$

where we use the error function defined by $\mathrm{erf}(x) = (2/\sqrt{\pi})\int_0^x \mathrm{e}^{-t^2}\mathrm{d}t$. Using a change of variable with $x = u/\sigma_u$ and the fact that the error function is an odd function, we can compute the following:

$$\int_{-D}^{D} ug'(u)g(u)\mathrm{d}u \;=\; -\frac{1}{2\pi\sigma_u^2\left[\mathrm{erf}\left(\frac{D}{\sqrt{2\sigma_u^2}}\right)\right]^2}\int_{-D}^{D}\frac{u^2}{\sigma_u^2}\mathrm{e}^{-\frac{u^2}{\sigma_u^2}}\mathrm{d}u$$

$$= -\frac{1}{2\pi\sigma_u^2\left[\operatorname{erf}\left(\frac{D}{\sqrt{2\sigma_u^2}}\right)\right]^2}\int_{-\frac{D}{\sigma_u}}^{\frac{D}{\sigma_u}} \sigma_u x^2 \mathrm{e}^{-x^2}\mathrm{d}x$$

$$= \frac{1}{4\pi\sigma_u\left[\operatorname{erf}\left(\frac{D}{\sqrt{2\sigma_u^2}}\right)\right]^2}\int_{-\frac{D}{\sigma_u}}^{\frac{D}{\sigma_u}} x\left(-2x\mathrm{e}^{-x^2}\right)\mathrm{d}x$$

$$= \frac{1}{4\pi\sigma_u\left[\operatorname{erf}\left(\frac{D}{\sqrt{2\sigma_u^2}}\right)\right]^2}\left[xe^{-x^2}\Big|_{-\frac{D}{\sigma_u}}^{\frac{D}{\sigma_u}} - \int_{-\frac{D}{\sigma_u}}^{\frac{D}{\sigma_u}} \mathrm{e}^{-x^2}\mathrm{d}x\right]$$

$$= \frac{1}{4\pi\sigma_u\left[\operatorname{erf}\left(\frac{D}{\sqrt{2\sigma_u^2}}\right)\right]^2}\left[\frac{2D}{\sigma_u}\mathrm{e}^{-\frac{D^2}{\sigma_u^2}} - \sqrt{\pi}\operatorname{erf}\left(\frac{D}{\sqrt{\sigma_u^2}}\right)\right]$$

$$= \frac{1}{\left[\operatorname{erf}\left(\frac{D}{\sqrt{2\sigma_u^2}}\right)\right]^2}\left[\frac{D}{2\pi\sigma_u^2}\mathrm{e}^{-\frac{D^2}{\sigma_u^2}} - \frac{1}{4\sqrt{\pi\sigma_u^2}}\operatorname{erf}\left(\frac{D}{\sqrt{\sigma_u^2}}\right)\right].$$

Therefore, we can write explicitly the probability of democratization as

$$\Pr(\text{democracy}) = \frac{1}{2} - \frac{\nu}{2b} + \frac{1}{2bc}\frac{\frac{D}{2\pi\sigma_u^2}\mathrm{e}^{-\frac{D^2}{\sigma_u^2}} - \frac{1}{4\sqrt{\pi\sigma_u^2}}\operatorname{erf}\left(\frac{D}{\sqrt{\sigma_u^2}}\right)}{\left[\operatorname{erf}\left(\frac{D}{\sqrt{2\sigma_u^2}}\right)\right]^2}.$$

Let A denote the last term of the above equation:

$$A = \frac{\frac{D}{2\pi\sigma_u^2}\mathrm{e}^{-\frac{D^2}{\sigma_u^2}} - \frac{1}{4\sqrt{\pi\sigma_u^2}}\operatorname{erf}\left(\frac{D}{\sqrt{\sigma_u^2}}\right)}{\left[\operatorname{erf}\left(\frac{D}{\sqrt{2\sigma_u^2}}\right)\right]^2}. \tag{G.1}$$

We then have the following comparative statics:

$$\frac{\partial \Pr \text{democracy}}{\partial \sigma_u^2} = \frac{1}{2bc} \frac{\partial A}{\partial \sigma_u^2}.$$

We can show numerically that under the restriction that D and σ_u^2 are positive (see the Mathematica Notebook in Appendix J below) :

$$\underset{\{D, \sigma_u^2\} \in \mathbb{R}_+^*}{\arg\min} \frac{\partial A}{\partial \sigma_u^2} > 0.$$

More specifically, the numerical solution for the minimization problem is:

$$\underset{\{D, \sigma_u^2\} \in \mathbb{R}_+^*}{\arg\min} \frac{\partial A}{\partial \sigma_u^2} = 5.8 \times 10^{-58}, \text{ as } D \longrightarrow 1.4 \times 10^{23} \text{ and } \sigma_u^2 \longrightarrow 2.5 \times 10^{37}.$$

Therefore, for all values of the parameters, $\partial \Pr(\text{democracy})/\partial \sigma_u^2 > 0$.

H Costly autocrat replacement

We show here that the initial decision the economic elites faces in choosing an autocrat is of first-order importance, implying that a static model is sufficient to analyze the dynamics outlined in the paper. Intuitively, this is because a bad first draw increases the likelihood that the elites will be stuck in an equilibrium with a low type autocrat whose policies run contrary to elite interests. For ease of exposition, we formally illustrate this point in a simplified version of the model in which autocrats cannot signal their types to the economic elites. There are two periods, $t = 1, 2$. In period 1, the elites make their decision without observing the autocrat's type. That is, they compare their expected payoff under dictatorship and democracy, μ versus $1 - \widetilde{T}$. As in the paper, we focus on the case in which the elites generally prefer autocracy, i.e., $\nu = \mu - \left(1 - \widetilde{T}\right) > 0$ (see footnote 9). The elite's payoff at the end of period 1 is the realized value of the chosen autocrat's type, θ.

At the beginning of period 2, the economic elites may choose to keep the current autocrat or draw a new one.[4] However, a new draw requires ousting the current autocrat through a coup. A coup has a fixed cost c_p, and its probability of success is a function of the elites' first draw $\lambda(\theta)$, with $\lambda'(\cdot) > 0$, $\lambda(0) = 0$ and $\lambda(1) = 1$. That is, a coup is more likely to succeed when the elite is more powerful and have a higher θ. Therefore, at the beginning of period 2, the elites will choose to foment a coup in order to get a new draw if the following holds:

$$\lambda(\theta)\mu + [1 - \lambda(\theta)]\,\theta - c_p \geq \theta,$$

where, as at the beginning of period 1, the elites' expectation over the new draw is μ. We therefore assume that the first draw is not informative of the second one (i.e., draws are i.i.d.

[4]We abstract from the possibility of pushing for democratization for the sake of the argument here; the mechanism is the same when allowing for that possibility.

across periods). We define a "Coup Constraint" function as:

$$CC\left(\theta;\mu,c_p\right)=\lambda(\theta)(\mu-\theta)-c_p. \tag{H.1}$$

When the coup constraint function is positive, it is incentive compatible for the elites to attempt a new draw. Otherwise, they are stuck with their draw from period 1.

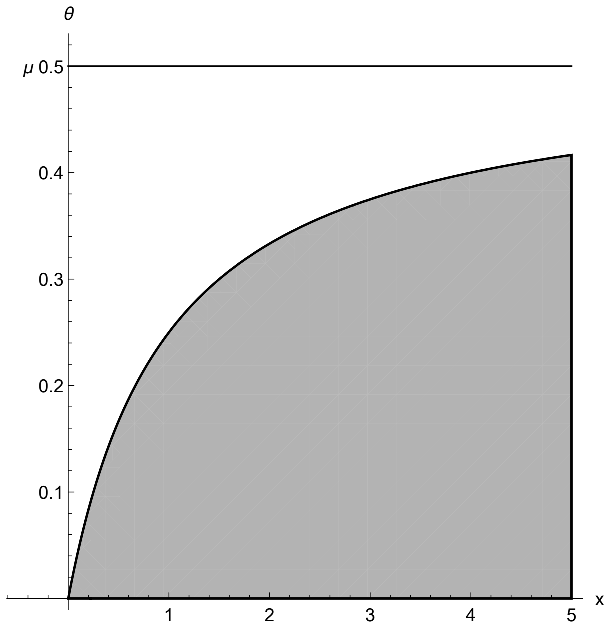
The main point is that whenever the elites get a low draw θ in period 1, this coup constraint is less likely to be met. That is, $\partial CC\left(\theta;\mu,c_p\right)/\partial(-\theta)<0$. How low the first draw needs to be for the elites to be stuck in the bad equilibrium in turn depends on the coup technology: if it exhibits strongly increasing returns, $\lambda(\cdot)$ is strongly convex, and a draw just below the mean μ implies that it is not incentive compatible for the elites to attempt a new draw. To see why, consider a general coup technology of the form $\lambda(\theta)=\theta^x$. Then,

$$\frac{\partial CC\left(\theta;\mu,c_p,x\right)}{\partial(-\theta)}<0\iff\lambda(\theta)-\lambda'(\theta)(\mu-\theta)<0\iff\theta<\frac{x}{1+x}\mu.$$

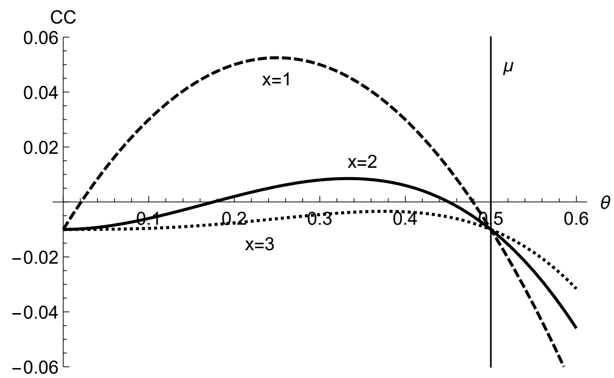
Because $\partial\left[x/(1+x)\right]/\partial x>0$, the stronger the returns in the coup technology, the larger the set of draws in period 1 that will make the elites less likely to be able to switch autocrat. At the limit, any draw below the mean μ will lead the elites to be stuck with the current autocrat. We illustrate this point in Figure H.1a where we show the set of first period draws for which the elites are less likely to be able to switch the autocrat as a function of the magnitude in the returns in the coup technology x when $\mu=0.5$.

We then show in Figure H.1b several simulations of the coup constraint defined in Equation H.1 as a function of the draw in the first period when $\mu=0.5$ and the coup cost is equal to 1% of the endowment flow (i.e., $c_p=0.01$). We do this for three levels of returns in the coup technology: a linear technology (dashed curve, $x=1$), a quadratic technology (solid curve, $x=2$) and a cubic technology (dotted curve, $x=3$). As one can see, the lower the initial draw, the more likely it is for the elites to have a negative coup function, and therefore to be stuck in the bad equilibrium. As suggested by Figure H.1a, the larger the returns in the coup technology, the more likely it is that the coup function is negative. Finally, note that the coup function is always negative when the draw is above the mean μ. This is because the elites are happily "stuck" in the good equilibrium in this case.

Figure H.1: Costly autocrat replacement



(a) Set of draws for which $\partial CC/\partial(-\theta) < 0$



(b) Coup Constraint Function

Parameters: $\mu = 0.5$, $c_p = 0.01$

I Allowing for an arbitrary number of periods

We show here that introducing an arbitrary number of periods to the model does not change its fundamental implications. For simplicity of the exposition, we first abstract from signalling and suppose that the elites observe perfectly the type of the potential autocrat once its value is realized. We show at the end of this appendix that the results also hold when signaling is allowed. We assume throughout that autocrats only live one period so that the elites can draw a new autocrat in each period without paying a replacement cost – we therefore abstract from the case examined in appendix H.

In each period t, the autocrat's type θ_t is realized and observed by the elites, who then decide whether to install him in power ($a_t = 1$) or to push for democratization ($a_t = 0$). If $\theta_t > 1 - \widetilde{T}$, the elites install the autocrat in power. The probability of democratization in period t is:

$$\begin{aligned}
\Pr\left(\text{democracy}_t\right) &= \Pr\left(a_1 = 1 \cap \ldots \cap a_{t-1} = 1 \cap a_t = 0\right) \\
&= \Pr\left(a_t = 0\right) \prod_{\tau=1}^{t-1} \Pr\left(a_\tau = 1\right) \\
&= \Pr\left(\theta_t \leq 1 - \widetilde{T}\right) \prod_{\tau=1}^{t-1} \left[1 - \Pr\left(\theta_\tau \leq 1 - \widetilde{T}\right)\right] \\
&= \left(\frac{1}{2} - \frac{\nu}{2b}\right) \left(\frac{1}{2} + \frac{\nu}{2b}\right)^{t-1},
\end{aligned}$$

where we assume that draws are i.i.d. across periods. That is, a draw in a given period is not informative about draws in later periods. Now, because democratization in each period is a mutually exclusive event, the overall probability of democratization when there are T periods is:

$$\begin{aligned}
\Pr\left(\text{democracy}\right) &= \Pr\left(\text{democracy}_1 \cup \ldots \cup \text{democracy}_T\right) \\
&= \sum_{t=1}^{T} \Pr\left(\text{democracy}_t\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{T}\left(\frac{1}{2}-\frac{\nu}{2b}\right)\left(\frac{1}{2}+\frac{\nu}{2b}\right)^{t-1} \\
&= 1-\left(\frac{1}{2}+\frac{\nu}{2b}\right)^{T}.
\end{aligned}
$$

Therefore, the intuition that the likelihood of democratization increases as the variance in the pool of potential autocrats σ_θ^2 increases still holds:

$$
\frac{\partial \Pr(\text{democracy})}{\partial \sigma_\theta^2} = \frac{\partial \Pr(\text{democracy})}{\partial b}\frac{\partial b}{\partial \sigma_\theta^2} = \left[\frac{T\nu}{2b^2}\left(\frac{1}{2}+\frac{\nu}{2b}\right)^{T-1}\right]\frac{\partial b}{\partial \sigma_\theta^2} > 0,
$$

where $\partial b/\partial \sigma_\theta^2 = 3/2(3\sigma_\theta^2)^{1/2} > 0$.

Indeed, the implications of the model are even reinforced by the inclusion of multiple periods since this increases the likelihood that the elites will eventually have a bad draw and democratize:

$$
\frac{\partial \Pr(\text{democracy})}{\partial T} = -\log\left(\frac{1}{2}+\frac{\nu}{2b}\right)\left(\frac{1}{2}+\frac{\nu}{2b}\right)^{T} > 0,
$$

where we use the fact that $1/2 + \nu/2b < 1$.

Adding the possibility of signaling does not alter this logic. Using Proposition 2 along with Equation E.1, the probability of democratization when signalling is allowed and when there are T periods can be written as:

$$
\Pr(\text{democracy}) = 1 - \left[\frac{1}{2}+\frac{\nu}{2b}-\frac{1}{2bc}\int ug'(u)g(u)\mathrm{d}u\right]^{T}.
$$

Again, we have the following comparative statics:

$$
\frac{\partial \Pr(\text{democracy})}{\partial \sigma_\theta^2} = T\left[\frac{\nu}{2b^2}-\frac{1}{2b^2c}\int ug'(u)g(u)\mathrm{d}u\right]\left[\frac{1}{2}+\frac{\nu}{2b}-\frac{1}{2bc}\int ug'(u)g(u)\mathrm{d}u\right]^{T-1}\frac{\partial b}{\partial \sigma_\theta^2}.
$$

Using the results from Appendix G, we know that the second group of terms is strictly positive, so the comparative statics are also strictly positive.

We can also show that the comparative statics with respect to the variance of the noise still hold in this dynamic setting. Using the results from Appendix G and the term A as defined in equation G.1, we can write the probability of democratization as:

$$\Pr(\text{democracy}) = 1 - \left[\frac{1}{2} + \frac{\nu}{2b} - \frac{1}{2bc}A\right]^{T}.$$

Therefore, we get the following comparative statics:

$$\frac{\Pr(\text{democracy})}{\partial\sigma_u^2} = \frac{T}{2bc}\frac{\partial A}{\partial\sigma_u^2}\left[\frac{1}{2} + \frac{\nu}{2b} - \frac{1}{2bc}A\right]^{T-1} > 0,$$

where we use the result from Appendix G that $\partial A/\partial\sigma_u^2 > 0$.

J *Mathematica* Code for the Proof of Proposition 3

In[1]:=
```
A[d_, σ2_] = ((d*Exp[-d^2/σ2])/(2*π*σ2) - Erf[d/Sqrt[σ2]]/(4*Sqrt[π*σ2]))/(Erf[d/Sqrt[2*σ2]])^2
```

Out[1]=
$$\frac{\frac{d\, e^{-\frac{d^2}{\sigma 2}}}{2\,\pi\,\sigma 2} - \frac{\mathrm{Erf}\left[\frac{d}{\sqrt{\sigma 2}}\right]}{4\sqrt{\pi}\,\sqrt{\sigma 2}}}{\mathrm{Erf}\left[\frac{d}{\sqrt{2}\,\sqrt{\sigma 2}}\right]^2}$$

In[2]:=
```
dA[d_, σ2_] = D[A[d, σ2], σ2]
```

Out[2]=
$$\frac{d\, e^{-\frac{d^2}{2\,\sigma 2}}\sqrt{\frac{2}{\pi}}\left(\frac{d\, e^{-\frac{d^2}{\sigma 2}}}{2\,\pi\,\sigma 2} - \frac{\mathrm{Erf}\left[\frac{d}{\sqrt{\sigma 2}}\right]}{4\sqrt{\pi}\,\sqrt{\sigma 2}}\right)}{\sigma 2^{3/2}\,\mathrm{Erf}\left[\frac{d}{\sqrt{2}\,\sqrt{\sigma 2}}\right]^3} + \frac{\frac{d^3\, e^{-\frac{d^2}{\sigma 2}}}{2\,\pi\,\sigma 2^3} - \frac{d\, e^{-\frac{d^2}{\sigma 2}}}{4\,\pi\,\sigma 2^2} + \frac{\mathrm{Erf}\left[\frac{d}{\sqrt{\sigma 2}}\right]}{8\sqrt{\pi}\,\sigma 2^{3/2}}}{\mathrm{Erf}\left[\frac{d}{\sqrt{2}\,\sqrt{\sigma 2}}\right]^2}$$

In[3]:=
```
assumptions = d > 0 && σ2 > 0
```

Out[3]= d > 0 && σ2 > 0

In[4]:=
```
NMinimize[{dA[d, σ2], assumptions}, {d, σ2}]
```

Out[4]= $\{5.76058 \times 10^{-58}, \{d \to 1.43497 \times 10^{23}, \sigma 2 \to 2.46554 \times 10^{37}\}\}$

References

Caselli, Francesco, Tom Cunningham, Massimo Morelli & Inés Moreno Barreda. 2014. "The Incumbency Effects of Signalling." *Economica* 81(323):397–418.

Cunningham, Tom & Inés Moreno de Barreda. 2015. "Equilibrium Persuasion." *Mimeo* http://www.economics.ox.ac.uk/materials/events/13655/cm2015oxford.pdf.

Milgrom, Paul R. 1981. "Good News and Bad News: Representation Theorems and Applications." *The Bell Journal of Economics* 12(2):380–391.