Political Economics - Lecture notes

Kristian Urup Olesen Larsen[1]

[1]Department of Economics, University of Copenhagen

December 16, 2018

Abstract.

Table of Contents

This is my lecture notes for the course *political economics* taught at the Department of Economics at the University of Copenhagen. I've written these notes as a student, and as such some of the material might be wrong. If you find any mistakes please create an issue on github and i will correct it.

Part 1

1 Lecture 1 - Introduction

This course is about political *economics*, i.e. studying how political systems affect outcomes. The course is not about political *economy* which is the old-school Marx/Smith economics. Also this course is not about international political economics, which is concerned with the functions of international organizations and regimes.
What we want to do is to understand and explain differences and similarities in economic policy across time, varying political regimes and geography.

1.1 Political preferences and majority voting (the median voter model)

The median voter model is a workhorse of the field, and we will use it throughout the course. In it we assume that policy is determined by simple majority voting.
Consider a society with N agents, who each have quasi-linear utility w^i of consumption c^i and leisure x^i such that

$$w^i = c^i + V(x^i), \quad V' > 0, V'' < 0 \tag{1.1}$$

Individuals are all taxed by some fixed share $(1-q)$ of their income l^i, and in return receive a fixed transfer f, meaning we can write an individuals budget constraint as

$$c^i = (1-q)l^i + f \tag{1.2}$$

Normalizing the wage to 1 across all individuals we can simply read l^i as the labor supply. A simple way to introduce agent heterogeneity with singular wages is to let the available time for labor be heterogeneous, so instead of having $1 = l^i + x^i$, i.e. total time is fully spent either on labor or leisure we induce some variation in the total time available α^i

$$1 - \alpha^i = l^i + x^i \tag{1.3}$$

We denote the mean of α^i's in the population by α and the median by α^m.

1.1.1 Agent maximization problem

From the perspective of a single agent the tax rate q is fixed, so they simply solve

$$\begin{aligned} &\max_{l^i} c^i + V(x^i) \\ &\quad \text{s.t. } c^i = (1-q)l^i + f \\ &\quad \text{and } 1-\alpha^i = l^i + x^i \end{aligned} \tag{1.4}$$

Inserting the constraints and taking the derivative w.r.t l^i we get a first order condition of

$$(1-q) = V'(1-\alpha^i - l^i) \tag{1.5}$$

That is agents choose their labor supply to balance the net income gain from supplying an additional unit of labor with the marginal increase in utility from having one more unit of leisure. Isolating the labor supply we find that

$$\begin{aligned} l^{i*} &= 1-\alpha^i - V'^{-1}(1-q) \\ &= 1-\alpha^i - V'^{-1}(1-q) - \alpha + \alpha \\ &= L(q) - (\alpha^i - \alpha) \end{aligned} \tag{1.6}$$

where $L(q) \equiv 1-\alpha - V'^{-1}(1-q)$ can be shown to be the average labor supply in the population (in equilibrium). To see this notice

$$l \equiv \frac{1}{N}\sum_{i=1}^{N} l^i = \sum_{i=1}^{N} \left[L(q) - (\alpha^i - \alpha)\right] = L(q) \tag{1.7}$$

1.1.2 Indirect utility

Let us assume that the government runs a balanced budget, meaning the transfer individual must be equal to the taxes raised from the average individual, that is

$$f = qL(q) \tag{1.8}$$

The indirect utility of an agent is then utility which is achieved by optimally setting the labor supply given q, put intuitively: now we know how agents behave for any given q and we know how the government sets q, so we can write the indirect utility each agent gets for any choice of q:

$$\begin{aligned} W(q,\alpha^i) &= c^i + V(x^i) \\ &= (1-q)l^i + f + V(1-\alpha^i - l^i) \\ &= (1-q)\underbrace{\left[L(q) - (\alpha^i - \alpha)\right]}_{\text{optimal } l^i \text{ given } q} + \underbrace{qL(q)}_{\text{GBC}} + \underbrace{V(1-L(q)-\alpha)}_{\text{insert } l^{i*}} \end{aligned} \tag{1.9}$$

This expression shows us how each agents utility will be in optimum given some q and some preferences $V(\cdot)$. Now in an election individuals recognize that they should not simply vote for the q that optimizes their utility given static labor supply, instead they should vote for the q that yields the highest utility after making changes to ones labor supply. The policy which satisfies

$$q(\alpha^i) = \arg\max_q W(q,\alpha^i) \tag{1.10}$$

is called the *bliss point* of agent i, as this is the ideal choice of policy for this agent. In general this will depend on α^i (imagine someone with $\alpha^i = 1$, they have no change of earning labor income and might as well prefer a very high q).

1.2 Majority rule voting

Now knowing how each individual would prefer the tax rate to be set we need a way to aggregate individual wishes into a single global tax rate which applies to everyone. One option is to implement pure majority voting, that is a system with

1. Direct democracy, whatever citizens vote for will be the outcome.
2. Sincere voting, agents vote for the policy that is their bliss point (no strategic voting).
3. Open agenda voting, for all possible pairs q_1, q_2 citizens vote for their prefered option untill all combinations have been tried against each other.

This naturally is not exactly how democracies work, but it serves as a useful framework for modelling the dynamics at play in voting systems.

Arrows imposibility theorem Arrows impossibility theorem states that there is three or more choices for voters to choose from, no ranked voting system (i.e. majority voting) can convert individual votes into a community choice which fulfils

1. Unrestricted domain: all preferences are allowable, there is no requirement for consistency in policy preferences.
2. non-dictatorship: Voters actually matter for the final choice (it is not simply a benevolent social planner who chooses the best option).
3. Pareto efficency: there is no pareto improvements to be made.
4. Independence of irrelevant alternatives: if some given policy is prefered, it should still be prefered in an election with only a subset (including q) of the options available.

Condorcet winners Without any restriction on the election it is possible to have condorcet cycles, which is essentially a set of preferences that is cyclic over some set of alternatives, so $a > b$, $b > c$ and $c > a$ would constitute a condorcet cycle.
A *condorcet winner* on the other hand is a policy q^* which can beat any alternative q in a pairwise vote of (q^*, q). It can be shown that under majority voting, with $q \in \mathbb{R}$ a condorcet winner exists if voters preferences are single peaked, that is if

$$q'' \leq q' \leq q(\alpha^i) \text{ or } q'' \geq q' \geq q(\alpha^i) \Rightarrow W(q'', \alpha^i) \leq W(q', \alpha^i) \tag{1.11}$$

This equation essentially states that for each individual the indirect utility is monotonically decreasing on both sides of $q(\alpha^i)$. This assumption is strong - it is the indirect utility that must behave nicely, not the direct utility. The condorcet winner will furthermore be equal to the median voters prefered policy q^m. To see this see that any $q' < q^m$ will have support from less than half the population and so q^m would win in a pairwise vote. This argument is then identical for $q' > q^m$.

Returning to the indirect utility derived in (1.9) we can show that this does not satisfy single peakedness as defined in (1.11) for sufficiently large $\frac{\partial^2}{\partial q \partial q} L(q)$. To see this note that a positive second derivative corresponds to an graph that is convex implying at least no equilibrium, or of this property is only piecewise that there are multiple equilibria. The second derivative is straight forward to derive as

$$\frac{\partial^2}{\partial q \partial q} = (1 + V'(\cdot))L''(q) + V''(\cdot)L'(q)^2 \tag{1.12}$$

By assumption $V' > 0$ and $V'' < 0$ so we loose single peakedness when the first term is larger than the second, or with a bit of rearranging:

$$L''(q) > \frac{V''(\cdot)}{1 + V'(\cdot)} L'(q)^2 \tag{1.13}$$

Single crossing preferences An alternative to the single peaked preferences assumption is to assume the *single crossing* property. This is different form single peakedness as it not only involves assumptions about the shape of individual agents utility functions, but makes assumptions about the distribution of voter types, specifically assume $\alpha^i \in \upsilon$ where υ is some set of voters. The single crossing property is then that if

$$\begin{aligned}
&(\alpha^i < \alpha^{i'} \text{ and } q > q') \text{ or } (\alpha^i > \alpha^{i'} \text{ and } q < q') \\
&\text{then } W(q, \alpha^i) \geq W(q', \alpha^i) \Rightarrow W(q, \alpha^{i'}) \geq W(q', \alpha^{i'})
\end{aligned} \tag{1.14}$$

That is, if we consider two agents and two possible policies such that a) the "stronger" agent prefers the highest tax or b) the "weaker" agent prefers the lower tax, then we can infer that a) the "weaker" agent also prefers the highest tax or b) The "stronger" agent also prefers the lower tax. In other words relatively more extreme individuals prefer policies that are also more extreme. Notice how this is an assumption on the distribution of preferences across α^i's while single-peakedness was an assumption about the individual preferences for a given α^i. When individuals satisfy the single crossing asusmption the median policy q^m will be the condorcet winner and equilibrium policy.

Proof that single crossing leads to q^m: Think of the median type α^m who naturally has prefered policy q^m. Any $q < q^m$ will be dismissed by α^m and any agents with $\alpha^i > \alpha^m$ giving a majority to q^m. Likewise for any $q > q^m$.

We can show that the expression in (1.9) does satisfy the single crossing property in (1.14). To see this notice that when adding and subtracting $(1-q)\alpha^i$ to (1.9) we get

$$\begin{aligned}
W(q, \alpha^{i'}) &= L(q) - V(1 - L(q) - \alpha) - (1-q)(\alpha^{i'} - \alpha^i) - (1-q)(\alpha^i - \alpha) \\
&= W(q, \alpha^i) - (1-q)(\alpha^{i'} - \alpha^i)
\end{aligned} \tag{1.15}$$

So consider a case where individual $\alpha^{i'}$ prefers q to q', i.e.

$$W(q, \alpha^{i'}) \geq W(q', \alpha^{i'}) \tag{1.16}$$

From the rewrite in (1.15) this directly implies

$$W(q, \alpha^i) - (1-q)(\alpha^{i'} - \alpha^i) \geq W(q', \alpha^i) - (1-q')(\alpha^{i'} - \alpha^i) \tag{1.17}$$

Rearanging we then have

$$W(q, \alpha^i) \geq W(q', \alpha^i) - (q' - q)(\alpha^{i'} - \alpha^i) \tag{1.18}$$

Now notice that if $(q' - q)(\alpha^{i'} - \alpha^i) \geq 0$ we have shown that $W(q, \alpha^{i'}) \geq W(q', \alpha^{i'})$ implies $W(q, \alpha^i) \geq W(q', \alpha^i)$. Since $(q' - q)(\alpha^{i'} - \alpha^i) \geq 0$ is logically identical to the requirements posed in the single crossing definition, and the implication is exactly what the single crossing property entails, we have shown that our model does have the single crossing property.

1.3 Reflections on the median voter theorem

Everything above relies on the policy being single dimensional, which is probably a poor fit for the real world. However we could consider this single dimension to not be as conrete as a tax rate, but rather a "left-right" leaning, in which case one could argue for some kind of single-dimensionality. Without a single dimensional policy, it is however not generally possible to find a condorcet winner.

2 Electoral competition I

In this lecture we will begin studying a model of electoral competition, that is a model for how politicians propose platforms in the competition to be elected. We consider two party competition where there is one policy objective (e.g. the size of the public sector) and politicians are opportunistic i.e. they have only one goal, which is getting elected. Politicians must commit to a proposal before the election and must provide this policy after being elected (this doesn't matter to much when their only goal is getting elected).

Model setup Consider a continuum of citizens with quasi-linear utility from consumption c_i and a public good g so

$$w^i = c^i + H(g) \tag{2.1}$$

Individuals are taxed and accordingly have a budget constraint of

$$c^i = (1-\tau)y^i \tag{2.2}$$

Here y^i is distributed according to $F(\cdot)$ so that $E[y^i] = y$, the median income is y^m so $F(y^m) = 1/2$ and we assume that $y^m < y$ implying a right skewed distribution of income.
The government earns taxes τ and uses these to provide the public good g so

$$g = \int \tau y^i f(y^i)\ dy^i = \tau y \tag{2.3}$$

The *indirect utility* can then be found by inserting the governments budget constraint in (2.1):

$$\begin{aligned} W^i(g) &= \left(1 - \frac{g}{y}\right) y^i + H(g) \\ &= (y-g)\frac{y_i}{y} + H(g) \end{aligned} \tag{2.4}$$

This equation shows the central tradeoff in the model, namely that individuals gain utility from increasing g but then face paying a higher tax reflected in the fact that higher g decreases the utility from private consumption. The loss of private consumption is more costly for individuals with a higher income.

Prefered policy We can solve for an agents preferred policy by maximizing the indirect utility w.r.t. g,

$$\frac{\partial W^i}{\partial g} = -\frac{y^i}{y} + H_g(g) \qquad (=0) \tag{2.5}$$

where H_g is the derivative of H w.r.t. g. Solving for g shows that the individuals optimal policy level is

$$g^i = H_g^{-1}\left(\frac{y^i}{y}\right) \tag{2.6}$$

By assumption H is strictly concave and since the first term in W^i is linear, we can infer that this too is strictly concave. This in turn tells us that *preferences are single peaked* in this model. By this logic we can also infer that H_g^{-1} is decreasing in y^i. So higher income individuals prefer less of the public good, intuitively this is because high income implies a large marginal cost of increased taxes, while the marginal benefit of g is identical across all individuals.

2.1 A benchmark for election performance

Before we study how elections affect the choice of g, let's first consider what level a benevolent social planner (i.e. a good dictator) would choose. The utilitarian welfare function is simply

$$SWF^U = \int_{y^i} W^i(g) f(y^i) \mathrm{y}^i \tag{2.7}$$

i.e. the aggregated indirect utility over all individuals. Inserting in this equation we can simplify the expression quite a bit

$$\begin{aligned} SWF^U &= \int_{y^i} \underbrace{\left((y-g)\frac{y^i}{y} + H(g) \right)}_{W^i(g)} f(y^i)\ dy^i \\ &= W(g) \end{aligned} \tag{2.8}$$

where $W(g)$ is the average indirect utility, or equivalently the indirect utility of the individual with $y^i = y$. The social planner accordingly would provide a level of g equal to

$$g^* = H_g^{-1}(1) \tag{2.9}$$

which is simply (2.6) evaluated for the average voter (we get the 1 from dividing y/y).

2.2 Downsian electoral competition

Now instead of a social planner let us consider a situation where there is held an election for the political position to choose g. There are two parties $P = A, B$ and the probability that candidate P wins the election we denote p_P. Candidates receive an "ego-rent" R from holding office and get 0 if they loose, consequentially they seek to maximize $p_P \cdot R$. The share of votes each candidate receive we denote π_P, and using this we can write p_A and p_B as

$$\begin{aligned} p_A &= Pr[\pi_A > 1/2 | g_A, g_B] \\ p_B &= 1 - p_A \end{aligned} \tag{2.10}$$

Once again we assume that voters vote sincerely. The proposals g_A, g_B are announced before the election.

Simple example Let us first assume that all voters have the same income so $y^i = y$ and $W^i(g) = W(g)$ for all i. In this case

$$p_A = \begin{cases} 1, & \text{if } W(g_A) > W(g_B) \\ 1/2, & \text{if } W(g_A) = W(g_B) \\ 0, & \text{if } W(g_A) < W(g_B) \end{cases} \tag{2.11}$$

and of course $p_B = 1 - p_A$. In this case there is an unique Nash equilibrium where $g_A = g_B = g^*$. To see this notice that g^* maximizes $W(g)$ by definition. Any deviation from this level by either candidate will therefore immediately set $p_P = 0$ which gives the politician utility $0 < \frac{1}{2}R$. In this case we can thus say that political competition induces politicians to propose good policies.

Adding variation in y^i Now let us assume that $y^i \sim F(y^i)$. Because voters preferences are single peaked we know from the median voter theorem that the median voters preferred policy will be a condorcet winner and have majority support. Therefore we now have

$$p_A = \begin{cases} 1, & \text{if } W^m(g_A) > W^m(g_B) \\ 1/2, & \text{if } W^m(g_A) = W^m(g_B) \\ 0, & \text{if } W^m(g_A) < W^m(g_B) \end{cases} \tag{2.12}$$

Where W^m is the indirect utility of the median voter. Following a similar argumentation as before the Nash equilibrium will therefore be for both politicians to propose $g_A = g_B = g^m$.
Now because of the skewness in the income distribution we have that $y^m/y < 1$. Because H_g^{-1} is decreasing this implies

$$g^m = H_g^{-1}(y^m/y) > H_g^{-1}(1) = g^* \tag{2.13}$$

which is to say with varying income politicians propose a higher level of g than what is optimal. This is because every voters vote counts equally so politicians can disregard the fact that taxation is extremely costly to those with the highest income. The social planner on the other hand takes this into account. This model gives us a testable prediction, which is that a more skewed income distribution should produce larger public sectors.

Voters vs. tax payers Notice that politicians only care about the part of their population that is eligible to vote. So the relevant measure of y^m should be calculated only within the voters. Notice that y should be calculated within the full population as this enters the equations through the taxes, which we assume you must pay regardless of your voting status. If some of the population is not eligible to vote we have to take this into account. One way to test the consequences of a more skewed distribution is thus to consider changes to voter eligibility laws which allow poor people to vote. These purely affect y^m while leaving y unaffected.

2.3 Beyond the median, Gerber and Lewis (2004)

The paper by Gerber and Lewis gather data from 2.8 million individuals votes from Los Angeles county in the 1992 general election. The data contains information on individuals voters choices in elections on all levels and includes votes on concrete policy proposals. From these data they estimate a county-level distribution of voters policy preferences. This lets the authors infer the policy position of the median candidate in each district.
The authors also measure politicians behavior in whatever chamber they are elected to. This allows them to estimate the position on the liberal-conservative axis of each politician (both elected in LA and not).

With these informations the authors can ask whether the position of the median voter in the local district affects the position of politicians. According to the median voter theorem politicians should vote in accordance with their local median voter, but it is obvious that peer and party effects can be competing explanations. The authors regress the preferred policy of LA politicians on the local median voters position as well as the median position of the own-party delegation in whatever chamber the politician has been elected to.

The authors also investigate the role of district heterogeneity by interacting the median voters preferred policy with the variance of voters positions within the district.

The authors find a significant role of median voters preferences *in districts with homogeneous voters*. In heterogeneous districts this effect doesn't seem to exist to nearly the same degree. The authors also find evidence from peer effects in political stances.

3 Electoral competition II

So far we've studied a version of the downsian model with full policy convergence which implies that the median voter theorem also applies to representative voting democracies. This has revealed an important mechanism which forces politicians to run on popular platforms, often called the *affect* mechanism.

There are however good arguments against the downsian model. First of it's conclusions are mainly based on the discontinuity in probability of winning, disregarding any uncertainty about voters preferences. Additionally the model must obviously be incomplete, as other things than the preferences of the median voter surely must matter to some degree.

3.1 Probabilistic voting

We will now relax the assumption about perfect information about voters preferences. We do this by introducing a candidate specific trait (e.g. ideology) which voters have preferences about. Importantly we assume uncertainty on the distribution of voter preferences with regards to this dimension.

We depart from exactly the same point as last time, so the modelling framework is identical. This time however we assume there are three population groups $J = R, M, P$. Within groups income is identical at y^J and they therefore also have the same indirect utility function W^J. Each groups share of the total population is given by $\alpha^J = \alpha^R, \alpha^M, \alpha^P$, and we assume that $y^R > y^M > y^P$. The average income in the population is

$$y = \sum_J \alpha^J y^J \tag{3.1}$$

The voters care about g as well as ideology, captured by the two bias-parameters σ^{iJ} and δ which capture bias in direction of candidate B (w.l.o.g.). Thus voters in group J prefer candidate A iff

$$W^J(g_A) > W^J(g_B) + \sigma^{iJ} + \delta \tag{3.2}$$

We assume that

$$\sigma^{iJ} \sim U\left(\frac{-1}{2\phi^J}, \frac{1}{2\phi^J}\right) \qquad \delta \sim U\left(\frac{-1}{2\psi}, \frac{1}{2\psi}\right) \tag{3.3}$$

The bounds of the distributions are simply chosen to simplify calculations. The important parts are that a) bias parameters are equally likely to be positive or negative, so the bias can be both in favor and disfavor of candidate B with equal probability. And b) a higher ϕ^J (or ψ) implies more moderate voters, in the sense their votes are less likely to be tilted by ideology.
In this setup σ^{iJ} measures varying degrees of ideological focus within each income group, while δ measures an aggregate bias across the whole population, e.g. from scandals, campaigning etc.

Probability of winning The parties want to maximize the probability of winning. Notice first that within each group J we can identify the swing voter by the σ^{iJ} that solves the equation in (3.2) with equality, that is the swing voter will have

$$\sigma^J \equiv W^J(g_A) - W^J(g_B) - \delta \tag{3.4}$$

As all voters in J with $\sigma^{iJ} < \sigma^J$ will vote for candidate A. Since σ^{iJ} is uniformly distributed we can calculate the share that votes for candidate A as[1]

$$\begin{aligned} F^J(\sigma^J) &= \frac{\sigma^J + (2\phi^J)^{-1}}{2(2\phi^J)^{-1}} \\ &= \phi^J\left(\sigma^J + \frac{1}{2\phi^J}\right) \\ &= \phi^J\sigma^J + \frac{1}{2} \end{aligned} \tag{3.5}$$

Aggregating this vote share over all three groups we then find

$$\begin{aligned} \pi_A &= \sum_J \alpha^J\left(\phi^J\sigma^J + \frac{1}{2}\right) \\ &= \sum_J \alpha^J\left(\phi^J(W^J(g_A) - W^J(g_B) - \delta) + \frac{1}{2}\right) \end{aligned} \tag{3.6}$$

This expression is continuous and we have thus been able to alleviate the discontinuous nature of the simple model by introducing the stochastic preferences for ideology. Now candidate A wins whenever

[1] Recall in the uniform distribution on $[a, b]$, $F(x) = \frac{x-a}{b-a}$

$\pi_A \geq 1/2$, which happens with probability p_A. This probability is w.r.t δ which is the final stochastic term we haven't done anything with yet. We can write

$$\begin{aligned}
p_A &= Pr[\pi^A \geq 1/2] \\
&= Pr\left[\sum_J \alpha^J \left(\phi^J(W^J(g_A) - W^J(g_B) - \delta) + \frac{1}{2}\right) \geq 1/2\right] \\
&= Pr\left[\sum_J \alpha^J \phi^J \left(W^J(g_A) - W^J(g_B)\right) \geq \sum_J \alpha^J \phi^J \delta\right] \\
&= Pr\left[\delta \leq \frac{1}{\sum_J \alpha^J \phi^J} \sum_J \alpha^J \phi^J \left(W^J(g_A) - W^J(g_B)\right)\right]
\end{aligned} \tag{3.7}$$

where the third step can be reached by writing out the sum in additive parts, and seeing that $\sum_J \alpha^J \frac{1}{2} = \frac{1}{2}$. Now using the distribution of δ we have that

$$\begin{aligned}
p_A &= \psi\left[\frac{1}{\sum_J \alpha^J \phi^J} \sum_J \alpha^J \phi^J \left(W^J(g_A) - W^J(g_B)\right) - \left(-\frac{1}{2\psi}\right)\right] \\
&= \frac{1}{2} + \frac{\psi}{\phi} \sum_J \alpha^J \phi^J \left(W^J(g_A) - W^J(g_B)\right)
\end{aligned} \tag{3.8}$$

where $\phi \equiv \sum_J \alpha^J \phi^J$. Once again here we are reaffirmed that this model does not feature the discontinuous jump we had in the simple model.

Politicians proposed *g* Having derived the probability of winning we can now find a Nash equilibrium between the two politicians. Assume like before that each of them gets a rent from holding office R, so the best response function of party A is to maximize $p_A \cdot R$, i.e.

$$\max_{g_A} p_A \cdot R \tag{3.9}$$

which has first order condition

$$\frac{\psi}{\phi}\left(\sum_J \alpha^J \phi^J W_g^J(g_A)\right) \cdot R = 0 \tag{3.10}$$

Clearly this expression is only 0 when $\sum_J \alpha^J \phi^J W_g^J(g_A) = 0$. This expression represents the best response of A given some fixed g_B. In a parallel way we can solve the best response of party B given any g_A. This is the solution to

$$\max_{g_B} (1 - p_A) \cdot R \tag{3.11}$$

Taking the derivative of this w.r.t g_B will quite easily yield a similar expression to the one for g_A, namely $\sum_J \alpha^J \phi^J W_g^J(g_B) = 0$. Now since these two equations are symmetric the solutions must be as well, so we can conclude that $g_A = g_B = g^S$ where g^S is simply shorthand for the symmetric equilibrium policy.

The problem politicians solve is essentially maximizing a weighted social welfare function, where both population shares α^J and ideological parameters ϕ^J are important. In particular a higher ϕ^J implies less ideological dispersion in group J and this in turn implies a higher weight to this groups preferences W^J. The intuition in this is that for highly ideological groups (low ϕ) changes in g are less important for tipping votes, meaning it requires large changes in proposed g to gain additional votes in these groups. If a group is moderate (high ϕ^J) g is important in determining what they vote and many voters will flip if g is modified.

Equilibrium policies In equilibrium the implemented g will be one which solves

$$\sum_J \alpha^J \phi^J W_g^J(g^S) = 0 \tag{3.12}$$

Recalling the definition of W^J from (2.4) we have that $W_g^J = -\frac{y^J}{y} + H_g(g)$ which we can insert in the FOC above to get

$$\begin{aligned} \sum_J \alpha^J \phi^J \left(-\frac{y^J}{y} + H_g(g^S)\right) = 0 \qquad \Leftrightarrow \\ \sum_J \alpha^J \phi^J H_g(g^S) = \frac{1}{y}\sum_J \alpha^J \phi^J y^J \end{aligned} \tag{3.13}$$

using once again that $\phi \equiv \sum_J \alpha^J \phi^J$ and defining $\tilde{y} = \frac{1}{\phi}\sum_J \alpha^J \phi^J y^J$ which is essentially a weighted average of group incomes, we can rearrange to derive

$$H_g(g^S) = \frac{\tilde{y}}{y} \tag{3.14}$$

as a characteristic of the equilibrium, from which we can directly derive the equilibrium strategy as being

$$g^S = H_g^{-1}(\frac{\tilde{y}}{y}) \tag{3.15}$$

This is immediately similar to the result we derived in the simple case. Now notice that if either α^J or ϕ^J increases $\tilde{y}$ will move closer to y^J, meaning politicians will propose policies that are more favorable to that group of voters.

In this model we have a single policy dimension and single peaked preferences, so the median voter theorem tells us that a condorcet winner exists, and this is equal to the median voters preferred policy $g^m \neq g^S$. The inequality might seem odd, but recall that the median voter theorem addresses the existence of a condorcet winner, it does not state that this will be the implemented policy.

Reflections on testing model predictions The central prediction of this new model is that politicians align their policies with large and moderate voter groups. One way to get good variation in the population groups is to study changes in voter eligibility legislation which at an instant changes the relative sizes of voting groups, although one could argue that changes in voters legislation is endogenous to the political process. To get variations in the voter heterogeneity one could consider cases where voting district boundaries are redrawn.

3.2 On the vote-purchasing behavior of incumbent governments, Dahlberg and Johansson (2002)

The central question posed in Dahlberg and Johansson (2002) is whether incumbent government use their position to increase spending in districts with many swing voters as the probabilistic voting model suggests (or in districts with many party supporters, as suggested by alternative theories). They find that the number of swing voters increase the likelihood of getting the grant; evidence that incumbent politicians does use grants to target swing voters.
To show this they use a special Swedish grant administered by the central government to municipalities in 1997. The nature of the grant is suitable to study their question because, unlike most other government funding, there are no clearly stated purpose of the funds (except for furthering "ecological" development, which was a quite new idea in 1997). The grant was awarded close to an election further increasing the incentive to misuse the grant.

The authors regress a binary variable for a municipality getting the grant on two variables indicating there are many swing voters in a municipality: a) an estimate of the density of "cutpoint voters" (voters close to σ^J) and b) the vote difference between blocs in the previous election (small distance⇒close election). The authors find evidence that many swing voters does increase the probability of getting the grant, while the number of core voters does not. These findings are in line with the probabilistic voting model.

Another piece of evidence is from Strömberg (2004) which investigates the distribution of New Deal relief funds throughout the USA in the period after the great depression. Strömberg finds evidence that the number of radio listeners in a county increased the amount of funds received while controlling for county poverty and unemployment. This indicates politicians target "well informed" counties, suggesting this too is an example of spending targeted at areas with non-ideological voters (the reasoning is that radio access gives people information about g, making it more important than σ).

4 Redistributive politics I

In this and the next lecture we will shift to another important question, namely what determines the level of redistribution in society, and why this differs so much across time and countries. In this section we discuss the Meltzer-Richard model. This is a classical economists model which assumes rational voters with their own utility in mind when they vote. Naturally this approach yields a conclusion along the lines of an income-based theory for preferences to redistribution in which the relative income determines ones preferences for redistribution. Combining this with the median voter theorem, the difference between mean and median income becomes important for determining the level of redistribution.

Model setup The model setup follows the one in section 1, voters occupy a continuum and each has utility $w^i = c^i + V(x^i)$ subject to the budget constraint $c^i = (1-\tau)l^i + f$ and the time constraint $1 + e^i = l^i + x^i$[2]. $e^i \sim F(\cdot)$ with mean e and median e^m. Just like in section 1 we can solve the voters utility maximization by inserting the constraints in the utility and maximizing w.r.t l^i.

Once again we can derive a FOC of $l^i = 1 + e^i - V_x^{-1}(1-\tau)$ and we can define the average labor supply of the average individual as

$$L(\tau) = 1 + e - V_x^{-1}(1-\tau) \tag{4.1}$$

implying the FOC can be rewritten $l^i = L(\tau) + (e^i - e)$. The government follows a budget constraint $f = \tau l = \tau L(\tau)$ and the indirect utility is then given by

$$W^i(\tau) = L(\tau) + (1-\tau)(e^i - e) + V(1 - L(\tau) + e) \tag{4.2}$$

Taking the derivative of this and setting equal to 0 we can then solve for τ to derive the individuals bliss point τ^i, first take the derivative

$$\frac{\partial W^i}{\partial \tau} = L_\tau(\tau) - (e^i - e) - V'(1 - L(\tau) + e)L_\tau(\tau) \tag{4.3}$$

From the FOC we have that in equilibrium $V'(1 - L(\tau) + e) = 1 - \tau$. Using this and rearranging gives us the bliss point

$$\tau^i = \frac{e - e^i}{-L_\tau(\tau)} \tag{4.4}$$

Since $L_\tau(\tau) < 0$ this expression is decreasing in e^i. For all individuals with $e^i > e$ therefore prefer negative taxes while those with $e^i < e$ prefers positive taxes.

[2] Here we add e^i, in section 1 we subtracted α^i so the sign is flipped.

Downsian voting As we saw in section 1 this model satisfies the single crossing property, so pure majority voting will ensure that the condorcet winner is chosen, and this will be the preferred tax rate of the median voter $\tau^m = \frac{e-e^m}{-L_\tau(\tau)}$. This expression shows the main conclusion; more inequality increases the difference between e and e^m yielding a higher tax rate. Notice this is not the same as saying income inequality causes more redistribution, if for example the income distribution skews because the middle class gains income, this raises e along with e^m. Oppositely if the rich get extremely rich, this changes e^m without affecting e a lot. Also note that the denominator $|L_\tau(\tau)|$ is the change in labor supply as a consequence of taxation, i.e. the deadweight loss of taxation. A higher cost of taxation reduces the desire for redistribution because individuals forego personal income in a tradeoff with the size of the public transfer.

Also take note that like before e^m is the median productivity of the average voter while e is the average income of the average *tax payer*. As a consequence extending voting rights to poorer citizens will decrease e^m but not affect e, resulting in more redistribution.

4.1 Empirical results

A first consideration is that our model assumes that higher income leads to lower desired taxes. However when regressing preferences for distribution on log income we observe a significant but not very predictive relationship, showing that many other variables also matter. With this in mind let us consider some empirical evidence investigating the conclusions from our model. Furthermore the model really compares the amount of redistribution to the would-be income distribution if a nation had no government at all. This is of course very difficult to observe meaning researchers have to come up with alternative measures of the "baseline inequality".

In general the evidence for a relation between the skewness of a nations income distribution and the amount of government redistribution is mixed. Some find a positive correlation while others dont. Aidt et al. (2006) investigates the consequence of "franchice-extentions" (e.g. broader voting rights for the poor/women) in 12 european countries and find that broadening the voter base to include poorer voters does induce an increase in government spending. This finding is consistent with our model. However their result suggest that the added spending is mainly directed towards infrastructure and internal security, which is not obviously increase redistribution, as our model would suggest.
Alesina et al. (2001) investigates perhaps the largest puzzle in the voting-spending question, namely how comes the european countries have build large welfare states, while the US remains a relatively slim state with limited spending, especially on redistributive efforts. The authors suggests explanations can be put in one of three bins: "economic", "political" and "behavioral" explanations.

Clearly the simple model is not able to explain the differences between Europe and the US, as the income distribution is more skewed in the US, which should suggest they had a higher level of redistribution. One extension considered by Alesina et al. is that income dynamics means y^i is not static, but can change over the life cycle. If the median voter expects y^i to increase in the future, this would dampen the demand for redistribution, as it 1) might reduce the possibility to climb the income ladder and 2) will be costly once the median voter earns a higher income. However there is little empirical evidence to support the idea that income is more upwards mobile in the US than in Europe. Americans do however believe more in the idea of upwards mobility than europeans do.

Another way to modify the model is to introduce altruistic agents. By assuming europeans care more for the well being of the poor than europeans one gets a straight forward explanation for the observed differences in redistribution. Alesina et al. give two explanations for why this might be the case. First of the USA is more racially fractionalized than Europe, and it is well established within psychology that people identify with groups, and that race is an easily visible marker for "group" membership.

Especially considering that minorities are over-represented among the poor in the USA, race might have become a group marker, justifying lower redistribution (although this would probably require that the politicians were primarily elected by whites, as otherwise minority votes should affect the policy in equal amount).
Another behavioral explanation is that americans and europeans differ in their beliefs about welfare recipients, so that americans are more likely to believe that social welfare recipients are lazy. There is some evidence that this is the case from survey questions.

5 Redistributive politics II

In the previous section we studies the simple Meltzer-Richard model predicting that a more skewed income distribution (where the mean was measured among the voters) would result in more government redistribution. This model has a major issue, namely that it's conclusions are exactly opposite of what we observe when comparing the US and Europe. The paper by Alesina et al. debated solutions to this "paradox". They propose that income mobility or perceptions about it can shape voters attitudes towards redistribution, that racial divides and "group thinking" might be to blame for low redistribution in the US, or that beliefs about the causes of poverty were central in shaping voters preferences for redistribution.

In this section we will focus on individual level preferences for redistribution. In particular we will attempt to answer three questions related to this, each studies by a separate paper

- Are people even aware of the degree of inequality and redistribution in society? (Gimpelson and Treisman, 2018)
- Are peoples preferences towards redistribution affected by their knowledge about the current level of inequality and redistribution? (Kuziemko et al., 2015)
- Do beliefs about social mobility play into these questions? (Alesina et al., 2001)

5.1 Misperceiving inequality, (Gimpelson and Treisman, 2018)

In the Meltzer-Richard model we assume that individuals are aware of the distribution of incomes in society, and that they know exactly their own position in the distribution. This entails a conclusion that individuals care about their relative position in the income distribution $e - e^i$ when determining what level of redistribution they would prefer. The paper by Gimpelson and Treisman asks to what degree individuals are even aware of their own place in the income distribution. They find that ordinary people generally does not know how they fit in the distribution, suggesting that theories that relies on some relation between income inequality and politics fails at a very basic level. When using individuals perceived income inequality instead of actual inequality, there is a clear relation between income (perceived) and the demand for redistribution.

Gimpelson and Treisman use survey dataset from several countries to elicit individuals knowledge of the income distribution in their country. They find that in almost all countries resident guesses the average income quite wrong. In most countries the average guessed mean income was even on the wrong side of the median income. Furthermore peoples guesses about the distributional shape are very varying, with the most common answer getting less than 50% of total answers in 29 of 40 countries.

Plotting actual and perceived GINI index against each other shows little or no correlation. The authors also asks respondents if they believe income inequality has increased or decreased over the past 5 years, and find that regardless of the actual development most people guess that inequality has increased. (This result is somewhat questionable, first of people rarely think about the economy in 5-year periods, and the distinction between income and wealth becomes blurry when asking ordinary people)

Life in transition survey The authors main point, that knowledge about ones position in the income distribution is very limited is however quite strong, as it seems people at both the top and bottom of the income distribution tend to estimate their position closer to the center than they are. In LiTS respondents are asked if they believe government should redistribute income between people, a measure which the authors regress using their "perceived position" measure while controlling for actual country GINI. They find that perceived GINI is significant in explaining respondents attitude towards redistribution, implying that a perceived higher inequality is associated with an increased interest in government redistribution.

5.2 How elastic are preferences for redistribution? (Kuziemko et al., 2015)

The next paper by Kuziemko et al. digs deeper into the *causal* link between perceived inequality and preferences for redistribution. Their main point is that as the US has become more unequal due to income concentration at the top of the income rank, one would from the Meltzer-Richard model expect the demand for redistribution to increase.

Opposite to the expectation Kuziemko et al. note that top income taxes have actually been decreasing, and they find no increased demand for redistribution in survey questions. To explain this puzzle authors propose three explanations

- Americans might not care about rising inequality.
- Americans might not know that inequality is rising.
- Americans dont believe the government can effectively redistribute income.

The central question for the authors is then to understand how knowledge about US income inequality and policies to change this affect peoples views. To do this the authors set up a randomized control trial experiment using Amazons Mechanical Turk. Using 4000 respondents they randomly assign either a treatment of interactive personalized information about US income inequality etc, while the control group receive no such information. Both groups are then asked to complete a questionnaire on their views on inequality, redistribution and their general view on government. The authors also conduct a followup survey with about 6000 respondents to analyze mechanisms behind the first results.
Their results show a strong effect from treatment on attitudes towards inequality (treatment individual perceive it as a more serious issue), but at the same time only a weak effect in favor of inequality reducing policies (except for the estate tax for which the effect is quite large). The treatment reduces participants trust in government but does not alter their voting intent in the coming election. The authors suspect two competing effects, 1) treatment increases concern about income inequality but also 2) reduces trust in government. To study these two competing effects the authors run a second experiment in which the treatment forces people to reflect on aspects of government that they dislike. This treatment reduces trust in government, while leaving views on income inequality unchanged. The authors also show that the reduces trust in government directly reduces support for government transfers to the poor.

In conclusion the authors find that more information can increase concern for an issue, but that there are complicated counteracting effects from learning about inequality. While the RCT is well carried out, the use of AmTurk most likely induces heavy skew in the sampled population towards low income or unemployed individuals, implying external validity might be low.

5.3 Causal Inference I

As a slight deviation from the main topic of political economics we will also study basic causal inference theory beginning with the potential outcomes framework. Consider a situation where individuals either receive treatment in which case their outcome is described by Y_{1i} or no treatment, resulting in an outcome of Y_{0i}. These variables describe potential outcomes, but naturally only one of the variables are observed. Let D_i be an indicator variable for being in the treatment group, the observed outcome is then

$$Y_i = \begin{cases} Y_{1i} \text{ if } D_i = 1 \\ Y_{0i} \text{ if } D_i = 0 \end{cases} = Y_{0i} - D_i(Y_{1i} - Y_{0i}) \tag{5.1}$$

We can never learn about this expression by studying a single unit, that is we can never observe the individual causal effect directly, instead we need to estimate the effect by comparing average observed outcomes for the treated with average observed outcomes for the untreated. In particular let us consider the expression

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{\text{observed difference}} = \underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{ATET} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\text{Selection bias}} \tag{5.2}$$

Here the first term measures the ATET $E[Y_{1i} - Y_{0i}|D_i = 1]$ which measures the causal effect of treatment on those who are actually receiving treatment. Note this is not the ATE $E[Y_{1i} - Y_{0i}]$.
The second term is the selection bias, which measures the difference in average in baseline outcome between the two groups. If we assume this selection bias is 0, that is groups are expected on average to fare equally well without treatment, OLS can estimate the ATET. Assuming that we achieve a selection bias of 0 further implies that the ATE is equal to the ATET.

6 Redistributive politics III

This lecture goes into the paper by Alesina et al. (2018). This paper studies the effect of perceived income mobility on attitudes towards redistribution. We have already seen evidence that americans are more optimistic about income inequality than europeans, but whether this optimism causally influences attitudes politics is unclear. The authors collect survey data from France, Italy, Sweden, UK and the US to get a cross country picture of the relation between perceived income mobility and demand for redistribution. Furthermore the authors add an element of randomization to the questionnaires to manipulate the respondents beliefs about income mobility. They also ask for respondent political beliefs allowing them to estimate any heterogeneity across the political spectrum.

Unlike Alesina et al. (2001) which focuses on the intragenerational mobility, Alesina et al. (2018) focus on intergenerational mobility. The authors use data on parents incomes matched to data on childrens income when adult to calculate transition probabilities towards each quintile for children with parents in the bottom quintile. (I.e. what is the probability of being born in the bottom five and ending up in the fourth quintile).

In the survey questions respondents are then asked to guess these probabilities. Comparing the actual numbers to the average questionaire answers. This shows that US citizens underestimate the probability of remaining in the bottom quartile while overestimating the probability of going from Q1 to Q5. In Europe respondents guess exactly opposite of this.
Correlating peoples guess of Q1 to Q5 income mobility with peoples preference for various redistributive policies show that believing more people make the Q1 to Q5 transition imply lower support for redistribution. To understand if this relation is causal the authors use their RCT setup in which some of the respondents are shown videos about income mobility designed to make them more pessimistic towards the level of mobility. They find the treatment significantly alter policy preferences for left-wing voters, while right wing voters change their perception of the issue but not their policy views. One way to interpret these results is that right wing voters have lower trust in the government, which is reinforced by the treatment videos, making them distrust the effectiveness of policy.

7 Incumbency advantage

The lecture beings by a rather basic observation that around 90% of US elections at most levels of government are won by the incumbent party.

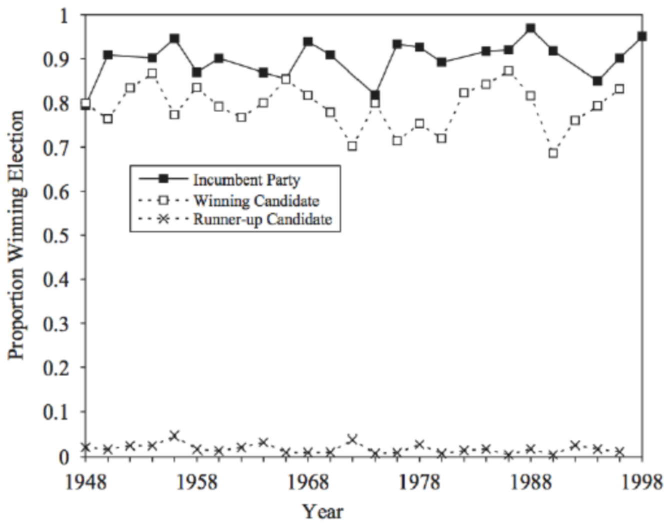

Figure 1: Incumbency advantage The figure shows the share of election won by candidates of incumbent and runner up parties. From Lee (2008)

The challenge with answering if incumbency causally affects the probability of winning is that there is no good random variation to use. After all randomly changing around which politicians are incumbent would be infeasible. Instead one can use a regression discontinuity design to extract the causal component. In particular we might worry that some districts simply lean either Republican or Democrat and this is what causes the large gap in election probabilities for incumbents, not the incumbency itself.

Lee does exactly this. They set up a RD design using data from elections for the US house of representatives from 1946 to 1998. In each distict they then calculate the democratic vote margin V in election at time t as a running variable (if this is above 0 democrats win, if it is below republicans win). Using this they create a "treatment" of $D = (V > 0.5)$ i.e. treatment is a democrat being incumbent in the period $t+1$ election. As outcome they have outcomes of the period $t+1$ election.

They also gather a long list of covariates X which are determined prior to the period t election (i.e. from year $t-1$ or earlier).

The authors find that the true causal effect of incumbency is somewhere between a 5 and 10% increase in the vote share and a corresponding 36% increase in probability of winning the election. Notice however that this setup defines winning at a party level, so while having a democratic incumbent increases the probability of a democrat being elected, it is uncertain if this effect also exists at the individual level.

To study if individual incumbency or party incumbency is the driving force is a different and harder question to answer, since there is no simple cutoff determining if a politician decides to run again, or attempts to pass the position on to another party member.

7.1 Regression discontinuity

RD designs are based on the existence of some running variable V with an observed threshold v_0 above which the treatment (i.e. winning the previous election) is given. The basic idea is then that around the threshold $[v_0 - \epsilon, v_0 + \epsilon]$ it is essentially random which party ends up on either side, and we can get an estimate of the causal effect of incumbency. This can be done either by running the regression only on observations where $V \in [v_0 - \epsilon, v_0 + \epsilon]$ or by adding a function $g(V)$ to the regression which is flexible enough to capture any nonlinearities outside the interval. Note that the RD design doesn't estimate the full ATE, but only ATE* which is a weighted ATE with high weight on observations close to v_0.

One might suspect that even close to v_0 parties have influence on who wins, but as long as there is some element of randomness in who actually wins the procedure will work. Given the RD estimates the correct result we can also infer that any predetermined characteristics of the running parties should be continuous around v_0, in other words there should not be any discontinuities in other variables X which were determined before the election. Checking if this is the case is essentially a test of random assignment around the threshold.

8 Affecting vs. electing policies I

So far we've seen two main models for political competition. The Downsian model which takes its starting point in a simple utility based model of consumption of private and public goods(/policy) and assumes politicians goals are to be elected. In this model voters know what policies politicians propose and only care about policy. Politicians furthermore know what policy voters would prefer. The Downsian element is the way politicians compete by proposing policy to be elected. Because of the perfect information the model has a large discontinuity in probability of election whenever a politician deviates from the equilibrium, which is to propose the preferred policy of the median voter.
In other words in the downsian model which politician is elected doesn't matter, as both politicians propose the same policy, which is the median voters preferred policy.
The second model we've seen is the probabilistic voting model in which voters not only care for policy but also have intrinsic preferences for either candidate. In this model politicians also fully converge, but not at the median voters prefered policy, but rather at a policy that weights the optimal policy of voter groups with their "importance" in being elected. (that is very ideological groups are difficult to persuade with policy, so they get little weight when setting the politicians proposal).

Both of these models show full policy convergence *before* the election takes place. This does not mean the election is not important, but does mean that the primary purpose of an election is not to choose the best politician, but to influence all running politicians before the election takes place.
Much of these results is driven by the assumption that politicians only goal is to be elected. Naturally when this is the case they dont have any incentives not to fully optimize their chance of being elected, but if instead politicians have ideological or partisan goals, they might be willing to risk loosing if the alternative is to big of a political compromise.

8.1 Model setup

Once again we begin with a version of the simple Downsian setup. Consider a continuum of voters indexed by i with incomes y^i. The elected politician will implement policy g and votes get different indirect utilities from this depending on their income $W^i = W^i(g; y^i)$. Voters preferred g^i is decreasing in y^i.
Whats new this time is that candidates now only care about policy, and give no weight to actually winning. Assume for this purpose that two candidates R, L are drawn from the pool of voters. W.l.o.g. we assume that $y^L < y^m < y^R$ meaning R has high income and prefers a low amount of public spending g^R etc. so $g_L^* > g^m > g_R^*$. Just like in the standard downsian model voters choose their preferred candidate to vote for and it is the median voter who is pivotal so the probability that L wins is

$$p_L = \begin{cases} 0 & \text{if } W^m(g_L) < W^m(g_R) \\ \frac{1}{2} & \text{if } W^m(g_L) = W^m(g_R) \\ 1 & \text{if } W^m(g_L) > W^m(g_R) \end{cases} \tag{8.1}$$

In this scenario we can then solve for a Nash equilibrium between the candidates, the expected utility for candidate L is

$$E[W^L(g)] = p_L W^L(g_L) + (1 - p_L) W^L(g_R) \tag{8.2}$$

from which we see the core tradeoff that is moving g_L towards g_L^* increases the benefit of being elected, but lowers the probability of being elected p_L. Similarly

$$E[W^R(g)] = (1 - p_L)W^R(g_R) + p_L W^R(g_L) \tag{8.3}$$

Now once again because of the discontinuous nature of p_L, if R is playing $g_R = g^m$ either L also plays g^m and gets $E[W^L(g)] = W^L(g^m)$, if L deviates towards g_L^* then immediately $p_L = 0$ and L is worse off. Naturally the same is true for deviating in direction of g_R^*. Thus in the Nash equilibrium both politicians propose $g_L = g_R = g^m$ and we again have full policy convergence like in the simple downsian model. The driving effects in this conclusion is the balance between the discontinuously changing probabilities and the continuously changing benefit of policy. In summary this model captures an interesting, probably realistic mechanism by which politicians are pulled towards the center simply by balancing their personal gains of winning office with their probability of getting elected. However the full convergence is a result driven by the knifeedge probability so an augmented model is likely to predict less than complete convergence.

(lack of) commitment One way to implement this augmentation is to remove politicians commitment to policy after they are elected. The only change from the above is that now politicians choose which policy g to implement after they're elected. Assuming this is a one-shot game, the proposed policy during election becomes unimportant as voters knows politicians are uncredible. Therefore the win-probabilities become

$$p_L = \begin{cases} 0 & \text{if } W^m(g_L^*) < W^m(g_R^*) \\ \frac{1}{2} & \text{if } W^m(g_L^*) = W^m(g_R^*) \\ 1 & \text{if } W^m(g_L^*) > W^m(g_R^*) \end{cases} \tag{8.4}$$

So the winning candidate will still be the one most aligned with the median voter, but whoever wins fully implement their desired policy g_L^* or g_R^*. This prediction is completely opposite to the previous one. In this case elections only alter policy outcomes through the ability to elect the most desired candidate.

In summary we've seen how two almost identical setups can predict both complete convergence (affect) and no convergence at all (elect). The probabilistic voting model is an example of an in-between of these two extremes we've already studied. Another example is a no-commitment repeated game model where previous policy affect the probability of being reelected.

9 Affecting vs electing policies II

In this lecture we will see a number of articles debating to what extend real politics operate after affect or elect mechanics. Essentially this is a question of whether it matters which party wins the election or not. To test this empirically we need to study the evolution of policies over election cycles. A major identification issue in this question is that the level of affecting and electing might vary across parties and time, including over the business cycle. One way to get around this is to use a RD design to compare close elections.

9.1 Do political parties matter? Evidence from US cities

The first paper estimates an RD model on election data from local elections in the US (Ferreira and Gyourko, 2007). They show the mayors political affiliation does not affect the size of government, the allocation of spending or on local crime rates. However they do find a significant incumbents advantage. For data they use information on mayoral election and policy outcomes from 1950 to 2005. Their sample covers all US cities with more than 50.000 citizens.

This first paper seems to suggest there is no real elect element in mayor election, but this does not rule out the affect mechanism. Furthermore one might argue mayorial elections are low-attention low-stakes elections compared to national and statewide elections.

9.2 Do voters affect or elect policies? Evidence from the US house

The paper by Lee et al. (2004) does almost the same as the paper by Ferreira and Gyourko, they set up a RD design but instead of mayoral elections they use data from the US House of Representatives. For outcomes they use the individual politicians ADA scores (a kind of "policy score"). In their dataset the RD shows a large discontinuity in voting behavior between democrats and nondemocrat politicians. This thus implies that the elect channel is very important in House elections. Furthermore using some clever econometrics they show that in fact the affect mechanism is of almost no importance, suggesting full policy divergence.

The way they do this is by noting that if affect is a mechanism at play we would expect an increase in democrats electoral strength to cause both parties to propose more left-wing policies. One way to think of this is that if many incumbents are democrats, they have an advantage and republicans have to make up this difference by moving their proposals for g towards g^m. So a democrat incumbent would result in more leftist policies after the election both by a) democrats being more likely to be reelected (the elect channel) and b) republicans proposing more leftist policy as well (the affect channel).
The first part can be estimated using a RD to estimate the increase in reelection probability and any change in period t voting behavior caused by incumbency. The authors can also estimate the total effect of by estimating the causal effect of incumbency on period $t+1$ voting behavior regardless of who gets elected. They can then residually calculate the affect component.

9.3 Do parties matter for economic outcomes? A regression discontinuity approach

The final paper for this lecture is by Pettersson-Lidbom (2008). Here the authors do pretty much the same as in Ferreira and Gyourko (2007), but using data for Swedish local government elections. The author finds a significant party effect on expenditures over total income, unemployment rates and the number of public servants. All effects go in the expected direction so for example left wing mayors have on average more public servants and higher expenditures than right wing mayors.

9.4 Summary

Of the three papers we've seen, two study the effect of partisanship on actual policy outcomes, while Lee et al. (2004) instead considers the effect of partisanship on voting behavior in parliament. Both of the two papers studying the effect on actual policy outcomes use local elections, Ferreira and Gyourko (2007) in the US and Pettersson-Lidbom (2008) in Sweden. In the US there seems to be no effect of partisanship (i.e. full policy convergence) while the Swedish data reveals large differences (little or no convergence). So at least in the US it is quite clear that national politics are driven primarily by the elect mechanism, while it is less certain what (if anything) matters for local elections. It would appear that the setting of elections matter a lot, but how or why is not clear.

10 Legislative bargaining

In this final lecture of part 1 of the course we study a part of the political process which we have so far left untouched, namely the politics that happen after an election has taken place. In a case with non-perfect affection of politicians, the elected will have positions and policies to bargain for, in favor of themselves or their supporters.

We still consider the question of how much governments should spend (g) but now we consider the bargaining game that takes place after politicians are elected. To give politicians different preferences for g we assume they vary in income. We can interpret this either as egoistic citizen candidates or as politicians representing different population groups which vary in y.

Specifically consider three politicians L, M, R and their incomes which are ranked $y^L < y^M < y^R$ and assume their preferences are given by $w^J = c^J + H(g)$. Each candidates preferred policy is then

$$g^J = H_g^{-1}(y^J/y) \tag{10.1}$$

implying $g^L > g^M > g^R$ because H_g^{-1} is decreasing in its argument.

One round bargaining We take as a given a specific protocol for bargaining, in this first case assume that nature selects a politician $a \in [L, M, R]$ to be agenda setter. This individual proposes a policy g_a as a take-it-or-leave-it offer. If two politicians agree on g_a it is accepted, but if only one (or 0) candidates support g_a some status quo policy $\bar{g}$ is implemented instead.
First assume $a = M$ and that in this case the proposer suggests $g_a = g^M$ and since M is the median voter in the legislature (not the population) this proposal is the Condorcet winner and passes. Why? Both of the two other politicians L, R cannot both at the same time support any other policy $\tilde{g}$ over g^M, as their ranking means one of them will always prefer g_M. Now knowing this it is obvious that M will play g^M to begin with.

While $a = M$ is trivial, next consider the case where $a = L$. Now the optimal decision for L depends on the value of $\bar{g}$.

- If $\bar{g} > g^L$ the proposer should suggest g^L which passes unanimously as everybody prefer this to the status quo.
- If $g^M \leq \bar{g} \leq g^L$ he should propose $\bar{g}$, this too passes unanimously as it is identical to the outside option.
- If $\bar{g} \leq g^M$ he should propose $\min[g^L, \tilde{g}^M]$ where $\tilde{g}^M$ is set such that $W^M(\tilde{g}^M) = W^M(\bar{g})$, that is $\tilde{g}^M$ is as bad as the outside option to the median politician, but no worse. This ensures M agrees to the policy.

Notice that in all cases M is part of the coalition, so while the median politician is not completely determining the outcome, the median politicians position but bounds on what proposals are possible to make by L or R. Furthermore the possibilities for L and R depend heavily on the value of $\bar{g}$.

Two round bargaining Now lets extend our reasoning to two round bargaining. In the first round everything goes as previously, except if g_a fails, a new random agenda setter $a_2 \neq a_1$ is selected, to propose a new policy g_2. Then if no agreement is reached in the second round the legislature implements $\bar{g}$. Otherwise g_2 will be implemented. The logic is the same as in the single round case, except the bounds of what can be offered to M is now not $W^M(\bar{g})$, but the *continuation value* for M from going on to a second round.
Say L is the proposer, this means L should propose $g \in [g^M, g^L]$ such that g is equal to the continuation value of M. This ensures the proposal passes with support from M. If the continuation value increases, this forces L to propose a g closer to g^M.

10.1 Multilateral bargaining (Baron and Ferejohn, 1989)

Consider a legislature with n single vote members, who each represents a district from which they were elected. The game considered by Baron and Ferejohn (1989) is then a kind of divide-the-dollar game where the legislature bargains to distribute a fixed amount of resources. This problem has $n-1$ dimensions, i.e. the number of districts among which to divide the resource, except for the last district in which the amount of resource is residually given. Thus no condorcet winner generally exists, and instead majority voting results in condorcet cycles. To give some structure the idea is therefore to alter the assumptions on the aggregation of preferences to something different than majority voting. In particular the proposed structure is a legislature with a set of rules

1. A recognition rule, which determines who in the legislature can propose policy.
2. An amendment rule, that sets up rules for making changes to the initial proposal.
3. A voting rule, which decides how voting is conducted and in particular what is required for a proposal to pass.
4. A deadline rule, setting the maximum number of bargaining rounds before an outside option is adopted.

A very simple set of such rules could be 1) equal probability of every legislator being selected as proposer 2) No amendments allowed 3) Majority rule and 4) Two sessions of bargaining. In this setup some random member P_1 is selected as proposer and proposes a distribution

$$x^{P_1} = (x_1^{P_1}, ..., x_n^{P_1}) \tag{10.2}$$

such that $\sum x_i \leq 1$. If a majority approves the proposal it becomes policy, otherwise a second round of bargaining is begin. If a majority still doesn't agree with the proposal the status quo $x^{SQ} = (0, ..., 0)$ is implemented.

We assume that a members payout of a result in round t is $\delta^t x_i$ where δ is some kind of common discount factor that measures the "cost of delay". This ensures some bias towards finishing the game early. The model can be solved by backwards induction. In the last period the continuation value

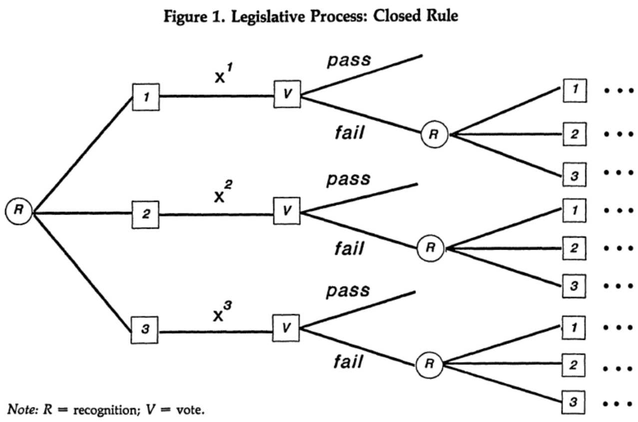

Figure 2: Figure 1 from Baron and Ferejohn (1989) showing the structure of the closed rule game.

for everyone is 0 so in this vote any voter will vote in favor of a distribution as long as it yields a small positive payoff. From the perspective of the proposer only $(n-1)/2$ votes are needed for the proposal to pass, so the proposer in the second round will propose some small payoff $\epsilon > 0$ to $(n-1)/2$ districts and payoff of 0 for the rest. This leaves the proposer with a value of ≈ 1 (technicalities about ϵ withstanding).

In period 1 all legislators know that there is an equal probability $1/n$ of begin selected as proposer in the second round of bargaining, and their continuation value is therefore δ/n. Anticipating this the proposer in period 1 will offer δ/n to exactly $(n-1)/2$ districts and 0 to the remaining. This leaves the payoff for the proposer $1-(n-1)\frac{\delta}{2n}$. Thus the equilibrium distribution is

$$x^{P_1} = (\underbrace{1-\frac{n-1}{2}\frac{\delta}{n}}_{P_1}, \underbrace{\frac{\delta}{n},\dots,\frac{\delta}{n}}_{(n-1)/2\text{ players}}, \underbrace{0,..,0}_{\text{remaining }(n-1)/2}) \tag{10.3}$$

which passes with a minimal majority. Notice that there are no veto players as the proposer can propose any of a whole set of equilibria in which different districts get 0. Furthermore we see quite a significant benefit from being awarded the initial proposer role. This is in part a result of the majority rule system, as the proposer can skim a benefit from the minority and in part because of the no-amendments rule which prevents modifying the proposal by P_1. Furthermore the closed rule model always have a solution in the first round meaning no delays.

10.2 Empirical evidence for the bargaining model (Ansolabehere et al., 2005)

The paper by Ansolabehere et al. tests aspect of the model by Baron and Ferejohn. Especially they test the theoretical observation that the proposer will gain a disproportionate gain from the bargaining. They do this in the setting of government formation. In this setting the prediction is that the formateur of the government should get a disproportionate amount of cabinet seats in the finished government.

An important part of the paper is the use of "minimum-integer voting weights" instead of raw seat shares in the respective parliament. This is essentially a mapping from the parliament seats to a measure of the number of possible coalitions a given party can be in to form a majority. In particular it is the minimum number of legislators each party could have while preserving the possible coalitions (without respect to the total size of parliament). The reason for this conversion is that while seat share and minimal-integer weight shares are correlated, the latter are a better measure of a political partys bargaining potential. Furthermore the theory is more accurately mirrored by minimal-integer voting weights than party seat shares. Notice the simple version of the Baron and Ferejohn does not include parties, but this extension is simple to implement, and predicts that parties with high voting weights get high bargaining power, that the formateur enjoys agenda power and overproportional placement in the government. Importantly alternative models does not provide the second implication, but instead predict proportional representation for all parties regardless of their role in the bargaining process.

TABLE 1 Voting Weight and Formateur Effects in the Allocation of Cabinet Posts in Parliamentary Governments, 1946–2001

	Unweighted (1)	PM Weighted (2)
Formateur (β_1)	.15* (.05)	.25* (.04)
Share of Voting Weight (β_2)	1.12* (.13)	.98* (.11)
Constant	.07* (.02)	.06* (.02)
R^2	.72	.82
# Observations	680	680

Dep. Var. = Share of Cabinet Posts
Clustered standard errors in parentheses, where each cluster is a country.
*statistically significant at the .01 level.

Figure 3: Key results from Ansolabehere et al. (2005)

The table shows results from regressing the number of cabinet posts awarded to each party on a dummy for being the formateur party as well as their share of voting weights. Their results are significant and suggest a formateur disproportionate advantage in accordance with the models predictions. Importantly studies that use seat shares instead of voting weights does not reach the same conclusion.

Part 2

11 Calculus of voting

So far we've been studying models with the implicit assumption that everybody votes build in. The first topic in the second half of the course will be to investigate why voters might not turn out to vote in reality. To motivate studying turnout one could argue a) that turnout in itself is required for a legitimate democracy (although one could argue that the affect mechanism means this is not necessarily the case), b) that elections institute a mechanism of information aggregation in which more voters implies better information or c) that elections serve to discipline politicians and a lack of turnout breaks this mechanism.

So far our models have primarily assumed two-candidate races with full turnout. The conclusions naturally hinge on the two-candidate assumption, but also that voters know which candidate they prefer and that voting is costless. To see how costly voting might affect turnout consider the choice for a single individual to vote for either candidate A or B. The utility from either candidate winning is U_A and U_B respectively. The probability that A wins depends on whether the individual votes or not. If the voter votes the probability of A winning is $P_{A|vote}$ and otherwise $P_{A|novote}$. We will assume that the voter never wants to vote for B. Also there is a cost c of voting. The expected utility when voting is therefore

$$V_{vote} = P_{A|vote}U_A + (1 - P_{A|vote})U_B - c \tag{11.1}$$

and when not voting

$$V_{novote} = P_{A|novote}U_A + (1 - P_{A|novote})U_B \tag{11.2}$$

Naturally the individual votes if

$$V_{vote} \geq V_{novote} \Rightarrow (P_{A|vote} - P_{A|novote})(U_A - U_B) \geq c \tag{11.3}$$

In other words if the change in probability of A winning from going to vote times the benefit from A winning $B \equiv U_A - U_B$ is greater than the costs, it is worth voting. Notice that $p \equiv P_{A|vote} - P_{A|novote}$ is both the change in probability that A wins and the probability that the voter is pivotal (think of the probabilities as discontinous depending on the state - if A is pivotal $P_{A|vote} = 1$ and $P_{A|novote} = 0$). Using the summarized terms we then have the inequality

$$pB \geq c \tag{11.4}$$

In itself this equation seems unlikely, it is almost never the case that one is pivotal, so the benefit must be very large, even with modest costs to drive voters to vote. The standard way to fix this is to add some independent benefits of voting D and instead postulate

$$pB + D \geq c \tag{11.5}$$

where D could be anything from bragging rights to a sense of civic duty. The only central restriction is that D should not depend on the outcome of the election.

11.1 Empirics on turnout (Gerber and Green, 2000)

A central question is what gets people out to vote. Gerber and Green conduct a large RCT in New Haven where they use canvassing, mail and phonecalls to try and persuade citizens to vote in the 1998 national election. They construct random messages to target different parts of the calculus of voting equation, including aluding to a civic duty (D), that some races are close (p) and neighborhood solidarity/benefits (B).

Their main finding is that personal contact has a larger effect than mail while telephone calls have no or maybe even a negative effect on turnout. They hypothesize on this basis that personal contact is important for turnout, and a reduction in canvassing can explain the reduced turnout in the US.

12 Elections as information aggregation

Consider a case where there are two policies 0 and 1 of which one must be chosen in an election. On top of this the world can also be in one of two states 0, 1 which occur as random with probability $\alpha = 1/2$. For notation we denote the chosen policy x and the realized state z. There are $N = 2n + 1$, $n \in \mathbb{N}$ voters with utility

$$U(x, z) = \begin{cases} -1 \text{ if } x \neq z \\ 0 \text{ if } x = z \end{cases} \tag{12.1}$$

To begin with assume voters *must* vote and that this is cost free.

This setup models a scenario in which voters does not per-se prefer any politician but only care about matching the policy outcome to the observed state of the world. For example we can think of the state as measuring the size of the dead weight loss of taxation, and people preferring a left wing politician when this is low, and vice versa.
Voters receive a signal $m_i \in [0, 1]$ which conditional on the true state z independent of other voters signals. With probability$r > \frac{1}{2}$ the signal is correct and $m = z$.

Prior to receiving the signal the rational expectation of voters is that $P(z = 1) = \alpha$ and after the signal $P(z = 1|m = 1) = P(z = 0|m = 0) = r$, so the signal is effective.
CONT. slide 8

13 Lecture 21 - Competence vs. representation

In this lecture we study the problem of selecting politicians who at the same time represent a voters preferred policies and carry the competencies to get these policies implemented. In the two extreme cases we might have either

1. All politicians are equally competent, but voters disagree on the optimal policy.

2. All voters(/politicians) agree on what is the optimal policy but politicians vary in their competence.

When considering some mix of these two extremes a tradeoff arises between competence and policy. In particular all voters wants to be represented by someone competent, but differ what policy attitudes they prefer.

13.1 Model by Mattozzi and Snowberg (2018)

The model is very similar to the regular downsian model. A continuum of citizens with measure 1 vary in their income and are taxed with some rate τ to provide a public good. Candidates for elections are drawn from the population and have no commitment to their proposals after election.
Voters are assumed to live in one of $2n+1$ districts, each of which elect a single legislator. The level of g is set independently for each district, but the tax rate is fixed at a single value, so legislators goal is to bargain a high level of expenditures in their own district.

Voters Voters are split in two groups, the "poor/unsuccessful" with income y_l and the "rich/sucessful" with income $y_h = \eta y_l$. Some share $\gamma > 1/2$ of voters are poor. The voters utility is quasi-linear in preferences over after-tax income and the level of public good provided in their district. I.e.

$$u_{ij} = (1-\tau)y^i + g(\pi^j \tau \bar{y}) \tag{13.1}$$

where $\bar{y}$ is the average income across all voters and π^j is the proportion of total revenue $\tau\bar{y}$ that goes to district j. From this we see that given any split π^j low type citizens prefer a higher tax rate than high type citizens[3] Letting τ_l^* and τ_h^* denote the preferred tax rates when revenue is split equally between the districts it must be that $\tau_l^* > \tau_h^*$.
Obviously given a τ the voters in district j prefers π^j to be set as high as possible.

Elections Elections are held at the same time across all districts. In each election one low type voter and one high type voter runs, and we assume that voters are distributed such that $\lambda > 1/2$ of districts have a majority of low type candidates. Applying the median voter theorem, any district with a majority of poor individual will have a median voter who is also poor who decides the election, and any rich district will have a rich median voter who decides the election.

Legislature The legislature might either be majority rich or majority poor depending on the elections. The way legislature works is that they first vote on a tax rate τ and then bargain for the shares π^j that goes to each district.

In the bargaining phase we assume that high income individuals are better at bargaining, by the reasoning that if you do well in the private sector you will also do well in the legislature. The abilities of legislators $(a_1, a_2, ..., a_N)$ are set such that $a_l = 1$ and $a_h = \eta$. The parameter γ governs how much ability matters for bargaining, and so the outcome of bargaining is $\pi^{j*}(a_1, a_2, ..., a_N|\gamma)$. Now if

[3] See $\partial_\tau u_{ij} = -y^i + \pi^j \bar{y} g'(\pi^j \tau \bar{y})$. Thus in optimum

$$\tau = \frac{1}{\pi^j \bar{y}} g'^{-1}\left(\frac{y^i}{\pi^j \bar{y}}\right) \tag{13.2}$$

bargaining ability has no influence the outcome for each district will be $\pi^{j*}(a_1, ..., a_N|\gamma = 0) \to 1/N$, that is districts simply split the revenue equally. We will further assume that $\partial_{a_j}\pi^{j*} \geq 0$ and $\partial_{a_k}\pi^{j*} \leq 0$ for all $j \neq k$. This simply says that higher bargaining ability of the legislator from district j increases the spending in district j won in bargaining, while higher bargaining abilities of legislators from other districts decrease the spending in district j.

The core question in this model is then who voters will elect, someone who share their views on the optimal τ or someone who can bargain a large share of the tax revenue to the home district. Naturally high type voters prefer to vote for a high type candidate, as they can get both low taxes and competent bargaining from the same candidate. Low income voters however face a tradeoff between getting a legislator who votes for higher taxes and one who bargains well. How this plays out depends on the value of γ:

1. if $\gamma = 0$ there os no tradeoff, so poor districts elect poor candidates. Since there is a majority of poor candidates the legislature will be majority poor and select a tax of τ_l^*. This is a fully representative equilibrium as the median voter is low, so is the median legislator and the outcome is in accordance with their preferences.

2. if γ is sufficiently high, voters concern with electing a competent politician dominates, so all districts select a high type candidate and the legislature implements τ_h^*. In this equilibrium there is no representativeness, as no legislators will be low type even though a majority of voters are.

3. if γ is moderate once a majority $n + 1$ have elected a low type it is certain that a low type candidate will by the MVT set the tax rate, and the remaining poor districts become free to select a candidate with high bargaining abilities meaning n high type candidates will enter the legislature. In this case the equilibrium is somewhat representative, as the tax rate is set by a low type, but there is a over representation of high type legislators in the legislature and as a consequence the tax rate will be set lower than τ_l^* in anticipation of bargaining, in which the districts with low type legislators will have to fund some of the high-type districts spending.

In summary this model predicts a tradeoff between electing competent politicians and getting a high tax rate for low-income individuals in district voting systems.

13.2 Empirics (Dal Bó et al., 2017)

Using purely descriptive analysis Dal Bó et al. (2017) use a dataset with the entire swedish population as well as various special information on politicians and candidates for elections.
This includes information from conscription tests including IQ and in some cases "leadership tests" undertaken with a psychologist.
To not mechanically measure a correlation between being in the "elite" (a politician) and competence they use parents background as a proxy for representation - i.e. if ones parents were regular people, it is fair to assume that one represents this group, and not the "elite" in the legislature (?).
The authors find a systematic tendency for politicians to be high competence individuals, with this trend being stronger the higher the office held. This selection is much less pronounced when comparing to parents background. In the sense that the relevant dimension for measuring representation is across generations this suggests that representation issues are not nearly as large as they could have been. In particular it seems that all parties select candidates who themselves are high competence, but tend to favor candidates whose parental background lies in different parts of the income distribution.
Importantly the lack of relation between parents background and being a politician does not mean that there is no intergenerational persistence in income and abilities etc. Instead the parties have stronger selection for candidates with background lower in the income distribution, evening out the correlation. In other words if your parents are rich you dont have to be very clever to become a politician, but if they're poor you need to be very smart. This selection removes the apparent relation between parents

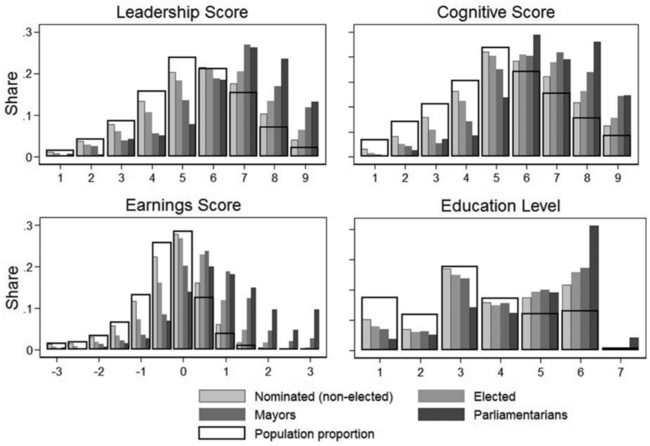

Figure 4: Results from Dal Bó et al. (2017) The results compares the distribution of various measures of competence with the population distribution of these traits.

abilities and being a politician.

In summary in Sweden politicians are highly competent compared to the population, but are representative in terms of their background.

References

Elisabeth R. Gerber and Jeffrey B. Lewis. Beyond the Median: Voter Preferences, District Heterogeneity, and Political Representation. *Journal of Political Economy*, 112(6):1364–1383, 2004. ISSN 0022-3808. doi: 10.1086/424737. URL https://www.jstor.org/stable/10.1086/424737.

Matz Dahlberg and Eva Johansson. On the Vote-Purchasing Behavior of Incumbent Governments. *The American Political Science Review*, 96(1):27–40, 2002. ISSN 0003-0554. URL https://www.jstor.org/stable/3117808.

Vladimir Gimpelson and Daniel Treisman. Misperceiving inequality. *Economics & Politics*, 30(1):27–54, March 2018. ISSN 1468-0343. doi: 10.1111/ecpo.12103. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecpo.12103.

Ilyana Kuziemko, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva. How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments. *American Economic Review*, 105(4):1478–1508, April 2015. ISSN 0002-8282. doi: 10.1257/aer.20130360. URL http://pubs.aeaweb.org/doi/10.1257/aer.20130360.

David P. Baron and John A. Ferejohn. Bargaining in Legislatures. *The American Political Science Review*, 83(4):1181–1206, 1989. ISSN 0003-0554. doi: 10.2307/1961664. URL https://www.jstor.org/stable/1961664.

Stephen Ansolabehere, James M. Snyder, Aaron B. Strauss, and Michael M. Ting. Voting Weights and Formateur Advantages in the Formation of Coalition Governments. *American Journal of Political Science*, 49(3):550–563, 2005. ISSN 0092-5853. doi: 10.2307/3647731. URL https://www.jstor.org/stable/3647731.

Alan S. Gerber and Donald P. Green. The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment. *The American Political Science Review*, 94(3):653–663, 2000. ISSN 0003-0554. doi: 10.2307/2585837. URL https://www.jstor.org/stable/2585837.

Andrea Mattozzi and Erik Snowberg. The Right Type of Legislator: A Theory of Taxation and Representation. Working Paper 24279, National Bureau of Economic Research, February 2018. URL http://www.nber.org/papers/w24279.

Ernesto Dal Bó, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne. Who Becomes A Politician? *The Quarterly Journal of Economics*, 132(4):1877–1914, November 2017. ISSN 0033-5533. doi: 10.1093/qje/qjx016. URL https://academic.oup.com/qje/article/132/4/1877/3859758.

David Strömberg. Radio's Impact on Public Spending. *The Quarterly Journal of Economics*, 119(1): 189–221, February 2004. ISSN 0033-5533. doi: 10.1162/003355304772839560. URL https://academic.oup.com/qje/article/119/1/189/1876059.

Toke Aidt, Jayasri Dutta, and Elena Loukoianova. Democracy comes to Europe: Franchise extension and fiscal outcomes 1830-1938. *European Economic Review*, 50(2):249–283, 2006. ISSN 0014-2921. URL https://econpapers.repec.org/article/eeeeecrev/v_3a50_3ay_3a2006_3ai_3a2_3ap_3a249-283.htm.

Alberto Alesina, Edward Glaeser, and Bruce Sacerdote. Why Doesn't the US Have a European-Style Welfare System? Working Paper 8524, National Bureau of Economic Research, October 2001. URL http://www.nber.org/papers/w8524.

Alberto Alesina, Stefanie Stantcheva, and Edoardo Teso. Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, 108(2):521–554, February 2018. ISSN 0002-8282. doi: 10.1257/aer.20162015. URL https://pubs.aeaweb.org/doi/10.1257/aer.20162015.

David S. Lee. Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics*, 142(2):675–697, February 2008. ISSN 03044076. doi: 10.1016/j.jeconom.2007.05.004. URL http://linkinghub.elsevier.com/retrieve/pii/S0304407607001121.

Fernando Ferreira and Joseph Gyourko. Do Political Parties Matter? Evidence from U.S. Cities. Working Paper 13535, National Bureau of Economic Research, October 2007. URL http://www.nber.org/papers/w13535.

D. S. Lee, E. Moretti, and M. J. Butler. Do Voters Affect or Elect Policies? Evidence from the U. S. House. *The Quarterly Journal of Economics*, 119(3):807–859, August 2004. ISSN 0033-5533, 1531-4650. doi: 10.1162/0033553041502153. URL https://academic.oup.com/qje/article-lookup/doi/10.1162/0033553041502153.

Per Pettersson-Lidbom. Do Parties Matter for Economic Outcomes? A Regression-Discontinuity Approach. *Journal of the European Economic Association*, 6(5):1037–1056, 2008. ISSN 1542-4766. URL https://www.jstor.org/stable/40283092.