8 Appendix: Not for publication

8.1 $z \in (\beta/4, \hat{z})$

Proposition 2 *Assume that (30) and $z \in (\beta/4, \hat{z})$. Then, the Markov perfect equilibrium is such that*

$$\tau^y(s) = \begin{cases} \frac{1}{2} & \text{if } s \leq s^2 \\ \lambda^-(s) & \text{if } s \in \left[s^2, s^R\right] \\ \frac{1+\beta}{2} & \text{if } s > s^R \end{cases}, \tag{54}$$

$$h = \begin{cases} \frac{1}{4} & \text{if } s \leq s^2 \\ \frac{1}{2} - (\rho s - z) & \text{if } s \in \left[s^2, s^R\right] \\ \frac{1+\beta}{4} & \text{if } s > s^R \end{cases}. \tag{55}$$

The proof is straightforward and immediately follows from Lemma 3, (23), (34) and (36). One can see that the implications of Proposition 2 are qualitatively the same as those of Propositions 1. It is worth noting that $\pi(\tau^y(s), s) = 0$ or 1. That is to say, there is no electoral uncertainty under a small z.

8.2 Vote Shares and the Size of Government

This subsection investigates the main empirical prediction of our model, which will be tested in Section 5. Although (36) has provided a prediction on the correlation between ideology and τ^y, there are two major difficulties in testing the prediction. The first one is how to measure ideology. A commonly used measure of ideology in political science literature is self-placement scores of the left-right position from opinion polls or survey data (Inglehart, 1990). This approach obviously suffers from limited comparative observations across countries and time.[40] Second, it is equally hard to find an empirical counterpart of τ^y, though age-dependent taxation contains some realistic components. Given these concerns, we adopt an alternative approach: looking at government size and vote shares, for which data can easily be obtained. Since s and τ^y are positively related to the right-wing's vote share and government size within each political regime, respectively, we expect the correlation between vote shares and government size qualitatively similar to that between s and τ^y.

To see this precisely, we aggregate the two types of taxes in (10) and (36), and then compute the size of government as a percentage of aggregate output, denoted by γ.

Corollary 1

$$\gamma \equiv \frac{2g}{y} = \begin{cases} \frac{h(s_{-1})+\tau^y(s)h(s)}{h(s_{-1})+h(s)} & \text{if } s \leq \frac{1}{2} - h(s_{-1}) \\ \frac{\tau^y(s)h(s)}{h(s_{-1})+h(s)} & \text{otherwise} \end{cases}, \tag{56}$$

[40] Moreover, it has long been questioned whether all respondents have consistent views on the location of the "left" and "right" (e.g., Levitin and Miller, 1979).

where $\tau^y(s)$ and $h(s)$ follow (36) and (37), respectively.

Note that γ is not only affected by the current ideological state s, but also depends on the past ideological state s_{-1}, which determines the current size of the old rich. Due to the predetermined size of the old rich, an analytical characterization of the correlation between γ and e (the right-wing's vote share) is not applicable. So, we first simulate the model and use simulated data to estimate the following linear equation.

$$\gamma = b^0 + b^1 R + b^2 e + \varepsilon, \tag{57}$$

where ε is an error term and R is a dummy variable which equals zero and one for the left- and right-wing regime, respectively. We run 1100 simulations 50 times with the benchmark parameterization.[41] The estimated results are $b^1 = -0.4950$ (-410.43) and $b^2 = 0.1010$ (55.16), with t statistics in brackets. Consistent with our theory, b^1 is negative and b^2 is positive, suggesting a negative partisan effect and a positive intertemporal effect, respectively. In addition, R^2 is 0.98, indicating a high degree of fitness of the linear specification.[42]

8.2.1 Sensitivity to Model Parameters

Now, we check the parameter sensitivity of the coefficient of interests, b^2. Specifically, we analyze sensitivity to two key model parameters: z and ρ.[43] Panel A of Figure 3 shows that an increase in z, implying more volatile ideology, tends to reduce b^2. This is consistent with (36); the intertemporal effect of ideology, captured by $\mathrm{d}\tau^y(s)/\mathrm{d}s = \beta\rho/4z$, mitigates as z increases. The effect of z on b^2, however, is non-monotonic. When z is sufficiently large, b^2 turns out to be increasing in z. A larger z increases the likelihood for s to fall into $\left(s_H^M, s^R\right]$, where the intertemporal effect gets stronger (recall that $\mathrm{d}\tau^y(s)/\mathrm{d}s = 2\rho$ for $s \in \left(s_H^M, s^R\right]$). This yields a larger b^2. The U-shaped estimates of b^2 in Panel A suggest that the second effect tends to dominate for large z.

[Insert Figure A-1]

Panel B shows the estimates of b^2 w.r.t. ρ. When $\rho = 0$, the intertemporal effect of ideology goes away. This implies a zero b^2. Adding persistence into the stochastic process of ideology ρ gives rise to the intertemporal effect and, therefore, a positive b^2. Similar to the effects of

[41] The first 100 observations are discarded to eliminate the effect of the initial ideological state.

[42] This exercise also suggests that the simple linear regression specification can be a good estimator of the intertemporal effect of ideology.

[43] When changing z or ρ, we recalibrate β accordingly so that the competitive political region remains symmetric around $s = 0$. See Appendix 7.4 for detailed discussion on calibration.

z on b^2, ρ also has two opposite effects on b^2. On the one hand, ρ increases the magnitude of the intertemporal effect, $\mathrm{d}\tau^y(s)/\mathrm{d}s$ for $s \in \left(s^1, s_H^M\right] \cup \left(s_H^M, s^R\right]$. On the other hand, the region $\left(s^1, s_H^M\right] \cup \left(s_H^M, s^R\right]$ shrinks with a larger ρ, suggesting a lower likelihood for the intertemporal effect to be functioning. This reduces the estimate of b^2. Different from the U-shaped estimates of b^2 w.r.t. z, Panel B shows that the aggregate effect of ρ seems monotonic; i.e., the positive effect always dominates the negative one.

To conclude, we find a positive b^2 , implying that an increase in the right-wing voter share leads to a larger government within each political regime. The result is robust to a wide range of parameter values. This allows us to test our theory against the standard partisan theory predicting a zero b^2. The empirical analysis will be conducted in Section 5.

8.3 Age-Independent Taxation

Throughout the paper, we maintain the assumption that the government can condition taxes on age. Although age-dependent taxation has its realistic counterpart and substantially simplifies the analysis, this assumption is not innocuous. One may wonder whether binary taxation (10), which obviously overstates the partisan effect of ideology, is crucial for the positive relationship between s and τ^y. An earlier version of this paper (Song, 2005, chapter 2) assessed the robustness of the main result under age-independent taxation and found that imposing the weaker policy instrument does not lead to any major change. The intuition is simple. Age-independent tax rates in the right-wing regime are, on average, lower than those in the left-wing regime. A right-wing ideology, as in our benchmark case with age-dependent taxation, reduces the expected tax rate by increasing π. This encourages investment and induces the incumbent to behave in a way similar to that described above.

8.4 Alternative Political Objectives

We have characterized equilibria of our benchmark model in which only the old vote. This assumption seems too extreme, as (old) politicians entirely ignore the welfare of the young in policy decision-making. A natural extension is to assume that politicians (party candidates) are altruistic towards the young. The political objective function can, therefore, be written as

$$W^L = u^{ou} + \hat{\omega} u^y, \quad (58)$$

$$W^R = u^{os} + \hat{\omega} u^y, \quad (59)$$

where $\hat{\omega} \geq 0$ is the intensity of altruism. Note that different from (7) and (8), here the altruism is independent of economic situation. Appendix 8.5 proves the following proposition.

Proposition 3 *Assume political objective functions (58) and (59). Then,*

(i) τ^o *follows (10).*

(ii) h *and* π *satisfy (16) and (21), respectively,*

(iii) Define

$$V \equiv \tau^y h + \omega \left(\underbrace{(1-\tau^y+\beta\pi)\, h}_{\textit{life-time earnings}} + \underbrace{\frac{a^o}{2}\beta\,(1-\pi)\, h}_{\textit{redistributive benefits}} - \underbrace{h^2}_{\textit{investment costs}} \right). \tag{60}$$

Then, τ^y *solves*

$$\tau^y = \arg\max_{\tau^y \in [0,1]} V. \tag{61}$$

(iv) The political outcomes associated with (58) and (59) are identical to those under probabilistic voting à la Lindbeck and Weibull (1987), in which politicians maximize a weighted average of individual utilities:

$$(h_{-1} + \hat{s})\, u^{os} + (1 - h_{-1} - \hat{s})\, u^{ou} + \hat{\omega} u^y,$$

where $\hat{s} \equiv s + (a^o + \hat{\omega} a^y - 1)/2$.

Four remarks are in order. First, the first and second parts of the proposition come directly from age-dependent taxation. Second, when choosing τ^y, politicians face a trade-off between tax revenues and the welfare of the young, shown as a sum of their lifetime earnings, redistributive benefits and costs of human-capital investment in (60).[44] Third, the last part of the proposition shows that the probabilistic voting *à la* Lindbeck and Weibull (1987) can be considered a micro-foundation for political objective functions (58) and (59). In particular, the political weight on the left-wing (right-wing) is equal to the population of the poor (rich) plus (minus) the ideological state (subject to a linear transformation). Finally, the ideology-independent altruism may originate from binding commitments to electoral platforms. If the left-wing (right-wing) cared about the young poor (rich) only and commitments were not binding, they would naturally maximize (7) and (8), rather than (58) and (59).[45]

The solid lines in Figure A-2 depict the equilibrium policy rules of τ^y, π and h. As in Figure 3, the dotted lines are those from the benchmark model in Figure 1. The results are qualitatively similar. This politico-economic equilibrium also features an increasing τ^y as ideology leans towards the right in the competitive political region. The human-capital investment h is, again, non-monotonically related to s. Nevertheless, two quantitative changes

[44] $(1-\pi)\,h$ in the middle term of the bracket in (60) is the next-period redistribution if the left-wing party wins the election.

[45] I thank a referee for pointing out the distinction.

deserve mention. First, τ^y is lower than that in the benchmark model. Note that ω may be a reflection of young population share. So a prediction of the model is that societies with an older population have larger governments, which is consistent with our empirical finding reported in footnote 32. Second, the competitive political region moves towards the left, and both π and h become larger. In other words, the right-wing party will be more likely to win the election when politicians are altruistic towards the young. The intuition is straightforward: The young dislike distortionary taxes. Therefore, altruistic politicians have the incentive to cut τ^y, resulting higher h and π.

[Insert Figure A-2]

8.5 Proof of Proposition 3

Using (1) to (4), (58) can be written as

$$\begin{aligned} W^L &= a^o g + \hat{\omega}\left(h\left(1-\tau^y\right) + a^y g - h^2 + \beta h E\left[1-\tau^{o\prime} + a^o g'\right] + \beta\left(1-h\right) a^o E\left[g'\right]\right) \\ &= \left(a^o + \hat{\omega} a^y\right) g + \hat{\omega}\left(h\left(1-\tau^y\right) - h^2 + a^o \beta E\left[g'\right] + \beta h \pi\right) \\ &= \frac{\left(a^o + \hat{\omega} a^y\right)}{2}\left(h_{-1}\tau^o + h\tau^y\right) + \hat{\omega}\left(h\left(1-\tau^y\right) - h^2 + \frac{a^o\beta}{2}\left(h\left(1-\pi\right)\right) + h\beta\pi\right). \end{aligned}$$

(6) is used to replace g and g' in the third line. We also drop the irrelevant term, $h'\tau^{y\prime}$, as it is fully determined by the next-period young households and government and, thus, independent of τ^o and τ^y. Similarly, (59) can be written as

$$\begin{aligned} W^R &= 1 - \tau^o + a^o g + \hat{\omega}\left(h\left(1-\tau^y\right) + a^y g - h^2 + \beta h E\left[1-\tau^{o\prime} + a^o g'\right] + \beta\left(1-h\right) a^o E\left[g'\right]\right) \\ &= 1 - \tau^o + \frac{\left(a^o + \hat{\omega} a^y\right)}{2}\left(h_{-1}\tau^o + h\tau^y\right) + \hat{\omega}\left(h\left(1-\tau^y\right) - h^2 + \frac{a^o\beta}{2}\left(h\left(1-\pi\right)\right) + h\beta\pi\right). \end{aligned}$$

Under probabilistic voting, the party candidates maximize

$$\begin{aligned} W &= \left(1 - h_{-1} - \hat{s}\right)\left(a^o g\right) + \left(h_{-1} + \hat{s}\right)\left(1 - \tau^o + a^o g\right) \\ &\quad + \hat{\omega}\left(h\left(1-\tau^y\right) + a^y g - h^2 + \beta h E\left[1-\tau^{o\prime} + a^o g'\right] + \beta\left(1-h\right) a^o E\left[g'\right]\right) \\ &= \left(h_{-1} + \hat{s}\right)\left(1-\tau^o\right) + \frac{\left(a^o + \hat{\omega} a^y\right)}{2}\left(h_{-1}\tau^o + h\tau^y\right) \\ &\quad + \hat{\omega}\left(h\left(1-\tau^y\right) - h^2 + \frac{a^o\beta}{2} h\left(1-\pi\right) + \beta h \pi\right). \end{aligned}$$

Clearly, maximizing W w.r.t. τ^o yields

$$\tau^o = \begin{cases} 1 & \text{if } \hat{s} + h_{-1} < \frac{a^o + \hat{\omega} a^y}{2} \\ 0 & \text{otherwise} \end{cases}.$$

The definition of $\hat{s}$ implies that the above equation is identical to (10). Then, the problem for τ^y reduces to (61).

Figure A-1: Sensitivity Analysis

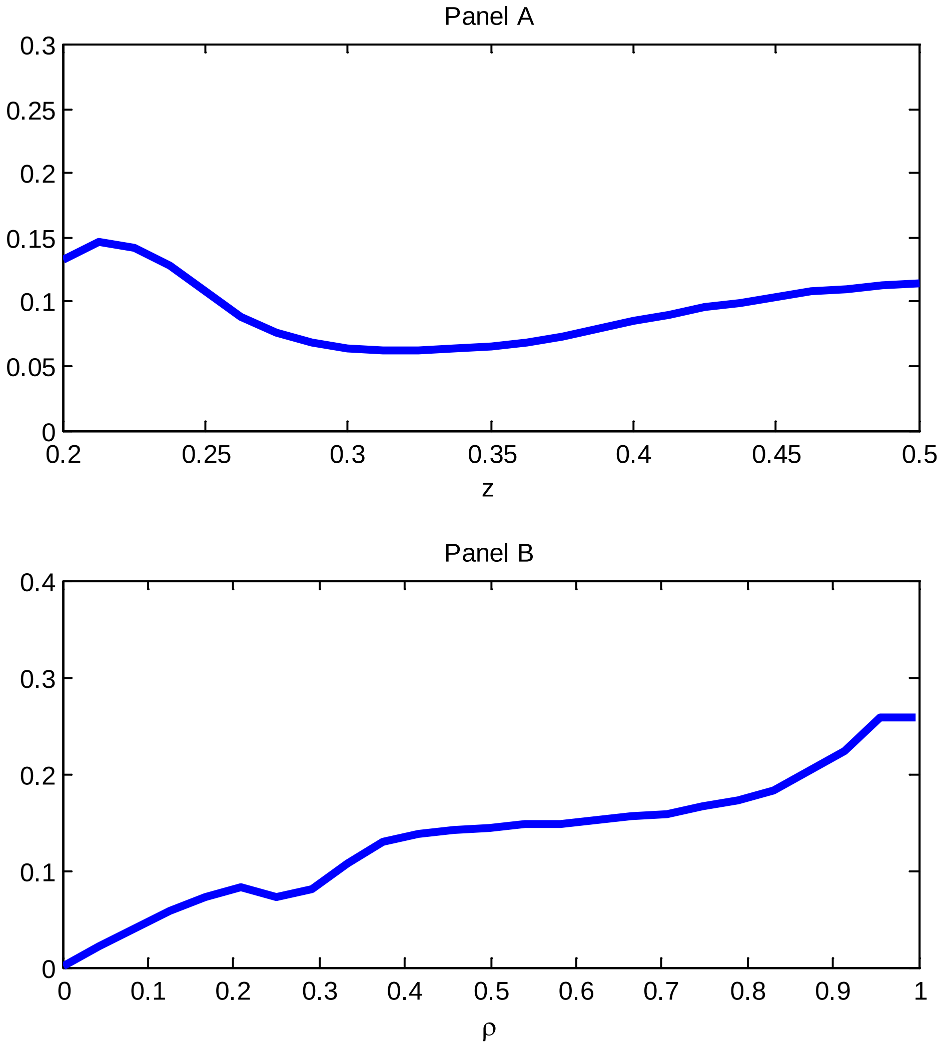

Figure A-1: Panel A and B plot the estimated b^2 with respect to z and ρ, respectively. When changing z or ρ, we recalibrate β accordingly so that the competitive political region remains symmetric at $s = 0$.

Figure A-2: Equilibrium Results with Ideology-Independent Altruism

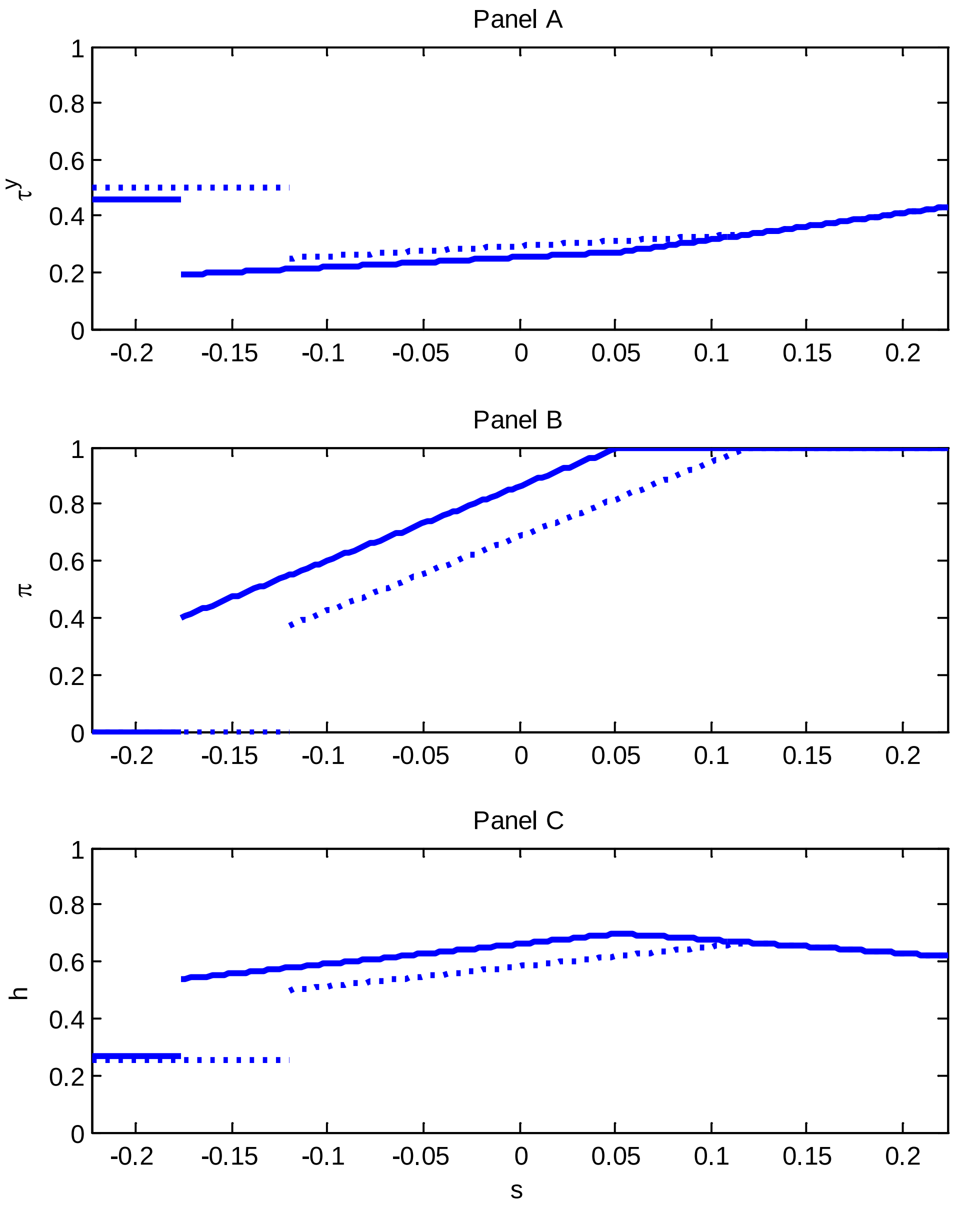

Figure A-2: Solid and dotted lines stand for equilibrium results with ideology-independent altruism and those in the benchmark case, respectively. Panel A represents the equilibrium policy rule $\tau^y(e,s)$. The probability for the right-wing to be elected, $\pi(\tau^y(e,s),s)$, is plotted in Panel B. Panel C corresponds to the equilibrium investment rule $h(\tau^y(e,s),s)$. $\omega = 0.1$ and the other parameter values are held constant as in the benchmark case.