WORKING PAPERS

N° 1050

November 2019

"Taxes and Turnout"

Felix Bierbrauer, Aleh Tsyvinski and Nicolas Werquin

Toulouse School of Economics

Taxes and Turnout*

Felix Bierbrauer University of Cologne

Aleh Tsyvinski Yale University

Nicolas Werquin Toulouse School of Economics

July 23, 2019

Abstract

We develop a model of political competition with endogenous platform choices of parties and endogenous turnout. A main finding is that a party that is leading in the polls has an incentive to cater primarily to the core voters of the opposing party. A party that is lagging behind, by contrast, has an incentive to cater to its own base. We analyze the implications for redistributive taxation and characterize the political weights that competing parties assign to voters with different incomes. Finally, we relate the comparative statics predictions of our model to the asymmetric demobilization strategy in the German elections in the era of Merkel.

Keywords: Political competition, Income Taxation, Turnout.

JEL classification: D72, D82, H21.

*We thank Gadi Barlevy, Marco Bassetto, Michael Becher, Luc Bridet, Micael Castanheira, Allan Drazen, Georgy Egorov, Hans-Peter Grüner, Anne Hannusch, Sinem Hidir, Bas Jacobs, Kai Konrad, Matthias Jung, Per Krusell, Alessandro Lizzeri, Salvatore Nunnari, Anna Paulson, Torsten Persson, Facundo Piguillem, Gerard Roland, Fernando Stipanicic Marquez, Andrea Moro, Sven-Oliver Proksch, Mattias Polborn, Sebastian Siegloch, Thomas Tröger, Justin Valazek, Karine Van der Straeten, Ngo Van Long, Marcelo Veracierto, Philipp Wangner, Matt Weinzierl, and John Weymark. Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866 is gratefully acknowledged. Werquin acknowledges funding from the French National Research Agency (ANR) under the Investissements d'Avenir program, grant ANR-17-EURE-0010.

Introduction

This paper has two main contributions. First, it develops a model of political competition in which the parties' platform choices and the voters' participation in elections are jointly determined in equilibrium. Second, it uses this framework for a political economy analysis of redistributive taxation.

Political competition. Most of the previous political economy literature has focused either on platform choices or on endogenous turnout. By combining the two we obtain a framework where parties face a tradeoff between, on the one hand, appealing to as many voters as possible, and on the other hand, ensuring that these potential voters turn out to vote. A potential voter of, say, party 1 is weakly better off if party 1 wins and implements its platform. Being among those who prefer party 1 over party 2 is, however, only a necessary condition for voting in favor of party 1. Potential voters are turned into actual voters only if the stakes are sufficiently high, i.e., they must be incentivized to fight for a victory of their party. A voter who is close to being indifferent between the two parties lacks such incentives, since in this case the gain in utility from having her preferred party elected does not justify incurring the voting cost. Thus, parties face a tradeoff between adopting polices that are good for the size of their electorate (or base) and policies that are good for mobilization. Consequently, best responses are such that parties don't aim at attracting as many supporters as they could, nor do they do everything they can to mobilize. The typical situation is that of a party that can enlarge its base by moving, say, to the right, but understands that such a move reduces its ability to mobilize its base.

First, we draw on the probabilistic voting model – see Coughlin and Nitzan (1981) and Lindbeck and Weibull (1987) – to determine how voters sort into the two parties' bases. Specifically, voters have both policy preferences and idiosyncratic party preferences. A voter can therefore be attracted to the base of party 1 because she likes the platform of party 1 better, or because she likes party 1 for exogenous reasons that we are not explicitly modeling. With well-behaved distributions of these preferences, a party's base responds in a continuous way to changes in the party's platform, and there are pure strategy equilibria even with multi-dimensional policy spaces. The probabilistic voting model is one of the workhorses in the formal analysis of party competition. However, this literature typically assumes that voter turnout is exogenous.

Second, we draw on models of ethical voting – originally proposed by Harsanyi (1980) and more recently analyzed by Coate and Conlin (2004) and Feddersen and Sandroni (2006) – to endogenize turnout. These models have been proposed as a way of adressing the paradox of voting.[1] It is assumed that voting is costly and that voting behavior is driven by a desire to fulfill a civic duty – formalized as a rule-utilitarian criterion for turnout. Thus, individuals choose a turnout rule (e.g., a probability of voting, or a threshold voting cost) that is optimal on the assumption that everyone with the same party preferences behaves according to the same rule.[2] Such group behavior is able to affect the outcome of the election, thus leading to non-trivial equilibrium turnout rates. These depend on how much voters have at stake: when their aggregate benefit from winning the election is higher, more individuals of a given group turn out to vote. This literature delivers predictions that are consistent with empirical facts on turnout, but it generally considers exogenous policy platforms.

Our formal analysis merges these two models so that both policies and turnout are endogenous outcomes. We focus on the implications of the trade-off between the number of potential voters (or "base") and mobilization. We establish conditions for equilibrium existence, fully characterize the equilibrium analytically, and provide a comparative statics analysis.

Political economy of redistributive taxation. Throughout, we use the political economy of redistributive income taxation as our main application. We have two reasons for this choice of policy space. First, it is a stylized fact that the rich are more likely to participate in elections than the poor. For a model with endogenous turnout, it is therefore interesting to investigate a policy domain that allows for a differential treatment of the rich and the poor. While our formal analysis does not rest on the assumption of differential turnout across income groups, the difference between welfare-maximizing and political equilibrium tax rates becomes very stark in this case. Second, while the literature developed a powerful *normative* theory of income taxation (see, e.g., Sheshinski (1972); Mirrlees (1971)), there is no such broadly

[1]The paradox is that observed turnout in elections is positive even though rational agents have no incentive to participate since the probability of being pivotal in large elections is negligible.

[2]Ethical voter models differ in some aspects: For instance, Feddersen and Sandroni (2006) model the electorate as being split between ethical and non-ethical voters. Coate and Conlin (2004) only have ethical voters in their framework. Our analysis is closer to Feddersen and Sandroni (2006), but we could as well have adopted the modelling choices of Coate and Conlin (2004). We provide a more detailed comparison of these approaches in the Online-Appendix, where we also show that these modelling choices are inconsequential for our main results.

accepted conceptual framework for a *positive* analysis of income taxation where tax rates arise as the outcome of political competition. Our theory can be viewed as a generalization of the seminal contribution by Meltzer and Richard (1981), allowing for the endogeneity of turnout and arbitrarily non-linear income tax schedules.

We show that the tax policy can be characterized as maximizing a virtual welfare function where the weights on the utilities realized by different types of voters arise endogenously in equilibrium. These *political weights* do not depend on whether taxes are linear or non-linear. Our main result is to characterize them in closed-form. When turnout is exogenous, parties place larger weights on income groups that have a larger number of *swing* voters, i.e., who are on the verge of indifference between the two candidates. When turnout is endogenous, the political weights at each income level are instead determined by the fractions of *core* supporters of both parties, i.e., their number of supporters when they propose the same platform. On the one hand, each party has an incentive to adopt a policy that is attractive to its own core voters, because increasing the stakes of its most loyal supporters is a way to get them out to vote. On the other hand, a policy that is appealing to the core voters of the competing party is also attractive, as this makes it harder for the rival party to mobilize its base. Crucially, we find that the extent to which a party caters to its rival's clientele, as opposed to its own, is increasing in its probability of winning the election. That is, the front-runner (resp., the runner-up) targets its campaign promises mostly to the opposition's (resp., to its own) core voters.

Asymmetric demobilization. We finally examine the German federal elections between 2005 and 2017 through the lens of our model. We focus on these episodes because the term *asymmetric demobilization* was coined by political analysts to characterize and explain the success of Angela Merkel's electoral campaigns. Following the near loss in the 2005 election where she catered primarily to her own core vorers, Angela Merkel adopted many positions of her main challenger, the social democratic party (SPD). For instance, the party manifesto for the 2013 election included demands for minimum wages, rent controls and a financial transactions tax, positions close to the heart of social democrats. We document this shift in positions using a quantitative text analysis of party manifestos. In all later elections (2009, 2013, and 2017) her party was at least 10 percent ahead of the SPD. Overall turnout in these elections fell to levels not witnessed before. Turnout was, moreover, asymmetric: the potential voters of the conservatives turned out in larger fractions than the potential voters of the SPD. We provide a detailed discussion of these elections using our theoretical framework and, importantly, show that these outcomes are aligned with the comparative statics predictions of our model.

Related literature. Our analysis relates to the literature on the paradox of voting. Downs (1957) and Riker and Ordeshook (1968) are classical references. Various approaches have since been explored to develop a model that explains both participation and abstentions in elections, see Feddersen (2004) for a survey.[3] We draw on one strand of this literature due to Harsanyi (1980), Coate and Conlin (2004) and Feddersen and Sandroni (2006). We square this ethical voter model with a model of party competition in which the parties' platform choices are endogenous variables, the probabilistic voting model due to Coughlin and Nitzan (1981) and Lindbeck and Weibull (1987).

Models of rational voting, often referred to as pivotal-voter models, are a prominent alternative.[4] An important contribution is by Ledyard (1984). In Ledyard's model, equilibrium turnout is positive if the alternatives are exogenously given and distinct. If the alternatives are instead endogenous variables, chosen by two competing candidates who seek to win an election, then the equilibrium has policy convergence, i.e. both candidates proposing the same platform, and, as a consequence, zero turnout. There are similarities and differences to our approach. Our equilibrium analysis also gives rise to policy convergence and we also find that policy convergence lowers turnout. We do not find that policy convergence implies zero turnout. In our framework, idiosyncratic party preferences imply that some voters turn out for party 1 and some voters for party 2, even if the two parties propose the same tax policy.

We contribute to the literature on the political economy of redistributive taxation. Many previous contributions have focused on the model of linear income taxation by Sheshinski (1972). Roberts (1977) has shown that the median voter's preferred policy is a Condorcet winner in this framework. Its empirical implications have been analyzed by Meltzer and Richard (1981). A prominent one is that redistributive taxes should be higher when pre-tax inequality – as measured by the gap between average and median income – is more pronounced. This prediction often fails in the data and

[3]More recent contributions include Callander and Wilson (2007), Degan and Merlo (2011) or Aldashev (2015).

[4]Coate et al. (2008) argue that the ethical voter model provides a better fit for data on turnout than the pivotal voter model.

this has led to a search for alternative explanations, see e.g. Alesina and Angeletos (2005) or Bénabou and Tirole (2006). Different turnout rates among the rich and the poor are one such explanation, see the discussion in Larcinese (2007); Sabet (2016) and the references therein. This literature has treated turnout as an exogenous variable; i.e., the possibility that turnout may depend on the parties' proposals has not been taken into account. Our analysis of the election campaigns in the era of Merkel in Germany shows that this feedback channel can be important.

Both for linear and non-linear income taxation, our equilibrium characterization uses the notion of generalized social welfare weights of Saez and Stantcheva (2016). Generalized social welfare weights facilitate a clarification of whether political equilibrium tax policies are as if a concave social welfare function was maximized, or, alternatively, whether political competition gives rise to a political failure, Besley (2006).

There is a rich literature in political science that investigates to what extent parties cater towards their core voters or to swing voters. Cox (2009) provides a survey of this literature. It has also been noted that parties may have an incentive to target their promises to the core voters of the competing party, see Erikson and Romero (1990) or Adams and Merrill III (2011). Our contribution relative to this literature is to provide a unified framework that nests these opposing forces. This makes it possible to have a systematic analysis of how parties should trade them off in an attempt to win an election. Our result that the front-runner in an election has stronger incentives to cater to the core voters of the competitor and that the runner up should focus primarily on its own core voters also arises under certain conditions in the recent paper by Bernhardt et al. (2018), but from a different channel: in their model, turnout is exogenous and parties care not only about their probability of winning but also their vote share.

Outline. The remainder is organized as follows. Section 1 introduces a general setup for an analysis of political competition that connects probabilistic voting with endogenous turnout. In Section 2 we apply this framework to characterize the political equilibrium tax rate and we derive empirically testable implications. In Section 3, we study German elections between 2005 and 2017 and interpret Merkel's *asymmetric demobilization* strategy through the lens of our model. Section 4 contains the analysis of non-linear taxes. Section 5 concludes with a discussion of generalized social welfare weights. Formal proofs are relegated to the Online-Appendix.

1 General framework

1.1 Endogenous turnout

Two political parties $j \in \{1, 2\}$ compete by choosing policies from a set of feasible policies $\mathcal{P}$. Party j's proposal is denoted by $p^j \in \mathcal{P}$. For now, we leave the set $\mathcal{P}$ abstract.

Preferences. There is a continuum of citizens of mass one. Citizens differ in their preferences over policies. To formalize this preference heterogeneity we distinguish different types of citizens. The set of types is denoted by Ω. For any $\omega \in \Omega$, we denote by $u(p, \omega)$ the utility that a type-ω citizen realizes under policy $p \in \mathcal{P}$. In the income tax application, ω will determine an individual's position in the income distribution and will thus shape preferences over redistributive taxation.[5] The cross-sectional distribution of types $\omega \in \Omega$ is common knowledge and represented by a cumulative distribution function F with density f.

Individuals not only have preferences over policy outcomes but also idiosyncratic party preferences. These preferences may be shaped by cultural and ethnic identities, party histories, or fixed party positions in certain policy domains. Formally, the random variable $\varepsilon \in \mathbb{R}$ denotes an agent's idiosyncratic preference for party 2. Conditional on ω, party preferences ε of different voters are independent and identically distributed. Thus, an individual with type ω and party preference ε supports party 1 if $u(p^1, \omega) \geq u(p^2, \omega) + \varepsilon$. We denote by $B(\cdot \mid \omega)$ the cumulative distribution function of party preferences ε among individuals of type ω, and by $b(\cdot \mid \omega)$ the corresponding density function. Therefore, the probability that a type-ω individual supports party 1 is given by

$$B(u(p^1, \omega) \; - \; u(p^2, \omega) \; \mid \; \omega).$$

In particular, if both parties propose the same policy $p^1 = p^2$, the fraction of agents with type ω who support party 1 is given by $B(0 \mid \omega)$. This formalism allows for the possibility that (say) party 2 receives more support than party 1 from high ω-types ("the rich" in the income tax application) and less from low ω-types ("the poor"), for instance, if the distributions $B(\cdot \mid \omega)$ can be ordered according to first order stochastic dominance (see also Dixit and Londregan (1998)).

[5] In the Online Appendix, we also sketch an application where ω is a measure of an individual's valuation of public goods.

Ethical voting and party bases. The mass of type-ω supporters of each party j is split into two groups: a fraction $1-\tilde{q}^j(\omega)$ of these agents always abstains, and a fraction $\tilde{q}^j(\omega)$ decides whether to vote based on a rule-utilitarian calculation.[6] The literature often refers to this last group as *ethical voters*.

We seek a framework where the election outcome is uncertain both from the voters' and the parties' perspectives. A convenient approach, also adopted by Feddersen and Sandroni (2006), is to assume that $\tilde{q}^j(\omega)$ is a random variable and that its realization is unknown when parties choose platforms and when potential voters decide whether or not to turn out. More specifically, we assume that $\tilde{q}^1(\omega)$ and $\tilde{q}^2(\omega)$ have the same expected value $\bar{q}(\omega) \in (0,1)$. That is, a type-$\omega$ supporter of party 1 is as likely to be of the ethical type as a type-ω supporter of party 2. Finally, we put the following structure on how realizations of $\tilde{q}^1(\omega)$ and $\tilde{q}^2(\omega)$ relate to the mean.

Assumption 1. *For each party j, there is a non-negative random variable η^j with mean 1 such that*

$$\tilde{q}^j(\omega) \quad = \quad \eta^j \cdot \bar{q}(\omega), \;\; \forall \omega \in \Omega.$$

This assumption stipulates that the uncertainty regarding the size of a party's base stems from a common shock that equally applies to all segments of the type distribution. This party-specific variable η^j can be interpreted as a characteristic of candidate j (say, likeability) or the exposition to negative advertising (see, e.g., Krupnikov (2011)) whose impact on voters' decision to participate is imperfectly measured and revealed only on election day. The possibility that, say, party 1 is affected by a positive shock $\eta^1 > 1$ and party 2 is affected by a negative shock $\eta^2 < 1$, or vice versa, generates uncertainty in election outcomes.[7] Assumption 1 is imposed in the sequel without further mention.

The ethical supporters are a party's potential voters. For ease of exposition, we also refer to the expected mass of these agents as a party's *base*. That is, given two

[6] We follow Coate and Conlin (2004) and Feddersen and Sandroni (2006) and assume that there are no "always-voters", i.e., individuals who come to the ballot irrespectively of how high their voting costs are. In the Online Appendix, we present a version of our model that includes such voters and gives rise to an equilibrium analysis that is equivalent to the one developed in the body of the text.

[7] Note that the shocks to the two parties' bases may be correlated. We do not impose an assumption of independence.

policies p^1 and p^2, the base of party 1 is given by

$$\mathbf{B}^1(p^1, p^2) \;=\; \mathbb{E}[\,\bar{q}(\omega)\, B(u(p^1, \omega) - u(p^2, \omega) \mid \omega)\,], \tag{1}$$

where the expectation operator $\mathbb{E}$ indicates the computation of a population average with respect to different types ω. We define the base of party 2 analogously, so that

$$\mathbf{B}^2(p^1, p^2) \;=\; \mathbb{E}[\bar{q}(\omega)] \,-\, \mathbf{B}^1(p^1, p^2). \tag{2}$$

Note that the two parties' bases add up to a constant. Hence, a change in the proposed policies that increases, say, the base for party 1, translates one-for-one into a decrease of party 2's base.

Net gains from winning – stakes. The *stakes* for the potential voters of party 1 are defined as the expected (utilitarian) welfare gain that is realized if a victory by party 2 is avoided. Formally,

$$W^1(p^1, p^2) \;=\; \mathbb{E}\left[\int_{\mathbb{R}} \max\left\{\, u(p^1, \omega) - [u(p^2, \omega) + \varepsilon]\ ,\ 0\right\}\, b(\varepsilon \mid \omega)\, \mathrm{d}\varepsilon\right]. \tag{3}$$

The integrand in equation (3) is the difference in utilities realized under the policies p^1 and p^2, including the welfare gains or losses due to the idiosyncratic party preferences ε. The max operator indicates that the summation over ε takes into account only the agents for whom this utility difference is positive, i.e., the supporters of party 1. Analogously, we define

$$W^2(p^1, p^2) \;=\; \mathbb{E}\left[\int_{\mathbb{R}} \max\left\{\, [\,u(p^2, \omega) + \varepsilon] - u(p^1, \omega)\ ,\ 0\right\}\, b(\varepsilon \mid \omega)\, \mathrm{d}\varepsilon\right]. \tag{4}$$

Voting costs. We denote by σ^j the fraction of ethical supporters of party j who actually turn out to vote. We define the (expected) aggregate voting cost of the ethical supporters of party j by $k(\sigma^j)\,\mathbf{B}^j(p^1, p^2)$, where for some $\mu \geq 0$,

$$k(\sigma^j) \;=\; \frac{\kappa}{1 + \frac{1}{\mu}}\,(\sigma^j)^{1+\frac{1}{\mu}}. \tag{5}$$

This isoelastic functional form unifies several cases. First, if all the ethical voters have the same per capita voting cost κ and choose an individual probability of voting

σ^j, we have $k(\sigma^j) = \kappa\, \sigma^j$, i.e., an infinite elasticity $\mu \to \infty$. Second, as in Coate and Conlin (2004) and Feddersen and Sandroni (2006), suppose that any one voter has a voting cost equal to $\kappa \times \sigma$, where σ is an idiosyncratic component that is i.i.d. across voters and uniformly distributed over the unit interval. In this case, σ^j is a cutoff so that all ethical supporters of party 1 with $\sigma \leq \sigma^j$ are turning out to vote. We then have $k(\sigma^j) = \kappa \int_0^{\sigma^j} \sigma \, \mathrm{d}\sigma = \frac{\kappa}{2}(\sigma^j)^2$, i.e., a unit elasticity $\mu = 1$. Third, as we argue below, an inelastic cost function with $\mu \to 0$ turns our setup into a standard probabilistic voting model with exogenous turnout.

Endogenous turnout. The ethical supporters of each party j adhere to a rule for participation in the election that maximizes their aggregate expected utility, taking the costs of voting into account. As a consequence, turnout depends on the parties' policy proposals. In the Online Appendix, we show that this problem of the ethical supporters of party j admits the following "calculus of voting" representation: taking as given the distributions of preferences and ethical voters, the policies (p^1, p^2), and the other party's turnout rule σ^{-j}, choose $\sigma^j \in [0, 1]$ to maximize

$$\pi^j(p^1, p^2, \sigma^1, \sigma^2)\, W^j(p^1, p^2) \; - \; k(\sigma^j)\, \mathbf{B}^j(p^1, p^2), \tag{6}$$

where $\pi^j(p^1, p^2, \sigma^1, \sigma^2)$ is the probability that party j wins the election. That is, the ethical supporters of party j face the following trade-off. On the one hand, a higher turnout σ^j raises the probability π^j that their favorite party wins, which in turn increases their aggregate welfare by W^j. On the other hand, a higher σ^j raises their aggregate voting costs as $k(\sigma^j)\, \mathbf{B}^j$ is increasing in σ^j.

Equilibrium turnout. Given p^1 and p^2, an equilibrium of the turnout game is a pair of turnout rates $(\sigma^{1*}(p^1, p^2), \sigma^{2*}(p^1, p^2))$ that are mutually best responses.

1.2 Equilibrium

Best responses and equilibrium policies. We assume that parties seek to maximize their probability of winning the election. For a given policy p^2, the best response problem of party 1 is to choose p^1 to maximize

$$\bar{\pi}^1(p^1, p^2) \;\; := \;\; \pi^1(p^1, p^2, \sigma^{1*}(p^1, p^2), \sigma^{2*}(p^1, p^2)).$$

A pair of equilibrium policies (p^1, p^2) satisfies $\bar{\pi}^1(p^1, p^2) \geq \bar{\pi}^1(\hat{p}^1, p^2)$, for all $\hat{p}^1 \in \mathcal{P}$ and $\bar{\pi}^1(p^1, p^2) \leq \bar{\pi}^1(p^1, \hat{p}^2)$, for all $\hat{p}^2 \in \mathcal{P}$. In such an equilibrium, parties take into account that alternative policy choices would also affect the equilibrium of the turnout subgame $(\sigma^{1*}(p^1, p^2), \sigma^{2*}(p^1, p^2))$. We are interested in equilibria that are interior, i.e., which are such that turnout responds at the margin to changes in proposed policies.[8] Equilibrium existence and uniqueness can be established by standard arguments – invoking Brouwer's fixed point theorem – when the policy space is one-dimensional. We cover this case in the context of the linear income tax model in Section 2. The non-linear income tax case needs a separate treatment, see Section 4.

1.3 Equilibrium characterization

We now turn to an equilibrium characterization that facilitates the analysis of specific policies in the context of our model.

Lemma 1 (**Relative turnout**). *Take the parties' proposals p^1 and p^2 as given. If the equilibrium of the turnout subgame $(\sigma^{1*}(p^1, p^2), \sigma^{2*}(p^1, p^2))$ satisfies the first-order conditions of the optimization problems (6) for $j \in \{1, 2\}$, then*

$$\frac{\sigma^{1*}(p^1, p^2)}{\sigma^{2*}(p^1, p^2)} \quad = \quad \left[\frac{W^1(p^1, p^2) \,/\, \mathbf{B}^1(p^1, p^2)}{W^2(p^1, p^2) \,/\, \mathbf{B}^2(p^1, p^2)} \right]^{\frac{\mu}{1+\mu}} . \tag{7}$$

The left-hand side of this equation is a measure of party 1's turnout advantage: the larger σ_1^*/σ_2^*, the larger the number of ethical supporters who turn out to vote for party 1, relative to the number of supporters who turn out to vote for party 2. The right-hand side is a ratio of the welfare gains per capita, $W^j/\mathbf{B}^j$, that the supporters of both parties can realize in case of winning the election. Thus, according to equation (7), the relative turnout for party 1 is increasing in the relative amounts that its supporters and the opposition have at stake.

The tradeoff between base and turnout. By combining Lemma 1 with Assumption 1 we find that party 1's probability of winning is a monotonic function of

[8] Corner solutions where all or none of the ethical voters participate are conceivable, but not very interesting. The analysis is then similar to the case with exogenous turnout.

a product of two terms:

$$R^{\mathbf{B}}(p^1, p^2) := \frac{\mathbf{B}^1(p^1, p^2)}{\mathbf{B}^2(p^1, p^2)} \quad \text{and} \quad R^{\sigma}(p^1, p^2) := \frac{\sigma^{1*}(p^1, p^2)}{\sigma^{2*}(p^1, p^2)} \, .$$

The first term is a measure of of the party's relative *base* advantage, the second term is a measure of its relative *turnout* advantage. We thus obtain the following lemma.

Lemma 2 (**Party objectives**). *Party 1's objective is to maximize*

$$R^{\mathbf{B}}(p^1, p^2) \; \times \; R^{\sigma}(p^1, p^2), \tag{8}$$

and party 2's objective is to minimize it.

Equation (8) reveals that parties face a trade-off between the size of their base and their turnout advantage. If turnout was exogenous, party 1 would simply focus on maximizing $R^{\mathbf{B}}(p^1, p^2)$, or equivalently its own base $\mathbf{B}^1(p^1, p^2)$ since equation (2) implies that $R^{\mathbf{B}}(p^1, p^2)$ is an increasing function of $\mathbf{B}^1$. If instead the base was exogenously given, party 1 would maximize its turnout advantage $R^{\sigma}(p^1, p^2)$. With both an endogenous base and endogenous turnout, party 1 faces a trade-off between maximizing the number of its supporters and maximizing their relative propensity to vote.

To illustrate this trade-off, suppose that p^{1*} is a best response of party 1 to the policy p^2 proposed by party 2. Also suppose that the policy space $\mathcal{P}$ is such that local deviations from p^{1*} are well defined. Now consider a deviation from p^{1*} that takes the form $p^{1*} + \epsilon\, \hat{p}$, where $\hat{p}$ describes the direction of the deviation from p^{1*} and $\epsilon > 0$ is a scalar that measures the size of this deviation. The deviation induces a payoff for party 1 equal to $R^{\mathbf{B}}(p^{1*} + \epsilon\, \hat{p}, p^2)\, R^{\sigma}(p^{1*} + \epsilon\, \hat{p}, p^2)$. Now, if p^{1*} is a best response, then it must be true that this expression is, for any feasible deviation $\hat{p}$, maximized by choosing the parameter $\epsilon = 0$. The corresponding first-order condition is

$$R^{\mathbf{B}}_{\hat{p}}(p^{1*}, p^2)\, R^{\sigma}(p^{1*}, p^2) \; + \; R^{\mathbf{B}}(p^{1*}, p^2)\, R^{\sigma}_{\hat{p}}(p^{1*}, p^2) \quad = \quad 0, \tag{9}$$

where $R^{\mathbf{B}}_{\hat{p}}(p^{1*}, p^2) \; := \; \frac{\partial}{\partial \epsilon}\, R^{\mathbf{B}}(p^{1*} + \epsilon \hat{p}^1, p^2)\big|_{\epsilon=0}$ denotes the (functional) derivative of $R^{\mathbf{B}}$ in direction $\hat{p}$ evaluated at the hypothetical pair of policies (p^{1*}, p^2), and $R^{\sigma}_{\hat{p}}(p^{1*}, p^2)$ is analogously defined.

To interpret this condition, suppose for simplicity that $\mathcal{P}$ is the unit interval and that p^{1*} lies in its interior. This applies, for instance, to the model of linear income taxation that we cover in Section 2. In this model, higher values of p^1 are associated with more redistributive or leftist policies. With $\mathcal{P}$ as the unit interval there is only one feasible direction of policy reform, so that $\hat{p}$ is simply a (positive or negative) constant. Now suppose that party 1's best response is such that its base could be increased by means of a more leftist policy, i.e., $R^{\mathbf{B}}_{\hat{p}}(p^{1*}, p^2) > 0$. Then, the first-order condition (9) implies that $R^{\sigma}_{\hat{p}}(p^{1*}, p^2) < 0$, so that by moving to the left in an effort to enlarge its base, the party sacrifices turnout. Thus, parties run into a trade-off that forces them to compromise the implications of their policy proposals for their base advantage with the implications of their policy proposals for their turnout advantage. Around the best response, what is good for the former is bad for the latter and vice versa.

Equilibrium characterization. Lemma 2 highlights that the tradeoff between base and turnout is central to our analysis. It does not yet illuminate, however, how parties resolve that tradeoff. Using equation (7) to substitute for the relative turnout σ^{1*}/σ^{2*} in (8) leads to our first main result.

Proposition 1 (Equilibrium characterization). *Party 1's objective is to maximize*

$$\psi^1(p^1, p^2) \quad := \quad \frac{1}{1+\mu} \log \frac{\mathbf{B}^1(p^1, p^2)}{\mathbf{B}^2(p^1, p^2)} \;+\; \frac{\mu}{1+\mu} \log \frac{W^1(p^1, p^2)}{W^2(p^1, p^2)}, \tag{10}$$

and party 2's objective is to minimize it. Thus, if (p^{1*}, p^{2*}) is a pair of interior subgame perfect equilibrium policies, then it is a saddle point of the function $\psi^1(p^1, p^2)$.

To understand Proposition 1, it is useful to focus in a first step on the polar cases $\mu \to 0$ and $\mu \to \infty$. Suppose that $\mu \to 0$, that is, the per capita cost of voting $k(\sigma^j)$ is inelastic. In this case, equation (10) implies that party 1 seeks to maximize its relative base $\mathbf{B}^1/\mathbf{B}^2$ or, equivalently, its own base $\mathbf{B}^1$. This objective is what party 1 would maximize in an environment with exogenous turnout. Thus, our framework nests the special case where turnout does not respond to the parties' policy platforms, or where parties do not consider the implications of their choices on voter participation.

Suppose next that $\mu \to \infty$, that is, the cost functions $k(\sigma^j)$ are linear. In this case, the bases no longer appear in expression (10), and party 1 only seeks to maximize W^1/W^2. This is because the influence of the base on σ^{1*}/σ^{2*} (equation (7)) exactly cancels with the expression $\mathbf{B}^1/\mathbf{B}^2$ that also appears on the right hand side of equation (8). Intuitively, suppose that party 1 proposes a policy that increases its base of supporters $\mathbf{B}^1$. The aggregate cost of participating $\kappa\,\sigma^1\,\mathbf{B}^1$ also increases proportionally with the size of the base. As a result, the relative turnout advantage for party 1 decreases one-for-one. Thus, the product of the relative base advantage and the relative turnout advantage, and hence the overall probability of winning, is unchanged. Therefore, if turnout is endogenous and voting costs are linear, parties do not care about the size of their base.

For intermediate values of $\mu \in (0, \infty)$, party 1 maximizes a weighted sum $\psi^1(p^1, p^2)$ of the relative base $\mathbf{B}^1/\mathbf{B}^2$ and the relative welfare gains W^1/W^2. Expression (10) is obtained by substituting for the relative turnout (7) in party 1's objective (8). It implies that the relative weight that party 1's platform choice puts on the objective W^1/W^2 is proportional to the elasticity μ of the cost function $k(\cdot)$.

Mobilizing one's supporters vs. demobilizing the opposition. Consider the case $\mu \to \infty$. Party 1 then has two competing aims. On the one hand, it would like to propose a policy p^1 that makes W^1 as large as possible, i.e., that makes its own supporters as well off as possible compared to the welfare they would obtain under the opposition's platform p^2. Doing so encourages its own supporters to turn out by increasing how much they have at stake in the election. On the other hand, party 1 would also like to propose a policy p^1 that makes W^2 as small as possible, i.e., that does not hurt party 2's supporters too much compared to the welfare they could obtain under *their* preferred policy p^2. Doing so discourages the opposition from turning out by lowering how much they have at stake. Therefore, the tradeoff for both parties is between *mobilizing their own base* (maximizing W^1) and *demobilizing their opponent's base* (minimizing W^2). With exogenous turnout these objectives of maximizing or minimizing what is at stake for each group would be irrelevant. Any positive, no matter how small, benefit W^1 for its supporters would deliver the same number of votes for party 1. Similarly, no attempt to soothe the opposition by lowering their stakes $W^2 > 0$ could ever increase party 1's vote share. With endogenous turnout, instead, the magnitude of W^1 and W^2 is crucial.

The tradeoff between mobilization and demobilization also appears for finite levels of μ. In that case, when choosing their platforms, parties also need to trade off these effects with the implications for the size of their parties' bases.

2 Linear income taxation

It is a stylized fact that individuals with higher incomes are more likely to participate in elections than individuals with lower incomes. Studying how the endogeneity of turnout affects equilibrium policies is therefore most interesting for polices that treat the rich and the poor differently. Our main focus is thus on the income tax-and-transfer system. We assume in this section that an admissible tax policy is affine, i.e., it consists of a constant labor income tax rate and a universal lump-sum transfer that balances the government budget. This is the setting analyzed by, e.g., Sheshinski (1972) and Meltzer and Richard (1981). In Section 4, we consider arbitrarily non-linear income taxation.

Setup. Individuals value after-tax income or private consumption denoted by c, and incur a disutility from the productive effort that is needed to generate pre-tax income y. Preferences of a type ω-individual over (c, y)-pairs are represented by the utility function $U(c, y, \omega)$, where u is twice continuously differentiable and satisfies $U_c > 0$, $U_y < 0$. For ease of exposition, we take the utility function to be quasilinear in c and the effort cost function to be isoleastic. Hence,

$$U(c, y, \omega) \quad = \quad c - \frac{1}{1+\frac{1}{e}} \left(\frac{y}{\omega}\right)^{1+\frac{1}{e}}. \tag{11}$$

A redistributive tax policy proposed by party j consists of a constant labor income tax rate τ^j and a universal lump-sum rebate $R^j \in \mathbb{R}$. Given such a policy (τ^j, R^j), voters choose their pre-tax income y so as to maximize their utility $U(c, y, \omega)$ subject to the budget constraint

$$c \quad = \quad (1 - \tau^j)\, y \, + \, R^j.$$

We denote by $y^j(\omega)$ the pre-tax labor income chosen by an agent with type ω given the policy τ^j of party j. It is easy to show that, since marginal effort costs are decreasing in the individual's type (i.e., the single-crossing condition $U_{y\omega} > 0$ holds), higher types ω earn higher (before- and after-tax) incomes than lower types. Thus, we refer to high types as "the rich" and low types as "the poor".

An admissible policy platform (τ^j, R^j) must balance the government budget.

Hence it must satisfy

$$R^j \;=\; \tau^j \, \mathbb{E}[\, y^j\,(\omega)\,]. \tag{12}$$

This requirement determines the transfer R^j as a function of the tax rate τ^j, so that the policy space $\mathcal{P}$ is one-dimensional. Without loss of generality, we assume in the following that admissible tax rates lie in a compact interval $\mathcal{P} = [\underline{\tau}, \overline{\tau}]$ that includes all Pareto-efficient ones. Thus, if $\tau \leq \underline{\tau}$ all individuals benefit from higher taxes, and if $\tau \geq \overline{\tau}$ all individuals benefit from lower taxes. In the sequel we simply denote party j's policy by τ^j, and let $u(\tau^j, \omega)$ denote the utility of agent ω realized under this policy.

Equilibrium existence. The following lemma provides a direct proof of the existence, uniqueness, and symmetry of a pure-strategy equilibrium under standard assumptions on preferences.

Lemma 3 (**Equilibrium existence, uniqueness and symmetry**). *Assume that the utility function $u(\tau, \omega)$ is single-peaked in the tax rate τ. Then there exists a unique subgame-perfect equilibrium. This equilibrium is symmetric, i.e., $\tau^1 = \tau^2$.*

The symmetry of the equilibrium is due to the fact that the first-order condition that τ^1 needs to fulfill to qualify as a maximizer of party 1's objective $\psi^1(\tau^1, \tau^2)$ is identical to the condition that τ^2 needs to fulfill to be a minimizer. That is, if for instance $\mu \to \infty$, in equilibrium the weight that party 1 places on mobilizing its own supporters is exactly equal to the weight that party 2 places on demobilizing the supporters of party 1. Conversely, the weight that party 1 places on demobilizing the supporters of party 2 is equal to the weight that party 2 places on mobilizing its own supporters. This shows that endogenous turnout does not by itself imply that parties have an incentive to differentiate from each other.[9]

Welfare-maximization. As a benchmark, we sketch the policy that would be chosen by a benevolent planner. Having this benchmark facilitates the interpretation of the political equilibrium tax rates that are characterized subsequently. Consider a

[9] If rather than maximizing their probabilities of winning the election, parties maximized their number of votes, or a weighted sum of these objectives, the equilibrium would no longer be symmetric.

social welfare function

$$\mathcal{SW}(\tau) \;:=\; \mathbb{E}\left[g(\omega)\, u(\tau,\omega)\right]; \tag{13}$$

where $g(\omega)$ is the *social welfare weight* placed on the utility of type ω-individuals. An unweighted utilitarian welfare function has $g(\omega) = 1$, for all $\omega \in \Omega$. A frequently considered alternative is to assign higher weights to poorer individuals so that the function $g : \Omega \to \mathbb{R}_+$ is decreasing. The following result is due to Sheshinski (1972).

Welfare-maximizing tax rate. *The welfare-maximizing tax rate τ^* satisfies*

$$\frac{\tau^*}{1-\tau^*} \;=\; -\frac{1}{e}\,\mathbb{C}\mathrm{ov}\left(\frac{g(\omega)}{\mathbb{E}[g(x)]}, \frac{y^*(\omega)}{\mathbb{E}[y^*(x)]}\right), \tag{14}$$

where e is the elasticity of the individual income $y^*(\omega)$ with respect to the retention rate $1-\tau^*$.[10]

Formula (14) highlights the following forces. First, the optimal marginal tax rate is decreasing in the elasticity of income e, which captures the efficiency costs of taxation. Second, and most importantly for our purposes, it is determined by the social welfare weights $g(\omega)$. Intuitively, for any given ω we have

$$g(\omega) \;=\; \frac{\partial\, \mathcal{SW}(\tau^*)}{\partial\, u(\tau^*,\omega)},$$

so that the term $\frac{g(\omega)}{\mathbb{E}[g(x)]}$ in equation (14) measures how much the planner values giving an additional unit of consumption (or utility) to agent ω – relative to giving an additional unit of consumption to everyone in the population. Now, the optimal tax rate τ^* depends on the covariance between these weights[11] and the agents' relative positions in the income distribution as measured by $\frac{y^*(\omega)}{\mathbb{E}[y^*(x)]}$. If the function g is decreasing, the covariance in (14) is negative, so that $\tau^* > 0$. The largest Pareto-efficient tax rate (i.e., the top of the Laffer curve) is equal to $\tau^* = \frac{1}{e}$ and is optimal for a Rawlsian planner with social welfare weights $g(\omega) = 0$ for all $\omega > \underline{\omega}$.

[10] *Under the functional form (11), we have $\frac{y^*(\omega)}{\mathbb{E}[y^*(x)]} = \frac{\omega^{1+e}}{\mathbb{E}[x^{1+e}]}$, so that equation (14) gives the welfare-maximizing tax rate in closed form as a function of the primitives of the model.*

[11] More generally, with a strictly concave utility function, the weights $g(\omega)$ would be multiplied in formula (14) by agent ω's marginal utility of consumption. In that case, even a utilitarian planner would choose a strictly positive tax rate since the marginal utility is decreasing.

2.1 Political equilibrium taxes

We now turn to the characterization of the tax rates that arise in a political equilibrium. To this end, we use Proposition 1: party 1's problem is to maximize the objective $\psi^1(\tau^1, \tau^2)$ defined in (10) and party 2's problem is to minimize this expression. Thus, if (τ^1, τ^2) is an equilibrium policy, then it is a saddle point of ψ. Proposition 2 below shows that party 1's platform choice is the same as if it were maximizing a welfare objective given by

$$\mathbb{E}\left[\, \gamma\left(\omega\right)\; u\left(\tau, \omega\right) \right], \tag{15}$$

where the weighting function $\gamma : \Omega \to \mathbb{R}_+$ is an endogenous object that we characterize in closed-form in Proposition 2. Put differently, the political equilibrium tax policy is given by a formula that looks formally like the one for welfare-maximizing taxes, but with endogenous *political weights* $\gamma\left(\omega\right)$ as opposed to exogenous welfare weights $g(\omega)$.

To highlight in the starkest possible way the novel insights implied by the endogeneity of turnout, we focus on the two polar cases where the cost function is either perfectly inelastic ($\mu \to 0$, exogenous turnout) or perfectly elastic ($\mu \to \infty$, endogenous turnout).

Proposition 2 (Political weights). *The tax policy τ^{eq} that arises in the symmetric equilibrium of the political game is given by formula (14), except that the social welfare weights $g\left(\omega\right)$ are replaced by the political weights $\gamma\left(\omega\right)$ given by:*

$$\gamma\left(\omega\right) \;\; = \;\; \bar{q}\left(\omega\right) b\left(0 \mid \omega\right) \tag{16}$$

if $\mu = 0$, and

$$\gamma\left(\omega\right) \;\; = \;\; \frac{1}{1+\psi_*^1} B\left(0 \mid \omega\right) + \frac{\psi_*^1}{1+\psi_*^1}\left(1 - B\left(0 \mid \omega\right)\right) \tag{17}$$

if $\mu \to \infty$, where

$$\psi_*^1 \;\; := \;\; \frac{\mathbb{E}\left[\, \int_{-\infty}^{0} \left|\varepsilon\right| \, \mathrm{d}B(\varepsilon \mid \omega) \right]}{\mathbb{E}\left[\, \int_{0}^{\infty} \varepsilon \, \mathrm{d}B(\varepsilon \mid \omega) \right]}. \tag{18}$$

While (14) is a *normative* formula that describes the optimal level of redistribution for a given social welfare objective, Proposition 2 gives a *positive* formula for the level of taxation. In particular, Proposition 2 has implications that can be tested empirically. Before we turn to these testable predictions, we provide a more detailed interpretation of the political weighting function $\gamma(\cdot)$.

Recall that party 1's objective $\psi^1(\tau^1, \tau^2)$, given by (10), is a weighted sum of three auxiliary, and possibly competing, objectives: (i) maximizing its base, $\mathbf{B}^1(\tau^1, \tau^2)$; (ii) maximizing what is at stake for its own supporters, $W^1(\tau^1, \tau^2)$; and (iii) minimizing what is at stake for its opponent's supporters, $W^2(\tau^1, \tau^2)$. We first study each of these auxiliary objectives separately.

Maximal size of the base. Suppose first that party 1 only seeks to maximize $\mathbf{B}^1(\tau^1, \tau^2)$. This is also the true objective when turnout is exogenous, i.e., $\mu = 0$. Given this objective, what is the political return to marginally raising the utility of agents with type ω? We know that equilibria are symmetric. So, to answer that question, we take as a starting point a situation with $\tau^1 = \tau^2$, so that $u(\tau^1, \omega) - u(\tau^2, \omega) = 0$ for all $\omega \in \Omega$. We then have

$$\left.\frac{\partial\, \mathbf{B}^1(\tau^1, \tau^2)}{\partial\, u(\tau^1, \omega)}\right|_{\tau^1 = \tau^2} \quad = \quad \bar{q}(\omega)\, b(0 \mid \omega).$$

That is, starting from the equilibrium, the number of voters of type ω that party 1 can expect to attract by marginally raising their utility is equal to the total mass of voters of that type, $\bar{q}(\omega)$, times the fraction of these voters that are on the verge of indifference between the two parties, $b(0 \mid \omega)$. This implies that it pays off for party 1 to place a higher weight on those types ω that have a higher participation rate and, crucially, a larger number $b\,(0 \mid \omega)$ of *marginal*, or *swing*, voters. This "swing voter" result is standard in probabilistic voting models with exogenous turnout (see, e.g., Lindbeck and Weibull (1987)).

Maximal mobilization of the base. Suppose now that party 1 only seeks to mobilize its own supporters by maximizing what they have at stake in the election, $W^1(\tau^1, \tau^2)$. We repeat the thought experiment above for this new objective: starting from a symmetric situation, the return to making agents with type ω marginally better off is given by

$$\left.\frac{\partial W^1(\tau^1,\tau^2)}{\partial u(\tau^1,\omega)}\right|_{\tau^1=\tau^2} \;=\; B\left(0 \mid \omega\right).$$

That is, the return is given by the the fraction of supporters of party 1 in this group, $B(0 \mid \omega)$. Therefore, party 1 has an incentive to place higher weight on those types ω among which it has a larger number $B(0 \mid \omega)$ of *inframarginal*, or *core*, supporters. Favoring these income groups is the most efficient way for party 1 to raise what its supporters have at stake in the election and, therefore, to boost their turnout. This "core voter" result, which strikingly contrasts with the "swing voter" result referred to above, has antecedents in the political science literature, in particular in models of indifference-based endogenous turnout (see, e.g., Cox and McCubbins (1986)).

Minimal mobilization of the opponent's base. Suppose finally that party 1 only seeks to demobilize its opponent's supporters by minimizing what they have at stake in the election, $W^2(\tau^1,\tau^2)$. We now have

$$\left.\frac{\partial W^2(\tau^1,\tau^2)}{\partial u(\tau^1,\omega)}\right|_{\tau^1=\tau^2} \;=\; -\left(1 - B\left(0 \mid \omega\right)\right).$$

That is, if party 1's platform makes agents ω marginally better off, the aggregate welfare benefit that party 2's ethical supporters could achieve by winning the election would decline by the fraction of agents that favor party 2, $1 - B(0 \mid \omega)$. Therefore, party 1 assigns more weight to those types ω among which its opponent, party 2, has a larger number $1 - B(0 \mid \omega)$ of *inframarginal*, or *core*, supporters. Favoring these income groups is the most efficient way for party 1 to lower what its opponent's supporters have at stake in the election and, therefore, to restrain their turnout. The importance of favoring the opposition's core voters has also been noted in the political science literature (see, e.g., Erikson and Romero (1990)).

Resolution of the trade-off. How does party 1 resolve the trade-off between these competing objectives? Suppose that turnout is endogenous with $\mu \to \infty$. Then, all that matters is the trade-off between mobilization (maximizing W^1) and demobilization (minimizing W^2). Catering marginally more to agents with type ω

raises party 1's objective $\psi^1(\tau^1,\tau^2) = \frac{W^1(\tau^1,\tau^2)}{W^2(\tau^1,\tau^2)}$ by

$$\begin{aligned} 0 &= \left.\frac{\partial \psi^1(\tau^1,\tau^2)}{\partial\, u(\tau^1,\omega)}\right|_{\tau^1=\tau^2} \\ &\propto \left.\left[\frac{\partial\, W^1(\tau^1,\tau^2)}{\partial\, u(\tau^1,\omega)} + \frac{W^1(\tau^1,\tau^2)}{W^2(\tau^1,\tau^2)} \times \left(-\frac{\partial\, W^2(\tau^1,\tau^2)}{\partial\, u(\tau^1,\omega)}\right)\right]\right|_{\tau^1=\tau^2}. \end{aligned}$$

Let τ^{eq} be the equilibrium tax rate. An inspection of this formula shows that, in equilibrium, the weight that party 1 places on the demobilization objective is given by $\frac{W^1(\tau^{eq},\tau^{eq})}{W^2(\tau^{eq},\tau^{eq})} = \psi^1(\tau^{eq},\tau^{eq}) =: \psi^1_*$, thus leading to formula (17). Note that ψ^1_*, given by (18), depends only on the model's primitives, not on the actual equilibrium policy τ^{eq} proposed by both parties. The numerator of ψ^1_*, $\mathbb{E}[\int |\varepsilon|\, \mathbb{I}_{\varepsilon<0} \mathrm{d}B]$, measures the average *intensity* of the idiosyncratic party preferences $\varepsilon < 0$ among those who belong to the base of party 1. The denominator, $\mathbb{E}[\int \varepsilon\, \mathbb{I}_{\varepsilon\geq 0} \mathrm{d}B]$, is the analogue for party 2. Thus, if the supporters of party 1 (resp., party 2) care more than the opposition about winning the election, because they have more at stake, then $\psi^1_* > 1$ (resp., $\psi^1_* < 1$).

Runner-up vs. front-runner. An implication of Proposition 2 is that, with endogenous turnout, the relative weight that party 1 places on the demobilization objective, W^2, is given by $\psi^1_* = \psi^1(\tau^{eq},\tau^{eq})$. Recall from Proposition 1 that there is an increasing relationship between this parameter and party 1's probability of winning the election. Formally, Lemma 2 and Proposition 1 imply that the odds of a victory by party 1 are equal to $P(\psi^1_*)$, where P is the c.d.f. of the random variable η^2/η^1. Suppose that ψ^1_* is large, so that party 1 is the likely winner of the election. Our analysis then implies that party 1 focuses primarily on minimizing what the core supporters of the opposition have at stake, i.e., W^2. Intuitively, this is because catering to the group that is unlikely to win (i.e., with small W^2) will lower its mobilization, in *percentage* terms, by a larger amount than that of the likely winner (i.e., with large W^1). Conversely, since party 2 is the runner-up, a symmetric analysis implies that it runs on a platform that focuses mainly on maximizing what its own base has at stake, i.e., W^1. Both parties cater to the same subset of the electorate, the core supporters of the runner-up party, but with different motives. The front-runner wants to secure an almost sure majority, while the runner-up seeks to boost the small probability of a surprise victory.

Implications for equilibrium taxes. To think through the implications of this equilibrium characterization for taxes, we consider a more specific example. Suppose that party 1 has little attachment from the rich. That is, the distribution of idiosyncratic political preferences ε is skewed towards the left ($\varepsilon < 0$) for small values of ω, and towards the right ($\varepsilon > 0$) for large values of ω, so that $\omega \mapsto B(0 \mid \omega)$ is decreasing. Suppose moreover that party 1 has a high chance of winning the election. Our analysis then predicts that party 1 will propose a policy that involves only small tax distortions and little redistribution. Symmetrically, the disadvantaged party with high attachment from the rich, party 2, also chooses to propose a moderate tax rate. The latter does so in an attempt to increase the stakes for its supporters, the former does so in order to undo this attempt. This observation rationalizes why even a more left-leaning party – a party with a comparatively high support from voters with low incomes – may shy away from heavier taxes. If the party is the likely winner of the election, low taxes on the rich are valuable as they help to discourage the supporters of the other, more right-leaning party. Only a left-leaning party with a small probability of winning chooses to propose high taxes on the rich, and hence a large amount of lump-sum redistribution, in order to raise the benefits that its own supporters would realize in the unlikely event that it wins the election. The same logic can explain why a more conservative or pro-market party that is supported to a large extent by voters with above average incomes may not propose to abandon a progressive tax system: if the party is the likely winner of the election, its focus is on weakening the support for the competing leftist party. It therefore chooses a platform that is appealing also to voters with below-average incomes. This *comparative statics* result on party positions with respect to their *ex-ante probability of winning the election* is our main novel result.

Political failures. It is possible that the political equilibrium weights $\gamma(\omega)$ are *not* monotonically decreasing along the income distribution. This is the case, for instance, in the model with exogenous turnout if the fraction of voters $\bar{q}(\omega)$ or the propensity to swing given identical policies $b(0 \mid \omega)$ are not decreasing. This is also the case in the model with endogenous turnout if the fraction of core supporters of the runner-up party, $B(0 \mid \omega)$ or $1 - B(0 \mid \omega)$, is not decreasing. In this situation, ceteris paribus, the equilibrium tax policy will be regressive (i.e., less redistributive than utilitarian) because each party j seeks to favor high incomes. This is either because the rich are more likely to turn out to vote, or because they are more likely to swing to a competing party that offers a lower amount of redistribution, or because the core supporters of the front-runner party are more numerous among poorer segments of the electorate. Such a regressive tax policy would never be chosen by a benevolent planner that allocates lower weights to agents with higher productivity in the social objective. We then say that there is a *political failure.* To be clear, the equilibrium policy is ex-post efficient in the set of linear tax policies, since $\gamma(\omega) \geq 0$ for all ω. It is, however, inefficient in an ex-ante sense: a risk-averse representative agent behind the veil of ignorance would be strictly better off with a higher tax rate.[12]

Equilibrium policies vs. best responses. Best responses can be analyzed along similar lines as equilibrium policies. We show in the Online Appendix that for a given policy τ^2 of party 2, the best response of party 1 is given by a formula identical to (14), except that the political weights $\gamma(\omega)$ derived in (16) and (17) are replaced by their off-equilibrium counterparts:

$$\gamma^1(\omega \mid \tau^1, \tau^2) \quad = \quad \bar{q}(\omega)\, b\,(\Delta u \mid \omega) \tag{19}$$

if $\mu = 0$, where Δu is a shorthand for $u(\tau^1, \omega) - u(\tau^2, \omega)$, and

$$\frac{1}{1 + \psi^1(\tau^1, \tau^2)} B\,(\Delta u \mid \omega) + \frac{\psi^1(\tau^1, \tau^2)}{1 + \psi^1(\tau^1, \tau^2)} \left(1 - B\,(\Delta u \mid \omega)\right) \tag{20}$$

if $\mu \to \infty$. To see how these expressions relate to the equilibrium tax in Proposition 2, note that $\tau^1 = \tau^2$ implies $\Delta u = 0$ and hence $\gamma^1(\omega \mid \tau^1, \tau^2) = \gamma(\omega)$ for all ω.

Discussion. The seminal work of Meltzer and Richard (1981) is known for its prediction that the "size of government" should be an increasing function of the level of inequality within the country, as measured by the ratio between median and average income. The analysis is based on a Downsian model of political competition with exogenous turnout. Empirical studies do not seem to support this prediction (see, e.g., Perotti (1996); Bassett et al. (1999)). Our framework predicts that the outcome of a political game is driven by different forces. Equilibrium taxes are shaped by the fraction of swing voters at each income level, and by the fractions of core voters of each party weighted by the ex-ante probability of winning the election.

[12] The distinction between ex ante, ex interim and ex post notions of efficiency is due to Holmström and Myerson (1983)).

Assessing whether campaign promises are primarily targeted at the swing voters or the core voters of either party is the object of a substantial empirical literature (see Cox (2009) for a survey).[13] Our analysis implies that all of these channels should be active empirically and, crucially, that the weight on each of them depends on the *ex-ante probability of a victory* by the two parties. The recent paper by Bernhardt et al. (2018) obtains, under certain conditions, a result with a similar flavor, but from a different channel: in their model, turnout is exogenous and parties care not only about their probability of winning but also their vote share. The literature is also informative on how the terms that appear in equations (16) and (17) can be measured empirically. For instance, the fraction of swing and core supporters of each party, as well as the parties' odds of winning the election, can be obtained using poll data. In Section 3 below, we provide more specific comparative statics results and illustrate them in the context of German federal elections. In particular, we will discuss how platform choices affect overall and relative turnout, and also how these effects can be traced in the data that is provided by election studies and by quantitative analyses of party positions.

2.2 Equilibrium turnout

So far we characterized the economic policy that emerges in equilibrium. We now study equilibrium turnout. Parties in our model do not care about turnout per se: they do only to the extent that it increases their probability of winning the election. For this purpose, only the turnout advantage over the competing party is important, i.e., parties prefer a large share in a small set of voters to a small share in a large set of voters. The following corollary enables us to formalize this claim.

Corollary 1 (Equilibrium turnout). *Suppose that $\mu \to \infty$. We then have*

$$\sigma^{1*}(p^1, p^2) \;=\; \frac{W^1(p^1,p^2)}{\kappa\,\mathbf{B}^1(p^1,p^2)}\,\psi^1(p^1,p^2)\,\rho(\psi^1(p^1,p^2))\,, \tag{21}$$

where $\rho(\cdot)$ is the density function of the random variable η^2/η^1.

[13] Several papers have exploited cross-electoral district (and, to a lesser extent, individual-voter) data to analyze this question. Evidence supporting the swing voter channel is provided by, e.g., Dahlberg and Johansson (2002); Stokes (2005). Evidence supporting the own-core voter channel is provided by, e.g., Levitt and Snyder (1995); Calvo and Murillo (2004). Whether campaign promises are also targeted to the opposition's core voters for demobilization purposes has been studied by, e.g., Erikson and Romero (1990); Adams and Merrill III (2011).

According to Corollary 1, if party 1 wanted to maximize the probability that its supporters actually turn out to vote – rather than its probability of winning – it would choose a policy p^1 so as to maximize the right hand side of (21). Analogously, if it wanted to maximize its number of voters, its objective would be $\sigma^{1*}\mathbf{B}^1$. In a subgame perfect equilibrium, by contrast, it chooses p^1 with the objective of maximizing $\psi^1(p^1, p^2)$. That is, in equilibrium parties have opportunities to increase their turnout rates and overall turnout in their favor, but choose not to use them as this would be detrimental for their probability of winning.

Previous analyses of ethical voter models by Coate and Conlin (2004) and Feddersen and Sandroni (2006) have shown that overall turnout increases if elections are close or if preferences over alternatives are more polarized. The observation that closeness is conducive to overall turnout has a resemblance to the implication of equation (21) that turnout is increasing in the semi-elasticity of the probability of winning with respect to ψ^1, which is given by $\psi^1(p^1, p^2)\,\rho(\psi^1(p^1, p^2))$. A close race is one in which this semi-elasticity is large: small changes in the proposed policies then have large consequences for relative turnout and each party's probability of winning. It is then implied by (21) that turnout tends to be larger if elections are close.[14]

3 Asymmetric demobilization in German elections

Our theoretical framework can be used to study the comparative statics of equilibrium policies or, more generally, of the parties' best response policies and turnout rates away from the equilibrium. At the level of generality of Sections 1 and 2, such an analysis is limited, however. A more specific application can provide the guidance to impose more structure on the general framework, which in turn makes it possible to derive sharper predictions. In this section, we analyze through the lens of our model the federal elections in Germany in the era of Angela Merkel, who centered her campaign strategy on *asymmetric demobilization*. The effect of this strategy on election outcomes and turnout rates is in line with the comparative statics predictions of a more structured version of our model.

[14]The observations that equilibria are symmetric and that a deviation from equilibrium policies would increase overall turnout can also be related to the finding in Feddersen and Sandroni (2006) that polarization drives turnout. Our analysis shows that a lack of polarization leaves room for an increase of overall turnout.

3.1 A more specific model

We start by applying the theoretical framework of Sections 1 and 2 in the context of German politics between 2005 and 2017. Merkel became the leader of the Christian democrats (CDU, center-right) in 2000 and successfully ran for the chancellory in 2005, 2009, 2013 and 2017. Her main competitor was the Socialdemocratic Party (SPD, center-left).

Policy space $\mathcal{P}$. As in Section 2 we consider a policy space of linear income taxes and let the variable ω index an individual's position in the income distribution. One can interpret a redistributive policy platform τ^j of party j either narrowly or broadly. First, as in Meltzer and Richard (1981), the tax rate τ^j can be interpreted as the "size of the government" or of the welfare state, which includes income taxes and monetary transfers but also social insurance, public education, etc. Second, τ^j can be interpreted more broadly as an index of the party's position on the "left-right" axis, a higher value of τ^j corresponding to a more leftist platform. In addition to the previous variables, this broad index would also account for, e.g., the party's stance on the minimum wage, gay marriage, or nuclear energy. While not strictly speaking redistributive across incomes, these policy choices played an important role in Merkel's campaign strategy discussed below.

Distribution of political biases $B(\cdot)$. By convention and without loss of generality, we interpret smaller values of ε as more "liberal" preferences, and larger values of ε as more "conservative" preferences. Thus, we identify party 1 with the SPD, and party 2 with the CDU: given identical policy platforms $\tau^1 = \tau^2$, party 1 (resp., party 2) is overly supported by voters with liberal preferences $\varepsilon < 0$ (resp., conservative preferences $\varepsilon > 0$). In practice, these party preferences may be shaped by party identities, e.g., roots in the worker's movement or the Christian churches, or the cultural milieu from which parties recruit their members. Party preferences may also reflect fixed party positions that are not adjusted in the political campaign. For instance, a salient issue in German politics in the Merkel era was whether families should be supported by direct transfers, as advocated by the CDU, or by publicly-provided childcare, as preferred by the SPD.

We assume that potential voters of the SPD have stronger preferences for redistributive policies than potential voters of the CDU.[15] Therefore, we suppose that the electorate of party 1 (SPD) is over-represented among the poor (i.e., low ω), while party 2 (CDU) is over-represented among the rich (i.e., high ω). Thus, for the comparative statics analysis that follows, we take as a starting point policies (τ^1, τ^2) such that:

(i) Party 1 has more potential voters with low incomes. Formally, the function $B(u(\tau^1, \omega) - u(\tau^2, \omega) \mid \omega)$ is decreasing in ω.

Status quo policy τ^2. The 2005 election was an early election called by Merkel's predecessor from the SPD, Gerhard Schröder. After an adoption of controversial labor market reforms, the SPD had lost various state elections. When the 2005 election was called, the CDU had a strong 21 percent lead over the SPD in opinion polls, and was expected to become by far the strongest party. The CDU decided to run on a pro-market platform, emphasizing the need for deregulation and lower taxes. Over the course of the election campaign, however, the SPD recovered and in the end the CDU won only by a tiny margin of victory: it was only 1 percent ahead of the SPD. Notice that this outcome is consistent with the predictions of our model. Because it was the clear front runner, the CDU would have maximized its chances of victory by focusing on demobilizing the opposition, rather than mobilizing its own electorate – i.e., by running on a redistributive platform more favorable to the SPD's core supporters rather than taking a fiscally conservative stance to benefit its own base. Therefore, the CDU's status quo policy τ^2 ahead of the 2009 election was "too far" to the right, i.e.:

(ii) A slight increase in the status quo tax rate τ^2 would raise party 2's probability of winning the election.

We show in the Online Appendix that this condition is satisfied if τ^2 is small enough, or if it is below and close enough to the best response $\tau^{2*}(\tau^1)$ to party 1's policy.

[15] For instance, the election outcomes in 2009 and 2013 show the following pattern: the vote shares of SPD and CDU among public servants and white collar workers were, by and large, in line with the parties' overall vote shares, see Jung et al. (2010, 2015). Hence, in absolute numbers, the CDU got more votes from these groups than the SPD. In relative terms, the CDU was stronger among the self-employed and the SPD among workers. The CDU voters also tend to be older and more formally educated. Thus, SPD voters benefit to a larger extent from redistributive policies.

Odds of winning $\bar{\pi}^1(\tau^1, \tau^2)$. In 2009, the CDU was clearly headed for reelection. The polls estimated that the CDU would get 35 percent of the votes, against 25 percent for the SPD and less than 15 percent for all the other parties. A week before election day, Merkel traded at 1.08 (1/12) in the "next Chancellor" market on Betfair – i.e., party 1 was given a chance of winning $\bar{\pi}^1(\tau^1, \tau^2)$ of 8 percent. This motivates the following assumption:

(iii) Party 2 is the likely winner of the election, i.e., the probability that party 1 wins is $\bar{\pi}^1(\tau^1, \tau^2) < 1/2$.

3.2 CDU's *asymmetric demobilization* strategy

The term *asymmetric demobilization* has been used to characterize Merkel's strategy vis à vis her main competitor (SPD) from 2009 onward.[16] This strategy is associated with Matthias Jung, an advisor of Angela Merkel and the head of the *Forschungsgruppe Wahlen* research institute that studies elections in Germany. He published on the rationale for this delibrate strategy and its impact on voting behavior and election outcomes: see Jung et al. (2010, 2015); Jung (2019). For an account in English language, see Schmidt (2014).

Definition and chronology. After the federal election in 2005 the CDU adopted the strategy of *asymmetric demobilization*. What defines this strategy is an avoidance of controversial positions or even an adoption of the rival's position in an attempt to lower the turnout of its potential voters. This strategy was successful and continued during the 2013 and 2017 campaigns. The clearest illustration is given by the 2013 official CDU program, which included many policies traditionally advocated by the SPD including the creation of a minimum wage, rent control in tight city areas, a financial transactions tax, a floor on pensions, or tax credits for families and single mothers. In addition, in 2011 Merkel had announced a plan to shut down all nuclear reactors by 2022, a measure traditionally favored by the left-leaning Green party. In 2017, the CDU avoided controversial topics on economic and social policy, and Merkel initiated a parliamentary decision on the question of gay marriage that her SPD opponent had made a central campaign issue – at the cost of alienating her own base. Narrative records of this strategy abound in the national and international press.

[16] It seems that the term *asymmetric demobilization* had its first appearance in an analysis of a regional election in Catalonia, see Lago et al. (2007).

To give but one example, Josef Joffe, a well-known German journalist, commented on the CDU's strategy in 2013: "*Ms Merkel's plan is to lull the other side; don't rile them and win by keeping them at home. How did she do it after the near-disaster of 2005? By shifting to the left. An apostle of free markets and low taxes ten years ago, Merkel simply outflanked the left on the left ... She is the best Social Democrat the SPD could have asked for.*"[17] While such journalistic documentation of the CDU's asymmetric demobilization strategy is overwhelming, it is also apparent in systematic quantitative analyses of party positions by political scientists, as we now describe.

Data sources. The *Manifesto Project*, see Volkens et al. (2018), provides a quantitative text analysis of party manifestos. The text is split into quasi-sentences, units of text that contain one political statement. Quasi-sentences are then assigned to categories such as *Free Market Economy*, *Market Regulation*, *Welfare State Expansion* or *Welfare State Limitation*.[18] Following our discussion of the policy space $\mathcal{P}$ discussed in Section 3.1, we focus on two such indices. See Volkens et al. (2018) for a detailed description of the data set and the methodology.[19]

First, we use the *Welfare State* index, which corresponds to our narrower interpretation of a policy platform τ^j. This index aggregates all of the favourable mentions of the "need to introduce, maintain or expand any public social service or social security scheme ... for example: government funding of health care, child care, elder care and pensions, social housing"; and of "equality: concept of social justice and the need for fair treatment of all people". Second, we use the *Right-Left* index, which corresponds to our broader interpretation of a policy platform τ^j. This index positions a party manifesto on a one-dimensional policy space by taking the share of quasi-sentences that are indicative of rightist positions (e.g., favorable mentions of military, freedom and human rights, constitutionalism, political authority, free market economy, incentives, economic orthodoxy, welfare state limitation, national way of life, traditional morality, law and order, civic mindedness) and substracts the share of quasi-sentences

[17] Financial times, 08-05-2013.

[18] The overall analysis is not restricted to economic policy dimensions, but also contains categories for positions on foreign policy, migration, political corruption and others.

[19] An alternative data source is the *Chapel Hill Expert Survey*, see Polk et al. (2017); Bakker et al. (2015). It also provides an analysis of party positions in various dimensions, including a left-versus-right positioning for economic policy issues. It differs from the *Manifesto Project* in that it is based on a survey of expert opinions as opposed to the text of party manifestos. This data set does not yet cover the most recent federal election in Germany in 2017. For the elections between 2002 and 2013 it shows the same pattern as the Party Manifesto data.

that are indicative of leftist positions (e.g., favorable mentions of anti-imperialism, peace, internationalism, democracy, economic planning, protectionism, nationalization, welfare state expansion, education expansion, labor groups).

Results. The table below describes how the positions of the CDU and the SPD evolved according to the two indices from the Manifesto Project for the federal elections since 2002. Both indices are normalized to 1 for the SPD in 2002. Larger (resp., smaller) values of the "Welfare State" index mean that the party's manifesto puts stronger (resp., weaker) emphasis on the expansion of the welfare state. Larger (resp., smaller) values of the "Right-Left" index mean that the party's manifesto is located further to the right (resp., left). This table shows clearly that the party positions diverged between the 2002 and 2005 elections. While the SPD reinforced its emphasis on welfare state expansion (the Welfare State index increased from 1 to 1.49) and overall moved further to the left (the Right-Left index decreased from 1 to -0.53), instead the CDU advocated a smaller welfare state (the corresponding index decreased from 0.85 to 0.58) and overall moved further to the right (the corresponding index increased from 5.06 to 6.25). From 2009 onwards, instead, the CDU moved to the left according to both indices: the welfare state index increased continuously from 0.58 in 2005 to 1.08 in 2017, and the right-left index decreased from 6.26 in 2005 to 0.67 in 2017. The two parties moved in parallel: when the SPD moved to the left, so did the CDU. Notice that according to both indices, the CDU was substantially more left-leaning in 2017 than the SPD was in 2002.

	Welfare State		**Right-Left**	
	SPD	**CDU**	**SPD**	**CDU**
2002	1	0.85	1	5.06
2005	1.49	0.58	-0.53	6.25
2009	1.76	0.74	-4.46	2.13
2013	2.14	0.83	-5.75	0.63
2017	1.83	1.08	-5.23	0.67

3.3 Analysis of turnout and election outcomes

In this section we analyze the impact of the CDU's asymmetric demobilization strategy on turnout rates and election results. Our goal is to confront the comparative statics predictions of our model with the outcomes of German elections from 2009 to 2017.

Comparative statics predictions. A major insight of our theoretical analysis is that a party that is leading in the polls has an incentive to adopt a platform that is appealing to the core supporters of its competitor. Thereby the potential voters of the competitor are demobilized. Corollary 2 below is an adaptation of this finding to the German context described above. For convenience, we invoke additional functional form assumptions.

Assumption 2. *Idiosyncratic party biases ε follow a uniform distribution at each income level: for any $\omega \in \Omega$ there exist $\mathfrak{B}(\omega) \in (0,1)$ and $\mathfrak{b}(\omega) > 0$ such that*[20] $B(x \mid \omega) = \mathfrak{B}(\omega) + \mathfrak{b}(\omega)\, x$.

Under Assumption 2, if both parties make the same proposal, so that $\tau^1 = \tau^2$ and hence $u(\tau^1, \omega) - u(\tau^2, \omega) = 0$, then $B(u(\tau^1, \omega) - u(\tau^2, \omega) \mid \omega) = \mathfrak{B}(\omega)$. Thus, $\mathfrak{B}(\omega)$ captures party 1's strength in the subset of type-ω citizens when the parties propose the same policies. Thus, it is a measure of its core supporters. Analogously, $1 - \mathfrak{B}(\omega)$ is a measure of party 2's core supporters. The parameter $\mathfrak{b}(\omega)$, by contrast, measures the fraction of swing supporters of type ω.

Assumption 3. *The random variables η^1 and η^2, defined in Assumption 1, are uniformly distributed on an interval $[1-\delta, 1+\delta]$ with $\delta > 0$.*

As the proof in the Online-Appendix makes clear, Assumptions 2 and 3 are sufficient, but by no means necessary, to obtain our next result.

Corollary 2 (Asymmetric demobilization). *Suppose that Assumptions 2 and 3 and Conditions (i)-(iii) above are satisfied. Then a slight increase in party 2's tax rate τ^2 has the following implications. Party 2's expected vote share increases. Overall turnout decreases. The demobilization is asymmetric, i.e., the relative turnout ratio σ^{1*}/σ^{2*} of parties 1 and 2 decreases.*

In the remainder of this section we confront the theoretical predictions in Corollary 2 with the election outcomes in Germany.

[20] Whenever we invoke this assumption, we assume that the support of the distribution is sufficiently wide so that all the values of $x = u(\tau^1, \omega) - u(\tau^2, \omega)$ that we are concerned with fall in the interior of the domain of $B(\cdot \mid \omega)$.

Empirical election outcomes. As we discussed above, the strategy of asymmetric demobilization was adopted in the 2009, 2013 and 2017 elections in response to the 2005 experience, in which Merkel learned that running on a platform that appeals to the core voters of her own party could jeopardize an almost sure victory. This strategy paid off: despite a similar lead in the polls in 2009, her margin of victory over the SPD increased from 1 percent in 2005 to more than 10 percent. Overall turnout (70.8 percent) went down by 6.9 percentage points compared to the 2005 election, and was at an all-time low. Crucially, turnout was lower among potential SPD voters than among potential CDU voters: 52 percent of the potential SPD voters indeed voted for the SPD, whereas 62 percent of the potential CDU voters voted for the CDU, see Jung et al. (2010); Forschungsgruppe Wahlen (2013b,a).[21]

In 2013, the CDU moved further left in parallel with the SPD. The election outcome was again a great success for the CDU: it gained 41.5 percent of the votes, was close to an absolute majority in parliament, and was 16 percent ahead of the SPD. Again, mobilization was asymmetric: turnout was 51 percent among the potential SPD voters and 69 percent among the potential CDU voters, see Forschungsgruppe Wahlen (2015). In 2017, the rise of a right-wing populist party implied large losses for the CDU relative to the 2013 election. The SPD also lost, however, and so the CDU stayed more than 12 percent ahead of the SPD. Moreover, it defended its dominant position in the German party system: as the only party with more than 30 percent of the votes, every realistic option for government formation had the CDU in the leading role with Merkel as the chancellor. Again, turnout of potential CDU voters (60 percent) was much higher than the turnout of the potential SPD voters (44 percent), see Forschungsgruppe Wahlen (2018). Overall turnout was slightly higher in 2013 and somewhat higher in 2017 than it was in 2009, at 76 percent, but still lower than any turnout ratio observed prior to 2009.

These outcomes are all consistent with our theoretical comparative statics predictions of Corollary 2.

[21] These numbers are obtained in the following way: The research institute Forschungsgruppe Wahlen runs a monthly survey with a representative sample of voters. The study is known as the *Politbarometer*. Shortly before an election it includes questions on prospective voting behavior. A person who plans to vote SPD or who includes the SPD in the set of conceivable parties is considered a potential SPD voter. Likewise for the CDU. The ratio of actual to potential voters then gives the numbers of 62 percent for the CDU and of 52 percent of the SPD. As a caveat, note that the Politbarometer is not a panel; i.e., it is not tracking the actual voting behavior of the participants in the survey.

4 Nonlinear income taxation

In this section, we extend our analysis of equilibrium tax rates to non-linear tax schedules. We show that our equilibrium characterization for linear taxes carries over to the non-linear setting, in the following sense: the political equilibrium is formally given by the same formula as the one that would be chosen by a benevolent social planner, except that the social welfare weights are again replaced by the political weights introduced in Proposition 2. The formal arguments in the Appendix are different though. The set of non-linear income tax schedules is an infinite-dimensional policy space, so that we need to involve functional derivatives. Moreover, the conditions for equilibrium existence are more restrictive.

As in our analysis of linear taxes, preferences of taxpayers are given by (11). A redistributive tax policy proposed by party j now consists of a twice continuously differentiable function $T^j : \mathbb{R}_+ \to \mathbb{R}$ that assigns a tax payment $T^j(y)$ to any level of pre-tax-income y. An admissible policy platform T^j must balance the government budget, i.e.,

$$\mathbb{E}[T^j(y^j(\omega))] \;=\; 0, \tag{22}$$

where $y^j(\omega)$ the pre-tax labor income of agent ω chosen under tax schedule T^j. We denote by $u(T^j, \omega)$ denote the utility of agent ω realized under this policy.

Equilibrium existence. The following lemma provides sufficient conditions on primitives ensuring the existence, uniqueness, and symmetry of a pure-strategy equilibrium when turnout is endogenous.

Lemma 4 (Equilibrium existence, uniqueness, and symmetry). *Let $\mu \to \infty$. Suppose that Assumption 2 holds, and that there are $\alpha > 0$ and $\beta > 0$ such that $\mathfrak{B}(\omega) \in [\frac{1}{2} - \alpha, \frac{1}{2} + \alpha]$ and $\mathfrak{b}(\omega) \leq \beta$ for all ω. Then there exists a unique equilibrium. This equilibrium is symmetric, i.e., $T^1 = T^2$.*

The assumptions of Lemma 4 imply that the race between the two parties is close, in the sense that both parties attract close to half of the electorate when they make the same proposal. They also imply that the distribution of idiosyncratic party preferences has a wide support, that is, $\mathfrak{b}(\omega)$ is sufficiently small. These distributional assumptions primarily serve to simplify the analysis and are stronger than necessary:

the equilibrium conditions are continuous in these parameters so that there is a range of close enough distributions for which they are also fulfilled.

The proof in the Appendix involves the following steps. First, we use functional derivatives to derive first-order conditions that characterize the parties' best responses. This gives us an equilibrium candidate. Second, we show that, under the conditions stated in Lemma 4, this equilibrium candidate satisfies also the second-order conditions. Thus, parties have no incentive to deviate locally. Finally, we invoke the contraction mapping theorem to show that, under the conditions in Lemma 4, there is one and only one intersection of the parties' best response functions. Hence, there is no incentive to deviate to a policy that is not in a neighborhood of the equilibrium candidate.

Welfare-maximization. Again, we start by describing as a benchmark the policy that a benevolent social planner would choose. The following result is due to Diamond (1998).

Welfare-maximizing tax schedule. *The welfare-maximizing tax schedule T^* satisfies*

$$\frac{T^{*\prime}\left(y^*\left(\omega\right)\right)}{1-T^{*\prime}\left(y^*\left(\omega\right)\right)} = \left(1+\frac{1}{e}\right)\frac{1-F\left(\omega\right)}{\omega f\left(\omega\right)}\left\{1-\mathbb{E}\left[\frac{g(x')}{\mathbb{E}[g(x)]}\mid x'\geq\omega\right]\right\}. \tag{23}$$

Formula (23) shows that optimal marginal taxes are, as in the linear case, decreasing in the intensity of behavioral responses to taxation as measured by the elasticity e. Also, the shape of the tax schedule is driven by the inverse hazard ratio of the wage distribution, $\frac{1-F(\omega)}{\omega f(\omega)}$. This term highlights an equity-efficiency tradeoff. It relates the mass $1-F(\omega)$ of people who pay a higher tax liability after a slight increase in the marginal tax rate at income $y^*\left(\omega\right)$ (i.e., the number of agents with income $y \geq y^*\left(\omega\right)$) to the fraction $f(\omega)$ of people whose incentives to exert effort are worsening after such an increase. Finally, as for linear taxation, the social welfare weights $g(\omega)$ shape the optimal tax schedule. These weights reflect inequality aversion if the function g is decreasing. With g decreasing, optimal marginal taxes are, ceteris paribus, higher for higher incomes.

Political equilibrium tax schedule. We finally turn to the analysis of the political competition game between two parties. Taking as given the opponent's policy T^2, the best response problem of party 1 is to choose a policy T^1 that maximizes the objective $\psi^1(T^1, T^2)$ defined in (10) subject to the budget constraint (22). Party 2 solves the analogous problem. In equilibrium, the two parties tax schedules are mutually best responses.

Proposition 3 (**Political equilibrium nonlinear tax policy**). *The equilibrium tax policy is given by formula (23), except that the social welfare weights $g(\omega)$ are replaced by the political weights $\gamma(\omega)$ defined by (16) if $\mu = 0$ and by (17)-(18) if $\mu \to \infty$.*

Proposition 3 shows that the equilibrium tax schedule is given by a formula that is formally identical to the one that would be chosen by a benevolent planner with a social welfare objective that allocates the weights $\gamma(\omega)$ on agents with wage ω. These weights are identical and hence have the same interpretation and economic implications as in the affine case. In other words, the political economy forces determine the relevant Pareto weights that appear in the tax formulas, but keep unchanged all of the other insights that the public finance literature has derived.

5 Conclusion: generalized social welfare weights

The theory of optimal taxation characterizes the tax policy that would be chosen by a benevolent social planner with the objective to maximize social welfare. However, there are many "reasonable" social welfare functions. The literature then often refers to "society's social welfare function" as the relevant one but typically fails to provide a precise formulation of how society would come to agree on a given social welfare function. Still there are attempts to identify society's social welfare function empirically. The idea is to use the formulas for welfare-maximizing taxation, data on the tax system, the earnings distribution and labor supply elasticities to map out society's social welfare weights; see, e.g., Blundell et al. (2009). Bargain et al. (2011), Bourguignon and Spadaro (2012), Zoutman, Jacobs and Jongen (2016) or Lockwood and Weinzierl (2016).

Our characterization of political equilibrium taxes provides a different interpretation of the weights that can be mapped out given data on the tax system, the earnings distribution and labor supply elasticities. This interpretation, moreover, depends on whether turnout in elections is viewed as exogenous or endogenous. With exogenous turnout, the weights reflect the political return to attracting swing voters. With endogenous turnout, they reflect the political return to mobilizing the own base and the political return to demobilizing the base of the competing party, as well as their interaction with the parties' odds of winning the election. This difference between the cases of exogenous and endogenous turnout raises the question which of these interpretations is more plausible.[22] We used our model to analyze the federal elections in Germany in the era of Angela Merkel to argue that the endogenous turnout played an important role. The core of Merkel's strategy was to deliberately demobilize the base of her main competitor. This strategy paid off: her party increased its margin of victory after the strategy of *asymmetric demobilization* was adopted. Moreover, it gained a relative turnout advantage: the potential voters of Merkel's party turned out in larger fractions than the potential voters of the main competitor (social-democrats). Finally, overall turnout went down. These observations are consistent with the comparative static predictions of our model for the endogenous turnout case.

A broader message is that it makes little sense to let the data speak about what society's social welfare function looks like without at the same time having an explanation for how society would come to settle on a specific social welfare function. In fact, Arrow's impossibility shows that it may be entirely infeasible to aggregate individual preferences over tax policies into a social welfare function. What is feasible, by contrast, is to reinterpret any process that selects a tax schedule from the set of Pareto-efficent tax schedules as if it was the choice of an agent with preferences over the set of Pareto-optima, see Samuelson (1967). Saez and Stantcheva (2016) recognize that observed tax policies are not entirely driven by welfare considerations, but also by non-welfarist value judgments or political economy forces. They propose to capture these forces by means of generalized social welfare functions. Thus, observed tax policies are interpreted as if there had been an agent who made a selection from the set of Pareto optima and who used a generalized social welfare function for this purpose.

We have shown that political equilibrium tax policies can be characterized by formulas akin to those for welfare-maximizing tax systems. Thus, we provide a detailed and microfounded proof of concept that political economy forces can be captured by

[22] Note that our general model with $\mu \in (0, \infty)$ has all of these channels active simultaneously.

political weights, that is, by specific generalized social welfare weights that emerge in the political process. We have also shown that political equilibrium outcomes may be incompatible with the maximization of a concave social welfare function. Generalized social welfare weights are therefore needed. "Conventional" social welfare weights cannot capture such political failures.

References

Adams, James and Samuel Merrill III, "Voter Turnout and Candidate Strategies in American Elections," *The Journal of Politics*, 2011, *65* (1), 161–189.

Aldashev, Gani, "Voter Turnout and Political Rents," *Journal of Public Economic Theory*, 2015, *17* (4), 528–552.

Alesina, A. and G.-M. Angeletos, "Fairness and Redistribution," *American Economic Review*, 2005, *95* (4), 960–980.

Bakker, Ryan, Catherine de Vries, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Vachudova, "Measuring party positions in Europe: The Chapel Hill expert survey trend file, 1999-2010.," *Party Politics*, 2015, (21.1), 143–152.

Bargain, O., M. Dolls, D. Neumann, A. Peichl, and S. Siegloch, "Tax-Benefit Systems in Europe and the US: Between Equity and Efficiency," *IZA Discussion Papers 5440*, 2011.

Bassett, William F, John P Burkett, and Louis Putterman, "Income distribution, government transfers, and the problem of unequal influence," *European Journal of Political Economy*, 1999, *15* (2), 207–228.

Bénabou, R. and J. Tirole, "Belief in a just world and redistributive politics," *Quarterly Journal of Economics*, 2006, *121* (2), 699–746.

Bernhardt, Dan, Peter Buisseret, Sinem Hidir et al., "The Race to the Base," Technical Report, University of Warwick, Department of Economics 2018.

Besley, T., *Principled Agents? The Political Economy of Good Government*, Oxford University Press, 2006.

Blundell, R., M. Brewer, P. Haan, and A. Shephard, "Optimal income taxation of lone mothers: an empirical comparison of the UK and Germany," *Economic Journal*, 2009, *119*, F101–F121.

Bourguignon, F. and A. Spadaro, "Tax-benefit Revealed Social Preferences," *Journal of Economic Inequality*, 2012, *10* (1), 75–108.

Callander, Steven and Catherine H. Wilson, "Turnout, Polarization, and Duverger's Law," *Journal of Politics*, 2007, *69* (4), 1047–1056.

Calvo, Ernesto and Maria Victoria Murillo, "Who delivers? Partisan clients in the Argentine electoral market," *American journal of political science*, 2004, *48* (4), 742–757.

Coate, Stephen and Michael Conlin, "A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence," *The American Economic Review*, 2004, *94* (5), 1476–1504.

_ , _ , and Andrea Moro, "The performance of pivotal-voter models in small-scale elections: Evidence from Texas liquor referenda," *Journal of Public Economics*, 2008, *92* (3), 582 – 596.

Coughlin, Peter and Shmuel Nitzan, "Electoral outcomes with probabilistic voting and Nash social welfare maxima," *Journal of Public Economics*, 1981, *15* (1), 113 – 121.

Cox, Gary W, "13 Swing voters, core voters, and distributive politics," *Political representation*, 2009, *342.*

_ and Mathew D McCubbins, "Electoral politics as a redistributive game," *The Journal of Politics*, 1986, *48* (2), 370–389.

Dahlberg, Matz and Eva Johansson, "On the vote-purchasing behavior of incumbent governments," *American political Science review*, 2002, *96* (1), 27–40.

Degan, Arianna and Antonio Merlo, "A Structural Model of Turnout and Voting in Multiple Elections," *Journal of the European Economic Association*, 2011, *9* (2), 209–245.

Diamond, P.A., "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 1998, *88*, 83–95.

Dixit, A. and J. Londregan, "Ideology, Tactics, and Efficiency in Redistributive Politics," *Quarterly Journal of Economics*, 1998, *113* (2), 497–529.

Downs, A., *An Economic Theory of Democracy*, New York, Harper and Row., 1957.

Erikson, Robert S and David W Romero, "Candidate equilibrium and the behavioral model of the vote," *American Political Science Review*, 1990, *84* (4), 1103–1126.

Feddersen, Timothy and Alvaro Sandroni, "A Theory of Participation in Elections," *The American Economic Review*, 2006, *96* (4), 1271–1282.

Feddersen, Timothy J., "Rational Choice Theory and the Paradox of Not Voting," *The Journal of Economic Perspectives*, 2004, *18* (1), 99–112.

Forschungsgruppe Wahlen, Mannheim, "Politbarometer Ost 2009 (Cumulated Data Set, incl. Flash)," 2013.

_ , "Politbarometer West 2009 (Cumulated Data Set, incl. Flash)," 2013.

_ , "Politbarometer 2013 (Cumulated Data Set, incl. Flash)," 2015.

_ , "Politbarometer 2017 (Cumulated Data Set)," 2018.

Harsanyi, John C., "Rule utilitarianism, rights, obligations and the theory of rational behavior," *Theory and Decision*, Jun 1980, *12* (2), 115–133.

Holmström, Bengt and Roger B Myerson, "Efficient and durable decision rules with incomplete information," *Econometrica: Journal of the Econometric Society*, 1983, pp. 1799–1819.

Jung, Matthias, "Modernisierung and asymmetrische Demobilisierung," in Karl-Rodolf Korte and Jan Schoofs, eds., *Die Bundestagswahl 2017*, Springer, 2019.

_ , Yvonne Schroth, and Andrea Wolf, "Wählerverhalten und Wahlergebnis," in Karl-Rodolf Korte, ed., *Die Bundestagswahl 2009*, Springer, 2010.

_ , _ , and _ , "Wählerverhalten und Wahlergebnis: Angela Merkels Sieg in der Mitte," in Karl-Rodolf Korte, ed., *Die Bundestagswahl 2013*, Springer, 2015.

Krupnikov, Yanna, "When does negativity demobilize? Tracing the conditional effect of negative campaigning on voter turnout," *American Journal of Political Science*, 2011, *55* (4), 797–813.

Lago, Ignacio, Jose Ramon Montero, and Mariano Torcal, "The 2006 Regional Election in Catalonia: Exit, Voice, and Electoral Market Failures," *South European Society and Politics*, 2007, *12* (2), 221–235.

Larcinese, Valentino, "Voting over redistribution and the size of the welfare state: the role of turnout," *Political Studies*, 2007, *55* (3), 568–585.

Ledyard, John O., "The Pure Theory of Large Two-Candidate Elections," *Public Choice*, 1984, *44* (1), 7–41.

Levitt, Steven D and James M Snyder, "Political parties and the distribution of federal outlays," *American Journal of Political Science*, 1995, *39*, 958–980.

Lindbeck, A. and J. Weibull, "Balanced-budget Redistribution as the Outcome of Political Competition," *Public Choice*, 1987, *52*, 273–297.

Lockwood, Benjamin B and Matthew Weinzierl, "Positive and normative judgments implicit in US tax policy, and the costs of unequal growth and recessions," *Journal of Monetary Economics*, 2016, *77*, 30–47.

Meltzer, A. and S. Richard, "A Rational Theory of the Size of Government," *Journal of Political Economy*, 1981, *89*, 914–927.

Mirrlees, J., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38*, 175–208.

Perotti, Roberto, "Growth, income distribution, and democracy: What the data say," *Journal of Economic growth*, 1996, *1* (2), 149–187.

Polk, Jonathan, Jan Rovny, Ryan Bakker, Erica Edwards, Liesbet Hooghe, Seth Jolly, Jelle Koedam, Filip Kostelka, Gary Marks, Gijs Schumacher, Marco Steenbergen, Milada Vachudova, and Marko Zilovic, "Explaining the salience of anti-elitism and reducing political corruption for political parties in

Europe with the 2014 Chapel Hill Expert Survey data," *Research & Politics*, 2017, pp. 1–9.

Riker, William H. and Peter C. Ordeshook, "A Theory of the Calculus of Voting," *The American Political Science Review*, 1968, *62* (1), 25–42.

Roberts, K., "Voting over Income Tax Schedules," *Journal of Public Economics*, 1977, *8.*

Sabet, Navid, "Turning out for redistribution: The effect of voter turnout on top marginal tax rates," Technical Report, Munich Discussion Paper 2016.

Saez, E. and S. Stantcheva, "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," *American Economic Review*, 2016, *106* (1), 24–45.

Samuelson, P., "Arrow's Mathematical Politics," in S. Hook, ed., *Human Values and Economic Policy*, New York University Press, 1967.

Schmidt, Manfred G., "The Policy of Strategic Demobilisation: The Social Policy of the Christian Democratic-Liberal Coalition, 2009-13," *German Politics*, 2014, *23* (4), 353–370.

Sheshinski, Eytan, "The optimal linear income-tax," *The Review of Economic Studies*, 1972, *39* (3), 297–302.

Stokes, Susan C, "Perverse accountability: A formal model of machine politics with evidence from Argentina," *American political science review*, 2005, *99* (3), 315–325.

Volkens, Andrea, Werner Krause, Pola Lehmann, Theres Matthieß, Nicolas Merz, Sven Regel, and Bernhard Weßels, "The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR)," 2018.

Zoutman, F.T., B. Jacobs, and E.L.W. Jongen, "Redistributive Politics and the Tyranny of the Middle Class," *Tinbergen Institute Discussion Paper 16-032/VI*, 2016.

A Proofs of Section 1

Proof of equation (6). The total number of votes for party 1 is a random variable equal to

$$\tilde{V}^1(p^1, p^2, \sigma^1, \tilde{q}^1) = \mathbb{E}[\sigma^1 \tilde{q}^1(\omega) B(u(p^1, \omega) - u(p^2, \omega) \mid \omega)] =: \sigma^1 \tilde{\mathbf{B}}^1(p^1, p^2). \tag{24}$$

Analogously, the total number of votes for party 2 equals

$$\tilde{V}^2(p^1, p^2, \sigma^2, \tilde{q}^2) = \mathbb{E}[\sigma^2 \tilde{q}^2(\omega)(1 - B(u(p^1, \omega) - u(p^2, \omega) \mid \omega))] =: \sigma^2 \tilde{\mathbf{B}}^2(p^1, p^2). \tag{25}$$

Given two party proposals p^1 and p^2, and given the turnout for party 2, σ^2, the best response problem of the rule-utilitarian supporters of party 1 is to choose σ^1 so as to maximize the expected value of the following expression

$$\mathbb{I}\left\{\tilde{V}^1(p^1, p^2, \sigma^1, \tilde{q}^1) \geq \tilde{V}^2(p^1, p^2, \sigma^2, \tilde{q}^2)\right\} \mathbb{E}\left[B(u(p^1, \omega) - u(p^2, \omega) \mid \omega) u(p^1, \omega)\right]$$

$$+ \left(1 - \mathbb{I}\left\{\tilde{V}^1(p^1, p^2, \sigma^1, \tilde{q}^1) \geq \tilde{V}^2(p^1, p^2, \sigma^2, \tilde{q}^2)\right\}\right) \times$$

$$\mathbb{E}\left[B(u(p^1, \omega) - u(p^2, \omega) \mid \omega) u(p^2, \omega) + \textstyle\int_{-\infty}^{u(p^1, \omega) - u(p^2, \omega)} \varepsilon\, b(\varepsilon \mid \omega) d\varepsilon\right]$$

$$-\ k(\sigma^1)\ \mathbb{E}[\tilde{q}^1(\omega) B(u(p^1, \omega) - u(p^2, \omega) \mid \omega)].$$

In this expression, $\mathbb{I}$ is an indicator function and the product

$$\mathbb{I}\left\{\tilde{V}^1(p^1, p^2, \sigma^1, \tilde{q}^1) \geq \tilde{V}^2(p^1, p^2, \sigma^2, \tilde{q}^2)\right\} \mathbb{E}\left[B(u(p^1, \omega) - u(p^2, \omega) \mid \omega) u(p^1, \omega)\right]$$

is utilitarian welfare realized by the supporters of party 1 in the event that their party 1 wins. Analogously,

$$(1 - \mathbb{I}\{\cdot\})\, \mathbb{E}\left[B(u(p^1, \omega) - u(p^2, \omega) \mid \omega) u(p^2, \omega) + \int_{-\infty}^{u(p^1, \omega) - u(p^2, \omega)} \varepsilon\, b(\varepsilon \mid \omega) d\varepsilon\right]$$

is utilitarian welfare realized by the supporters of party 1 in the event that party 2 wins, where the integral term in this expression is the sum of the gains (or losses) that the supporters of party 1 realize because of their idiosyncratic party preference.

Voting costs do not depend on which party wins the election. Upon exploiting the linearity of the expectations operator and dropping terms that do not depend on σ^1, we can equivalently write this optimization problem as follows: choose $\sigma^1 \in [0,1]$ to maximize

$$\pi^1(p^1, p^2, \sigma^1, \sigma^2) W^1(p^1, p^2) - k(\sigma^1) \mathbf{B}^1(p^1, p^2), \tag{26}$$

where $\pi^1(p^1, p^2, \sigma^1, \sigma^2)$ is the probability that $\tilde{V}^1(p^1, p^2, \sigma^1, \tilde{q}^1) \geq \tilde{V}^2(p^1, p^2, \sigma^2, \tilde{q}^2)$, $W^1(p^1, p^2)$ is defined by (3) and captures the welfare gain that is realized by the supporters of party 1 if their party wins, and $\mathbf{B}^1(p^1, p^2)$ is defined by (1) and captures the expected value of the mass of group-rule-utilitarian supporters of party 1. $\square$

Proof of Lemma 1. Using equations (24) and (25), we easily obtain that given p^1 and p^2, the probability that party 1 wins the election is equal to the probability of the event

$$\frac{\sigma^1}{\sigma^2} \geq \frac{\tilde{\mathbf{B}}^2(p^1, p^2)}{\tilde{\mathbf{B}}^1(p^1, p^2)}.$$

Given p^1 and p^2, denote by $\mathfrak{P}(\cdot \mid p^1, p^2)$ the c.d.f. and by $\mathfrak{p}(\cdot \mid p^1, p^2)$ the density of the random variable $\frac{\tilde{\mathbf{B}}^2(p^1,p^2)}{\tilde{\mathbf{B}}^1(p^1,p^2)}$. Thus,

$$\pi^1(p^1, p^2, \sigma^1, \sigma^2) = \mathfrak{P}\left(\frac{\sigma^1}{\sigma^2} \mid p^1, p^2\right). \tag{27}$$

We take the party platforms p^1 and p^2 as given and characterize equilibrium turnout. We say that the turnout subgame has an interior equilibrium if $0 < \sigma^{1*}(p^1, p^2) < 1$ and $0 < \sigma^{2*}(p^1, p^2) < 1$. An interior equilibrium is characterized by the first-order conditions

$$\pi^1_{\sigma^1}(\cdot) W^1(p^1, p^2) - \kappa \left(\sigma^1\right)^{1/\mu} \mathbf{B}^1(p^1, p^2) = 0, \tag{28}$$

and

$$-\pi^1_{\sigma^2}(\cdot) W^2(p^1, p^2) - \kappa \left(\sigma^2\right)^{1/\mu} \mathbf{B}^2(p^1, p^2) = 0. \tag{29}$$

Using equation (27), these first order conditions can also be written as

$$\mathfrak{p}\left(\frac{\sigma^1}{\sigma^2} \mid p^1, p^2\right) \frac{\sigma^1}{\sigma^2} \frac{1}{\sigma^1} W^1(p^1, p^2) - \kappa \left(\sigma^1\right)^{1/\mu} \mathbf{B}^1(p^1, p^2) = 0, \tag{30}$$

and

$$\mathfrak{p}\left(\frac{\sigma^1}{\sigma^2} \mid p^1, p^2\right) \frac{\sigma^1}{\sigma^2} \frac{1}{\sigma^2} W^2(p^1, p^2) - \kappa \left(\sigma^2\right)^{1/\mu} \mathbf{B}^2(p^1, p^2) = 0. \tag{31}$$

Equations (30) and (31) allow us to pin down the equilibrium value of relative turnout,

$$R^{\sigma}(p^1,p^2) := \frac{\sigma^{1*}(p^1,p^2)}{\sigma^{2*}(p^1,p^2)} = \left[\frac{W^1(p^1,p^2)/\mathbf{B}^1(p^1,p^2)}{W^2(p^1,p^2)/\mathbf{B}^2(p^1,p^2)}\right]^{\frac{\mu}{1+\mu}}, \tag{32}$$

as had to be shown. □

Proof of Lemma 2. Under Assumption 1,

$$\tilde{\mathbf{B}}^1(p^1,p^2) = \eta^1\mathbf{B}^1(p^1,p^2) \quad \text{and} \quad \tilde{\mathbf{B}}^2(p^1,p^2) = \eta^2\mathbf{B}^2(p^1,p^2).$$

The probability that party 1 wins the election is therefore equal to the probability of the event

$$\sigma^1\eta^1\mathbf{B}^1(p^1,p^2) \geq \sigma^2\eta^2\mathbf{B}^2(p^1,p^2)$$

or, equivalently,

$$\frac{\sigma^1}{\sigma^2}\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)} \geq \frac{\eta^2}{\eta^1}.$$

Let P be the c.d.f. of the random variable η^2/η^1. Then this probability can be written as

$$\bar{\pi}^1(p^1,p^2) = P\left(\frac{\sigma^1(p^1,p^2)}{\sigma^2(p^1,p^2)}\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\right).$$

Lemma 1 then implies that with an interior equilibrium of the participation subgame, the probability that party 1 wins the election is a non-decreasing function of $R^{\sigma}(p^1,p^2)R^{\mathbf{B}}(p^1,p^2)$. Therefore, party 1's objective is to maximize this expression and party 2's objective is to minimize it. □

Proof of Proposition 1. Lemma 2 implies that Party 1 seeks to maximize

$$R^{\sigma}(p^1,p^2)\frac{\mathbf{B}^2(p^1,p^2)}{\mathbf{B}^1(p^1,p^2)} = \left[\frac{W^1(p^1,p^2)}{W^2(p^1,p^2)}\right]^{\frac{\mu}{1+\mu}}\left[\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\right]^{\frac{1}{1+\mu}}$$

and that party 2 seeks to minimize this expression, where the equality follows from equation (32). This immediately leads to equation (10). □

B Proof of Lemma 3: Equilibrium existence with a one-dimensional policy space

In this section we establish equilibrium existence and uniqueness of an equilibrium for a model with a one-dimensional policy space, assuming that policy preferences are single-peaked and represented by a strictly concave utility function. We provide the following characterization: in equilibrium both parties propose the policy that maximizes the generalized social welfare function $\mathbb{E}[\gamma(\omega)u(p,\omega)]$, where the weights $\{\gamma(\omega)\}_{\omega\in\Omega}$ are determined by the equilibrium analysis below.

Setup. Let the policy space $\mathcal{P}$ be a an interval of real numbers, $\mathcal{P} = [\underline{p}, \overline{p}] \subset \mathbb{R}$. Suppose that preferences over policies are single-peaked.[23] Let $p^*(\omega) = \text{argmax}_{p\in\mathcal{P}}\, u(p,\omega)$ be the ideal policy for voter type ω. We assume, moreover, that the voters' utility functions $u(\cdot,\omega)$ are twice differentiable and concave. Moreover, we assume that ideal policies lie in the interior of $\mathcal{P}$ and satisfy the first order condition $u_1(p^*(\omega),\omega) = 0$. Assuming for simplicity that $\mu \to \infty$, the objective of party 1 is to maximize

$$\psi^1(p^1,p^2) = \frac{W^1(p^1,p^2)}{W^2(p^1,p^2)}$$

and the objective of party 2 is to minimize this expression. Focusing on this case simplifies the exposition, but as we clarify below the argument does not depend on it and extends to any value of μ. Let $\Delta u(p^1,p^2,\omega) = u(p^1,\omega) - u(p^2,\omega)$ and recall that

$$\begin{aligned} W^1(p^1,p^2) &= \mathbb{E}\left[G_W^1(\Delta u(p^1,p^2,\omega) \mid \omega)\right] \\ W^2(p^1,p^2) &= \mathbb{E}\left[G_W^2(\Delta u(p^1,p^2,\omega) \mid \omega)\right] \end{aligned}$$

where we denote by

$$\begin{aligned} G_W^1(x \mid \omega) &:= \int_{-\infty}^{x} (x-\varepsilon)\, b(\varepsilon \mid \omega) d\varepsilon \\ G_W^2(x \mid \omega) &:= \int_{x}^{\infty} (\varepsilon - x)\, b(\varepsilon \mid \omega) d\varepsilon. \end{aligned}$$

[23] As is well-known, on the assumption that all citizens participate in elections, this property implies that the median voter's preferred policy is a Condorcet winner, i.e., it wins a majority in any pairwise vote against an alternative from $\mathcal{P}$.

Note that the derivatives of the functions $G_W^1(\cdot \mid \omega)$ and $G_W^2(\cdot \mid \omega)$ are respectively given by

$$\begin{aligned} g_W^1(x \mid \omega) &:= B(x \mid \omega) \\ g_W^2(x \mid \omega) &:= -\left(1 - B(x \mid \omega)\right). \end{aligned}$$

Results. We now proceed to the proof of Lemma 3.

Lemma 5 (**Best responses exist and are interior**). *For any $p^2 \in \mathcal{P}$, there is a best response of party 1. Any best response of party 1 lies in the interior of $\mathcal{P}$ and satisfies the first order condition $\psi_1^1(p^1, p^2) = 0$, where ψ_1^1 is the partial derivative of ψ^1 with respect to p^1. Analogously, for any $p^2 \in \mathcal{P}$ there is a best response of party* 2*. Any best response of party 2 is interior and satisfies the first order condition $\psi_2^1(p^1, p^2) = 0$.*

Proof. We only prove the statements referring to the best responses of party 1. For any p^2, the function $\psi^1(\cdot, p^2)$ is continuous in p^1 and therefore attains a maximum on the compact policy space $\mathcal{P} = [\underline{p}, \overline{p}]$. The function $\psi^1(\cdot, p^2)$ is, moreover, differentiable. To prove that the maximum is interior and satisfies first-order conditions we show that, for any p^2,

$$\psi_1^1(\underline{p}, p^2) > 0 \quad \text{and} \quad \psi_1^1(\overline{p}, p^2) < 0 \, .$$

Given p^2, the derivative of $\psi^1(\cdot, p^2)$ with respect to p^1 can be written as

$$\begin{aligned} \psi_1^1(p^1, p^2) = {} & \frac{1}{W^2(p^1, p^2)} \mathbb{E}\left[g_W^1(\Delta u(\cdot) \mid \omega) u_1(p^1, \omega)\right] \\ & - \frac{W^1(p^1, p^2)}{\left(W^2(p^1, p^2)\right)^2} \mathbb{E}\left[g_W^2(\Delta u(\cdot) \mid \omega) u_1(p^1, \omega)\right] \\ = {} & \frac{1 + \psi^1(p^1, p^2)}{W^2(p^1, p^2)} \, \mathbb{E}\left[\gamma^1(\omega \mid p^1, p^2) \; u_1(p^1, \omega)\right] \end{aligned}$$

where

$$\gamma^1(\omega \mid p^1, p^2) = \frac{1}{1 + \psi^1(p^1, p^2)} \; B(\Delta u(\cdot) \mid \omega) + \frac{\psi^1(p^1, p^2)}{1 + \psi^1(p^1, p^2)} \; \left(1 - B(\Delta u(\cdot) \mid \omega)\right).$$

Let $p^1 = \underline{p}$, then $u_1(p^1, \omega) > 0$ for all ω and hence $\psi_1^1(p^1, p^2) > 0$. Analogously, if $p^1 = \overline{p}$, then $u_1(p^1, \omega) < 0$ for all ω and hence $\psi_1^1(p^1, p^2) < 0$. $\square$

Lemma 6 (**Best responses are continuous**). *For given p^2, let $p^{1*}(p^2)$ be a solution to the first order condition $\psi_1^1(p^1, p^2)$. Then p^{1*} is a continuous function.*

Proof. The implicit function theorem can be applied to the first-order condition for party 1 and implies that p^{1*} is a differentiable, and hence continuous, function of p^2. $\square$

Lemma 7 (**Existence and uniqueness of fixed point**). *The function p^{1*} has one and only one fixed point.*

Proof. The function p^{1*} is a continuous function from $\mathcal{P}$ to $\mathcal{P}$, where $\mathcal{P}$ is a non-empty, compact and convex set. Therefore, it has a fixed point by Brouwer's fixed point theorem. This proves existence. To establish uniqueness, let (p^1, p^2) be such a fixed point. It satisfies $\psi_1^1(p^1, p^2) = 0$, i.e.,

$$\mathbb{E}\left[\gamma^1(\omega \mid p^1, p^2)\ u_1(p^1, \omega)\right] = 0 \ ,$$

and

$$p^1 = p^2 \ .$$

These two equations uniquely pin down p^1. To see this, note first that $p^1 = p^2$ implies $\Delta u(p^1, p^2, \omega) = 0$ for all ω and that $\gamma^1(\omega \mid p^1, p^2)$ depends on p^1 and p^2 only via $\Delta u(p^1, p^2, \omega)$. Let $\gamma(\omega)$ be the corresponding value of $\gamma^1(\omega \mid p^1, p^2)$, i.e.,

$$\gamma(\omega) := \frac{1}{1+\psi_*^1} B(0 \mid \omega) + \frac{\psi_*^1}{1+\psi_*^1}\left(1 - B(0 \mid \omega)\right), \quad \text{with} \quad \psi_*^1 := \frac{\mathbb{E}\left[G_W^1(0 \mid \omega)\right]}{\mathbb{E}\left[G_W^2(0 \mid \omega)\right]}$$

Then p^1 solves

$$\alpha(p^1) := \mathbb{E}\left[\gamma(\omega)\ u_1(p^1, \omega)\right] = 0 \ .$$

To see that this equation has a unique solution, note that the auxiliary function $\alpha(p^1)$ is differentiable, and decreasing as $\alpha'(p^1) = \mathbb{E}\left[\gamma(\omega)\ u_{11}(p^1, \omega)\right] < 0$. Moreover, $\alpha(\underline{p}) > 0$ and $\alpha(\overline{p}) < 0$. Thus, there is one and only one solution to the equation $\alpha(p^1) = 0$. $\square$

Lemma 8. *The best response functions p^{1*} and p^{2*} have the same fixed points.*

Proof. As argued above, (p^1, p^2) is a fixed point of p^{1*} if it satisfies

$$\mathbb{E}\left[\gamma^1(\omega \mid p^1, p^2)\ u_1(p^1, \omega)\right] = 0 \tag{33}$$

and

$$p^1 = p^2 \ . \tag{34}$$

Given p^1, the best responses of party 2, $p^{2*}(p^1)$ is obtained as the solution to

$$\min_{p^2 \in \mathcal{P}} \quad \psi^1(p^1, p^2)$$

and solves the first-order condition

$$\psi^1_2(p^1, p^2) = -\frac{1 + \psi^1(p^1, p^2)}{W^2(p^1, p^2)} \, \mathbb{E}\left[\gamma^1(\omega \mid p^1, p^2) \; u_1(p^2, \omega)\right] = 0 \ .$$

Thus, a fixed point (p^1, p^2) of p^{2*} satisfies

$$\mathbb{E}\left[\gamma^1(\omega \mid p^1, p^2) \; u_1(p^2, \omega)\right] = 0 \ , \tag{35}$$

and

$$p^1 = p^2 \ . \tag{36}$$

If it satisfies these equations, then it also satisfies (33) and (34) and hence is a fixed point of p^{1*}. Conversely, if it satisfies (33) and (34), then it also satisfies (35) and (36) so that any fixed point of p^{1*} is also a fixed point p^{2*}. □

Proof of Lemma 3. We can now complete the proof of Lemma 3. The policy that maximizes $\mathbb{E}\left[\gamma(\omega) \; u(p, \omega)\right]$ satisfies the first order condition

$$\mathbb{E}\left[\gamma(\omega) \; u_1(p, \omega)\right] = 0 \ .$$

It follows from Lemmas 7 and (8) that this policy is both a fixed point of party 1's best response function and a fixed point of party 2's best response function. Thus, if both parties propose this policy they are giving mutually best responses. It remains to be shown that there can be no other equilibrium. To arrive at a contradiction, suppose that (p^1, p^2) is an alternative equilibrium. Then, p^1 is a best response to p^2 which implies that the first-order condition

$$\mathbb{E}\left[\gamma^1(\omega \mid p^1, p^2) \; u_1(p^1, \omega)\right] = 0 \tag{37}$$

is satisfied. Analogously, p^2 is a best response to p^1 and this implies that

$$\mathbb{E}\left[\gamma^1(\omega \mid p^1, p^2)\ u_1(p^2, \omega)\right] = 0\ . \tag{38}$$

If we assume that $p^1 = p^2$, these conditions imply that both parties propose the policy that maximizes $\mathbb{E}\left[\gamma(\omega)\ u(p, \omega)\right]$. Assume therefore that $p^1 \neq p^2$. But then there must be two different values of p that solve

$$\tilde{\alpha}(p) = 0$$

where the auxiliary function $\tilde{\alpha}$ is such that

$$\tilde{\alpha}(p) := \mathbb{E}\left[\gamma^1(\omega \mid p^1, p^2)\ u_1(p, \omega)\right]\ .$$

Note that

$$\tilde{\alpha}'(p) := \mathbb{E}\left[\gamma^1(\omega \mid p^1, p^2)\ u_{11}(p, \omega)\right] < 0\ ,$$

$\tilde{\alpha}(\underline{p}) > 0$ and $\tilde{\alpha}(\overline{p}) < 0$. Thus, there is one and only one value of p satisfying $\tilde{\alpha}(p) = 0$, a contradiction. $\square$

General Objective, $0 < \mu < \infty$**.** The preceding argument uses that $W^1(\cdot)$ and $W^2(\cdot)$ depend on p^1 and p^2 only via $\Delta u(p^1, p^2, \omega) = u(p^1, \omega) - u(p^2, \omega)$ and the derivatives of the objective function can be written as a weighted sum of the different types' marginal utilities, where the weights are all positive. These properties remain intact with a more general objective function of the form

$$\psi^1(p^1, p^2) = \frac{1}{1+\mu} \ln\left(\frac{\mathbf{B}^1(p^1, p^2)}{\mathbf{B}^2(p^1, p^2)}\right) + \frac{\mu}{1+\mu} \ln\left(\frac{W^1(p^1, p^2)}{W^2(p^1, p^2)}\right)\ .$$

For instance, the best response condition $\psi_1^1(p^1, p^2) = 0$ for party 1 can then be written as

$$\mathbb{E}\left[\gamma^{1,\mu}(\omega \mid p^1, p^2)\ u_1(p^1, \omega)\right] = 0\ ,$$

where

$$\gamma^{1,\mu}(\omega \mid p^1, p^2) \quad = \quad \tfrac{1}{1+\mu}\ \bar{q}(\omega)\ b(\cdot \mid \omega) + \tfrac{\mu}{1+\mu} \left(\tfrac{1}{\mathbf{B}^1(\cdot)} + \tfrac{1}{\mathbf{B}^2(\cdot)}\right)^{-1} \gamma^1(\omega \mid p^1, p^2).$$

Weakening concavity. If we drop the assumption that $u_{11}(p,\omega) < 0$, for all p and ω, but keep the assumptions $u_1(\underline{p},\omega) > 0$ and $u_1(\overline{p},\omega) > 0$, for all ω, we can no longer prove uniqueness. The arguments for the existence of the symmetric equilibrium go through, however. This observation is relevant because a generic version of the linear tax model gives us $u_1(\underline{p},\omega) > 0$ and $u_1(\overline{p},\omega) < 0$, for all ω, but not necessarily concavity. By contrast, a spatial model of political competition in which voters have bliss points and a quadratic loss function will give rise to strict concavity. In this paper, we focus on an equilibrium that exists for any generic linear tax model and this equilibrium is unique under ancillary conditions.

C Proofs of Sections 2 and 3

Proof of equation (14). The individual optimization problem reads

$$\max_{y} \; U\left(\left(1-\tau\right)y + R, y; \omega\right).$$

The first order condition is given by

$$\left(1-\tau\right)U_1\left(\left(1-\tau\right)y\left(\omega\right) + R, y\left(\omega\right);\omega\right) + U_2\left(\left(1-\tau\right)y\left(\omega\right) + R, y\left(\omega\right);\omega\right) = 0.$$

The implicit function theorem leads to

$$\frac{\partial y\left(\omega\right)}{\partial\left(1-\tau\right)} = -\frac{U_1 + \left[\left(1-\tau\right)U_{11} + U_{12}\right]y\left(\omega\right)}{\left(1-\tau\right)^2 U_{11} + 2\left(1-\tau\right)U_{12} + U_{22}}.$$

We define the uncompensated elasticity of labor supply with respect to the retention rate by

$$e\left(\omega\right) := \frac{1-\tau}{y\left(\omega\right)}\frac{\partial y\left(\omega\right)}{\partial\left(1-\tau\right)}.$$

The envelope theorem implies

$$\frac{\partial}{\partial\tau}U\left(\left(1-\tau\right)y\left(\omega\right) + R\left(\tau\right), y\left(\omega\right);\omega\right) = \left(-y\left(\omega\right) + \frac{\partial R\left(\tau\right)}{\partial\tau}\right)U_1\left(c\left(\omega\right), y\left(\omega\right);\omega\right).$$

Now, the government optimization problem reads

$$\max_{\tau,R} \; \mathbb{E}\left[g\left(\omega\right)U\left(\left(1-\tau\right)y\left(\omega\right) + R, y\left(\omega\right);\omega\right)\right],$$

subject to the budget balance constraint

$$R = \tau \mathbb{E}\left[y\left(\omega\right)\right].$$

The first-order condition of this problem is given by

$$0 = \mathbb{E}\left[g\left(\omega\right) U_1\left(c\left(\omega\right), y\left(\omega\right); \omega\right)\left(-y\left(\omega\right) + \mathbb{E}\left[y\left(\omega\right)\right] + \tau \mathbb{E}\left[\frac{\partial y\left(\omega\right)}{\partial \tau}\right]\right)\right],$$

which yields

$$\frac{\tau \mathbb{E}\left[\frac{y(\omega)}{1-\tau} e\left(\omega\right)\right]}{\mathbb{E}\left[y\left(\omega\right)\right]} = -\frac{\mathbb{E}\left[g\left(\omega\right) U_1\left(c\left(\omega\right), y\left(\omega\right); \omega\right) y\left(\omega\right)\right]}{\mathbb{E}\left[g\left(\omega\right) U_1\left(c\left(\omega\right), y\left(\omega\right); \omega\right)\right] \mathbb{E}\left[y\left(\omega\right)\right]} + 1.$$

Under the functional for (11), we have $U_1 = 1$ and $e\left(\omega\right) = e$ for all ω. We then easily obtain equation (14). □

Intuitive derivation of equation (14). Consider a small reform of the tax rate τ^* by $\delta\tau$, along with an adjustment in the rebate δR in order to ensure that the government budget remains balanced. First note that this reform $(\delta\tau, \delta R)$ induces a behavioral response of every agent: by definition of the elasticity of labor supply, the pre-tax income of type ω changes by $\delta y\left(\omega\right) = -\frac{y(\omega)}{1-\tau^*} e\left(\omega\right) \delta\tau$. Thus, the universal lump-sum rebate must be adjusted by

$$\delta R \;=\; \mathbb{E}[y\left(\omega\right)]\,\delta\tau \;+\; \tau^*\,\mathbb{E}[\delta y\left(\omega\right)] \;=\; \left\{1 \;-\; \frac{\tau^*}{1-\tau^*}\,\mathbb{E}\big[\frac{y\left(\omega\right)}{\mathbb{E}y}\,e\left(\omega\right)\big]\right\}\mathbb{E}[y\left(\omega\right)]\,\delta\tau.$$

The first term in this expression is the statutory change in revenue due to the tax change $\delta\tau$, i.e., the effect of the reform ignoring agents' responses to the policy. The second term accounts for these behavioral (substitution) effects: the actual rise in tax revenue due to a tax increase $\delta\tau > 0$ is different from the statutory effect, because agents respond by adjusting their labor supply. Next, the indirect utility of agent ω changes, in response to the tax reform $(\delta\tau, \delta R)$, by

$$\delta u\left(\tau, \omega\right) \;=\; \left\{-y\left(\omega\right)\,\delta\tau + \delta R\right\} U_1\left(c\left(\omega\right), y\left(\omega\right); \omega\right).$$

This follows from the envelope theorem: utility decreases one for one with the change in the total tax liability $(y\left(\omega\right)\delta\tau - \delta R)$ of the agent – in particular, the labor supply responses have only a first-order effect on the agent's welfare. Finally, social welfare $\mathbb{E}[g(\omega)\,u(\tau,\omega)]$ changes by $\mathbb{E}[g(\omega)\,\delta u(\tau,\omega)]$. Imposing that such a reform has a zero first-order impact on social welfare and rearranging terms easily leads to formula (14).

Proof of Proposition 2 and equations (19, 20). Suppose first that $\mu = 0$. In this case, party 1 chooses τ^1 to maximize

$$\mathbf{B}^1\left(\tau^1,\tau^2\right) = \mathbb{E}\left[\bar{q}(\omega) B\left(\Delta u\left(\tau^1,\tau^2,\omega\right) \mid \omega\right)\right]$$

subject to the budget balance constraint

$$R^1 = \tau^1 \mathbb{E}\left[y^1(\omega)\right],$$

where, as usual, we let

$$\begin{aligned} & \Delta u\left(\tau^1,\tau^2,\omega\right) \\ = \;& U\left(\left(1-\tau^1\right) y^1(\omega) + R^1, y^1(\omega);\omega\right) - U\left(\left(1-\tau^2\right) y^2(\omega) + R^2, y^2(\omega);\omega\right). \end{aligned}$$

By the envelope theorem, we have

$$\frac{\partial \Delta u\left(\tau^1,\tau^2,\omega\right)}{\partial \tau^1} = -y^1(\omega) + \mathbb{E}\left[y^1(\omega)\right] + \tau^1 \mathbb{E}\left[\frac{\partial y^1(\omega)}{\partial \tau^1}\right].$$

The first order condition to party 1's problem is then given by

$$0 = \frac{\partial \mathbf{B}^1\left(\tau^1,\tau^2\right)}{\partial \tau^1} = \mathbb{E}\left[\bar{q}(\omega) b\left(\Delta u\left(\tau^1,\tau^2,\omega\right) \mid \omega\right) \frac{\partial \Delta u\left(\tau^1,\tau^2,\omega\right)}{\partial \tau^1}\right].$$

Noting that the labor supply elasticity is given by $-\frac{1-\tau^1}{y^1(\omega)}\frac{\partial y^1(\omega)}{\partial \tau^1} = e$ and rearranging terms yields

$$\tau^1 \frac{\mathbb{E}\left[\frac{y^1(\omega) e}{1-\tau^1}\right]}{\mathbb{E}\left[y^1(\omega)\right]} = -\frac{\mathbb{E}\left[\bar{q}(\omega) b\left(\Delta u(\cdot) \mid \omega\right) y^1(\omega)\right]}{\mathbb{E}\left[\bar{q}(\omega) b\left(\Delta u(\cdot) \mid \omega\right)\right] \mathbb{E}\left[y^1(\omega)\right]} + 1.$$

This in turn leads to formula (19). Since the subgame perfect equilibrium of the political game is symmetric, we moreover have $\Delta u\left(\tau^1,\tau^2,\omega\right) = 0$ for all ω. This leads to formula (16).

Suppose next that $\mu \to \infty$. In this case, party 1's objective is

$$\psi^1\left(\tau^1,\tau^2\right) = \frac{W^1\left(\tau^1,\tau^2\right)}{W^2\left(\tau^1,\tau^2\right)} = \frac{\mathbb{E}\left[\int \left(\Delta u\left(\cdot\right) - \varepsilon\right) \mathbb{I}_{\{\varepsilon \leq \Delta u(\cdot)\}} b(\varepsilon \mid \omega) d\varepsilon\right]}{\mathbb{E}\left[\int \left(\varepsilon - \Delta u\left(\cdot\right)\right) \mathbb{I}_{\{\varepsilon \geq \Delta u(\cdot)\}} b(\varepsilon \mid \omega) d\varepsilon\right]}.$$

We have

$$\begin{aligned}
\frac{\partial W^1\left(\tau^1,\tau^2\right)}{\partial \tau^1} &= \frac{\partial}{\partial \tau^1}\mathbb{E}\left[\left(\Delta u\left(\cdot\right)\right) B\left(\Delta u\left(\cdot\right) \mid \omega\right) - \int_{-\infty}^{\Delta u(\cdot)} \varepsilon b(\varepsilon \mid \omega) d\varepsilon\right] \\
&= \mathbb{E}\left[B\left(\Delta u\left(\cdot\right) \mid \omega\right) \frac{\partial\left(\Delta u\left(\cdot\right)\right)}{\partial \tau^1}\right]
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial W^2\left(\tau^1,\tau^2\right)}{\partial \tau^1} &= \frac{\partial}{\partial \tau^1}\mathbb{E}\left[\int_{\Delta u(\cdot)}^{\infty} \varepsilon b(\varepsilon \mid \omega) d\varepsilon - \left(\Delta u\left(\cdot\right)\right)\left(1 - B\left(\Delta u\left(\cdot\right) \mid \omega\right)\right)\right] \\
&= -\mathbb{E}\left[\left(1 - B\left(\Delta u\left(\cdot\right) \mid \omega\right)\right) \frac{\partial\left(\Delta u\left(\cdot\right)\right)}{\partial \tau^1}\right].
\end{aligned}$$

Next, we have

$$\begin{aligned}
&\frac{\partial}{\partial \tau^1}\left\{\frac{W^1\left(\tau^1,\tau^2\right)}{W^2\left(\tau^1,\tau^2\right)}\right\} = \frac{1}{W^2\left(\tau^1,\tau^2\right)}\left\{\frac{\partial W^1\left(\tau^1,\tau^2\right)}{\partial \tau^1} - \frac{W^1\left(\tau^1,\tau^2\right)}{W^2\left(\tau^1,\tau^2\right)}\frac{\partial W^2\left(\tau^1,\tau^2\right)}{\partial \tau^1}\right\} \\
= &\frac{1+\psi^1\left(\tau^1,\tau^2\right)}{W^2\left(\tau^1,\tau^2\right)} \times \\
&\mathbb{E}\left[\left\{\frac{1}{1+\psi^1\left(\tau^1,\tau^2\right)} B\left(\Delta u \mid \omega\right) + \frac{\psi^1\left(\tau^1,\tau^2\right)}{1+\psi^1\left(\tau^1,\tau^2\right)}\left(1 - B\left(\Delta u \mid \omega\right)\right)\right\} \frac{\partial\left(\Delta u\left(\cdot\right)\right)}{\partial \tau^1}\right].
\end{aligned}$$

Following the same steps as in the case $\mu = 0$ shows that the best response of party 1 is given by equation (20). In the symmetric equilibrium, we have $\tau^1 = \tau^2$ and hence $\Delta u\left(\cdot\right) = 0$ and

$$\frac{W^1\left(\tau^1,\tau^2\right)}{W^2\left(\tau^1,\tau^2\right)} = \frac{\mathbb{E}\left[\int_{-\infty}^{0} \left(-\varepsilon\right) b(\varepsilon \mid \omega) d\varepsilon\right]}{\mathbb{E}\left[\int_{0}^{\infty} \varepsilon b(\varepsilon \mid \omega) d\varepsilon\right]} =: \psi^1_*.$$

This in turn leads to formula (17). □

Corollary 3 (**Strategic complementarities**). *In the affine tax policy space, suppose that turnout is exogenous ($\mu = 0$) and that $B\left(\cdot \mid \omega\right)$ is uniform for all ω. Then the two parties' policies are strategically neutral (neither substitutes nor complements). Suppose next that turnout is endogenous ($\mu \to \infty$) and that the status quo is the SPE.*

In response to a budget-neutral increase in party 2's tax rate, party 1's best response is to increase its tax rate (i.e., the policies are strategic complements) if and only if $\psi_*^1 < 1$, *that is, iff the intensity of political preferences among party 1's supporters is smaller than among party 2's supporters.*

Proof of Corollary 3. Suppose first that turnout is exogenous, i.e., $\mu = 0$. Using the budget balance requirement $R^1 = \tau^1 \mathbb{E}\left[y^1(\omega)\right]$, we get

$$\frac{\partial \mathbf{B}^1(\tau^1,\tau^2)}{\partial \tau^1} = -\mathbb{E}\left[\bar{q}(\omega)\, b\left(u\left(\tau^1,\omega\right) - u\left(\tau^2,\omega\right) \mid \omega\right) Y^1(\omega)\right]$$

where we let

$$Y^j(\omega) := y^j(\omega) - \mathbb{E}y^j + \frac{\tau^j}{1-\tau^j} e\mathbb{E}\left[y^j(\omega)\right].$$

In response to a deviation by party 2, $\frac{\partial \mathbf{B}^1(\tau^1,\tau^2)}{\partial \tau^1}$ changes by:

$$\begin{aligned}
&\frac{\partial}{\partial \tau^2}\left\{\frac{\partial \mathbf{B}^1(\tau^1,\tau^2)}{\partial \tau^1}\right\} \\
&= \mathbb{E}\left[\bar{q}(\omega)\, b'\left(u\left(\tau^1,\omega\right) - u\left(\tau^2,\omega\right) \mid \omega\right) Y^1(\omega) \frac{\partial u\left(\tau^2,\omega\right)}{\partial \tau^2}\right] \\
&= -\mathbb{E}\left[\bar{q}(\omega)\, b'\left(u\left(\tau^1,\omega\right) - u\left(\tau^2,\omega\right) \mid \omega\right) Y^1(\omega)\, Y^2(\omega)\right].
\end{aligned}$$

If B is uniform, we have $b'(\cdot) = 0$, and hence

$$\frac{\partial}{\partial \tau^2}\left\{\frac{\partial \mathbf{B}^1(\tau^1,\tau^2)}{\partial \tau^1}\right\} = 0.$$

Therefore party 1 does not have an incentive to raise or lower its tax rate, as this would have no effect on its base $\mathbf{B}^1(\tau^1,\tau^2)$. Therefore, the policies of parties 1 and 2 are strategically neutral.

Now suppose that turnout is endogenous with $\mu \to \infty$. We have

$$\frac{\partial \psi^1(\tau^1,\tau^2)}{\partial \tau^1} = -\mathbb{E}\left[\gamma^1\left(\omega \mid \tau^1,\tau^2\right) Y^1(\omega)\right]$$

where

$$\gamma^1\left(\omega \mid \tau^1,\tau^2\right) = \frac{1}{1+\psi^1(\tau^1,\tau^2)} B\left(u\left(\tau^1,\omega\right)-u\left(\tau^2,\omega\right) \mid \omega\right) + \frac{\psi^1(\tau^1,\tau^2)}{1+\psi^1(\tau^1,\tau^2)}\left[1-B\left(u\left(\tau^1,\omega\right)-u\left(\tau^2,\omega\right) \mid \omega\right)\right].$$

In response to a deviation by party 2, $\frac{\partial\psi^1(\tau^1,\tau^2)}{\partial\tau^1}$ changes by:

$$\begin{aligned}
&\frac{\partial}{\partial\tau^2}\left\{\frac{\partial\psi^1(\tau^1,\tau^2)}{\partial\tau^1}\right\} \\
=\;& \mathbb{E}\left[\left(\psi^1(\tau^1,\tau^2)-1\right) b\left(u\left(\tau^1,\omega\right)-u\left(\tau^2,\omega\right) \mid \omega\right) Y^1(\omega) \frac{\partial u\left(\tau^2,\omega\right)}{\partial\tau^2}\right] \\
=\;& -\mathbb{E}\left[\left(\psi^1(\tau^1,\tau^2)-1\right) b\left(u\left(\tau^1,\omega\right)-u\left(\tau^2,\omega\right) \mid \omega\right) Y^1(\omega)\, Y^2(\omega)\right]
\end{aligned}$$

where we used the fact that $\frac{\partial\psi^1(\tau^1,\tau^2)}{\partial\tau^1}=0$ if the status quo is the SPE. Since the SPE is symmetric, we have $\tau^1=\tau^2$ and $y^1(\omega)=y^2(\omega)$ for all ω, so that

$$\frac{\partial}{\partial\tau^2}\left\{\frac{\partial\psi^1(\tau^1,\tau^2)}{\partial\tau^1}\right\} = \left(1-\psi^1_*\right)\mathbb{E}\left[b\left(0 \mid \omega\right)\left(Y^1(\omega)\right)^2\right].$$

This expression is strictly positive if and only if $\psi^1_*<1$. In this case, if party 2's tax rate goes up, party 1 has an incentive to increase its own tax rate, as we would then have $\frac{\partial\psi^1(\tau^1,\tau^2)}{\partial\tau^1}>0$. Thus the two parties' policies are complements. Conversely, they are substitutes if $\psi^1_*>1$. □

Proof of Corollary 1. Using Lemmas 1 and 2 and the first order conditions of the optimization problems in (26), we can write

$$\sigma^{1*}(p^1,p^2) = \frac{W^1(p^1,p^2)}{\kappa\mathbf{B}^1(p^1,p^2)} R^\sigma(p^1,p^2)\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\rho\left(R^\sigma(p^1,p^2)\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\right) \tag{39}$$

with

$$R^\sigma(p^1,p^2) = \frac{W^1(p^1,p^2)/\mathbf{B}^1(p^1,p^2)}{W^2(p^1,p^2)/\mathbf{B}^2(p^1,p^2)}. \tag{40}$$

Substituting (40) into (39) yields equation (21). □

Proof of Corollary 2. Suppose that Assumptions 1, 2 and 3, and Conditions (i)-(iv) in Section 3.3 are satisfied. Suppose moreover that $\mu\to\infty$.

Party 1's probability of winning. Consider first an increase in party 2's tax rate on its probability of winning. We saw that

$$\frac{\partial\psi^1(\tau^1,\tau^2)}{\partial\tau^2} \quad = \quad \frac{1+\psi^1\left(\tau^1,\tau^2\right)}{W^2\left(\tau^1,\tau^2\right)}\mathbb{E}\left[\gamma^1\left(\omega \mid \tau^1,\tau^2\right)\frac{\partial\left(\Delta u\left(\cdot\right)\right)}{\partial\tau^1}\right]$$

where

$$\frac{\partial\Delta u\left(\tau^1,\tau^2,\omega\right)}{\partial\tau^2} = -\left\{\left(1-\frac{\tau^2}{1-\tau^2}\,e\right)\mathbb{E}\left[y^2(\omega)\right]-y^2(\omega)\right\}$$

This deviation raises party 2's probability of winning the election if $\frac{\partial\psi^1(\tau^1,\tau^2)}{\partial\tau^2}<0$, which holds if:

$$\frac{\tau^2}{1-\tau^2} < -\frac{1}{e}\,\mathbb{C}\mathrm{ov}\left(\frac{\gamma^1\left(\omega \mid \tau^1,\tau^2\right)}{\mathbb{E}\left[\gamma^1\left(\omega \mid \tau^1,\tau^2\right)\right]};\frac{y^2(\omega)}{\mathbb{E}[y^2(\omega)]}\right).$$

This inequality holds for τ^2 sufficiently small since the covariance is bounded, and, by assumption (ii), negative. Alternatively, suppose that $\tau^2<\tau^{*2}\left(\tau^1\right)$, where $\tau^{*2}\left(\tau^1\right)$ denotes party 2's best response to party 1's platform τ^1. Since the second-order conditions are satisfied in equilibrium, there exists $\epsilon>0$ such that the function $\tau^2\mapsto\frac{\partial\psi^1(\tau^1,\tau^2)}{\partial\tau^2}$ is increasing, and hence negative, on $[\tau^{*2}\left(\tau^1\right)-\epsilon,\tau^{*2}\left(\tau^1\right))$. This implies that a slight increase in party 2's tax rate strictly raises its probability of winning (i.e., lowers $\psi^1(\tau^1,\tau^2)$).

Party 2's vote share. Next, consider the impact of the deviation on party 2's vote share. The total number of votes for party j is equal to

$$\tilde{V}^j(p^1,p^2) \quad = \quad \sigma^{j*}(p^1,p^2)\tilde{\mathbf{B}}^j\left(p^1,p^2\right).$$

Hence party 2's expected vote share is equal to

$$\begin{aligned}\mathcal{S}^2 \quad &:= \quad \tilde{\mathbb{E}}\left[\frac{\sigma^{2*}\tilde{\mathbf{B}}^2}{\sigma^{1*}\tilde{\mathbf{B}}^1+\sigma^{2*}\tilde{\mathbf{B}}^2}\right] \quad = \quad \tilde{\mathbb{E}}\left[\left(1+\frac{\sigma^{1*}}{\sigma^{2*}}\frac{\eta^1\mathbf{B}^1}{\eta^2\mathbf{B}^2}\right)^{-1}\right]\\ &= \quad \tilde{\mathbb{E}}\left[\left(1+\frac{W^1}{W^2}\frac{\eta^1}{\eta^2}\right)^{-1}\right] \quad = \quad \int\frac{1}{1+\psi^1\left(\tau^1,\tau^2\right)\frac{1}{x}}\rho\left(x\right)dx.\end{aligned}$$

The derivative of party 2's vote share with respect to τ^2 is given by

$$\frac{\partial \mathcal{S}^2}{\partial \tau^2} = -\left(\int \frac{x}{\left[\psi^1\left(\tau^1,\tau^2\right)+x\right]^2}\rho\left(x\right)dx\right)\frac{\partial \psi^1\left(\tau^1,\tau^2\right)}{\partial \tau^2}$$

Therefore, since the deviation raises party 2's probability of winning, so that $\frac{\partial \psi^1(\tau^1,\tau^2)}{\partial \tau^2} < 0$, it also increases its expected vote share, $\frac{\partial \mathcal{S}^2}{\partial \tau^2} > 0$.

Overall turnout. Next, expected overall turnout is equal to

$$\begin{aligned}
\Sigma \quad &:= \quad \tilde{\mathbb{E}}\left[\frac{\sigma^{1*}(\tau^1,\tau^2)\tilde{\mathbf{B}}^1(\tau^1,\tau^2)+\sigma^{2*}(\tau^1,\tau^2)\tilde{\mathbf{B}}^2(\tau^1,\tau^2)}{\tilde{\mathbf{B}}^1(\tau^1,\tau^2)+\tilde{\mathbf{B}}^2(\tau^1,\tau^2)}\right] \\
&= \quad \frac{\sigma^{1*}(\tau^1,\tau^2)\mathbf{B}^1(\tau^1,\tau^2)+\sigma^{2*}(\tau^1,\tau^2)\mathbf{B}^2(\tau^1,\tau^2)}{\mathbb{E}\left[\bar{q}(\omega)\right]} \\
&= \quad \frac{\sigma^{1*}(\tau^1,\tau^2)\mathbf{B}^1(\tau^1,\tau^2)}{\mathbb{E}\left[\bar{q}(\omega)\right]}\left(1+\frac{W^2(\tau^1,\tau^2)}{W^1(\tau^1,\tau^2)}\right) \\
&= \quad \frac{W^1(\tau^1,\tau^2)}{\kappa\mathbb{E}\left[\bar{q}(\omega)\right]}\left(1+\psi^1(\tau^1,\tau^2)\right)\rho(\psi^1(\tau^1,\tau^2)),
\end{aligned}$$

where the last equality follows from (21). We now show that $\partial\Sigma/\partial\tau^2 < 0$, i.e., overall turnout decreases in response to party 2's deviation. Since all the terms in the expression for Σ are positive, the result follows if both $\rho(\psi^1(\tau^1,\tau^2))$ and $W^1(\tau^1,\tau^2)\left(1+\psi^1(\tau^1,\tau^2)\right)$ are decreasing in τ^2.

First, Assumption 3 allows us to compute the distribution of η^2/η^1 as follows. The joint distribution of (η^1,η^2) has the following density:

$$p_{\eta^1,\eta^2}\left(x^1,x^2\right) = \frac{1}{(\bar{\eta}-\underline{\eta})^2}\mathbb{I}_{\left\{(x^1,x^2)\in[\underline{\eta},\bar{\eta}]\times[\underline{\eta},\bar{\eta}]\right\}},$$

where $\underline{\eta} := 1-\delta$ and $\bar{\eta} := 1+\delta$. The random variable η^1/η^2 is non-zero on $[\underline{\eta}/\bar{\eta},\bar{\eta}/\underline{\eta}]$ and its density is given, for any $\psi \in [\underline{\eta}/\bar{\eta},\bar{\eta}/\underline{\eta}]$, by

$$\rho\left(\psi\right) = \int_{\underline{\eta}}^{\bar{\eta}} x^1 p_{\eta^1,\eta^2}\left(x^1,\psi x^1\right)dx^1 = \frac{1}{(\bar{\eta}-\underline{\eta})^2}\int x^1\mathbb{I}_{\left\{x^1\in[\underline{\eta},\bar{\eta}]\right\}}\mathbb{I}_{\left\{x^1\in[\underline{\eta}/\psi,\bar{\eta}/\psi]\right\}}dx^1.$$

Note that if $\psi < 1$, we have $\underline{\eta}/\psi > \underline{\eta}$ and $\bar{\eta}/\psi > \bar{\eta}$, so that $[\underline{\eta}/\psi,\bar{\eta}/\psi]\cap[\underline{\eta},\bar{\eta}] = [\underline{\eta}/\psi,\bar{\eta}]$.

Conversely, if $\psi > 1$, we have $[\underline{\eta}/\psi, \bar{\eta}/\psi] \cap [\underline{\eta}, \bar{\eta}] = [\underline{\eta}, \bar{\eta}/\psi]$. Therefore we have

$$\rho(\psi) = \frac{1}{(\bar{\eta} - \underline{\eta})^2} \begin{cases} \int_{\underline{\eta}/\psi}^{\bar{\eta}} x^1 dx^1 & \text{if } \psi < 1, \\ \int_{\underline{\eta}}^{\bar{\eta}/\psi} x^1 dx^1 & \text{if } \psi > 1. \end{cases}$$

We easily obtain that, for any $\psi < 1$,

$$\rho(\psi) = \frac{\bar{\eta}^1 - \underline{\eta}^1/\psi^2}{2(\bar{\eta} - \underline{\eta})^2},$$

which is increasing in ψ. Moreover, letting P denote the cdf of η^2/η^1 and summing over $[\underline{\eta}/\bar{\eta}, 1]$ implies

$$P(1) = 1/2.$$

Now recall that party 1's probability of winning decreases in response to party 2's deviation, i.e., $\partial\psi^1(\tau^1, \tau^2)/\partial\tau^2 < 0$. Since party 2 is the front runner by assumption (iv), party 1's probability of winning is given by $\bar{\pi}^1 = P(\psi^1(\tau^1, \tau^2)) < 1/2$, and therefore we have $\psi^1(\tau^1, \tau^2) < 1$. As a result, $\rho(\cdot)$ is locally increasing around the status quo, and we have $\partial\rho(\psi^1(\tau^1, \tau^2))/\partial\tau^2 < 0$ in response to party 2's deviation.

Next, we show that the term $W^1(\tau^1, \tau^2)(1 + \psi^1(\tau^1, \tau^2))$ is decreasing. We have

$$\begin{aligned} & \frac{\partial\left(W^1(\tau^1, \tau^2)\left(1 + \psi^1(\tau^1, \tau^2)\right)\right)}{\partial\tau^2} \\ = \; & \left(1 + \psi^1(\tau^1, \tau^2)\right) \frac{\partial W^1(\tau^1, \tau^2)}{\partial\tau^2} \\ & + W^1(\tau^1, \tau^2) \left\{ \frac{1}{W^2(\tau^1, \tau^2)} \frac{\partial W^1}{\partial\tau^2} - \frac{W^1(\tau^1, \tau^2)}{[W^2(\tau^1, \tau^2)]^2} \frac{\partial W^2}{\partial\tau^2} \right\} \\ = \; & \left(1 + 2\psi^1(\tau^1, \tau^2)\right) \frac{\partial W^1(\tau^1, \tau^2)}{\partial\tau^2} - \left(\psi^1(\tau^1, \tau^2)\right)^2 \frac{\partial W^2(\tau^1, \tau^2)}{\partial\tau^2} \end{aligned}$$

and recall that

$$\begin{aligned} \frac{\partial W^1(\tau^1, \tau^2)}{\partial\tau^2} &= \mathbb{E}\left[B\left(u\left(\tau^1, \omega\right) - u\left(\tau^2, \omega\right) \mid \omega\right) \frac{\partial\left(\Delta u(\cdot)\right)}{\partial\tau^2} \right] \\ \frac{\partial W^2(\tau^1, \tau^2)}{\partial\tau^2} &= -\mathbb{E}\left[\left(1 - B\left(u\left(\tau^1, \omega\right) - u\left(\tau^2, \omega\right) \mid \omega\right)\right) \frac{\partial\left(\Delta u(\cdot)\right)}{\partial\tau^2} \right] \end{aligned}$$

where

$$\frac{\partial\left(\Delta u\left(\cdot\right)\right)}{\partial\tau^{2}} \quad = \quad y^{2}\left(\omega\right)-\mathbb{E}y^{2}+\frac{\tau^{2}}{1-\tau^{2}}e\mathbb{E}\left[y^{2}\left(\omega\right)\right]:=Y^{2}\left(\omega\right).$$

Now note that overall turnout decreases as soon as

$$\frac{\partial\left(W^{1}(\tau^{1},\tau^{2})\left(1+\psi^{1}(\tau^{1},\tau^{2})\right)\right)}{\partial\tau^{2}}<0$$

which requires that

$$\mathbb{E}\left[\frac{\tilde{\gamma}\left(\omega\mid\tau^{1},\tau^{2}\right)}{\mathbb{E}\left[\tilde{\gamma}\right]}\left(y^{2}\left(\omega\right)-\mathbb{E}y^{2}+\frac{\tau^{2}}{1-\tau^{2}}e\mathbb{E}\left[y^{2}\left(\omega\right)\right]\right)\right]<0$$

or equivalently

$$\frac{\tau^{2}}{1-\tau^{2}} \quad < \quad -\frac{1}{e}\mathbb{C}\mathrm{ov}\left(\frac{\tilde{\gamma}\left(\omega\mid\tau^{1},\tau^{2}\right)}{\mathbb{E}[\tilde{\gamma}]},\frac{y^{2}\left(\omega\right)}{\mathbb{E}y^{2}}\right),$$

where

$$\begin{aligned}\tilde{\gamma}\left(\omega\mid\tau^{1},\tau^{2}\right) \quad = \quad & \left(2+\frac{1}{\psi^{1}(\tau^{1},\tau^{2})}\right)\frac{1}{\psi^{1}(\tau^{1},\tau^{2})}B\left(u\left(\tau^{1},\omega\right)-u\left(\tau^{2},\omega\right)\mid\omega\right)\\ & +\left(1-B\left(u\left(\tau^{1},\omega\right)-u\left(\tau^{2},\omega\right)\mid\omega\right)\right).\end{aligned}$$

But recall that party 2's probability of winning increases when τ^2 increases. We saw above that this implies

$$\mathbb{E}\left[\frac{\gamma^{1}\left(\omega\mid\tau^{1},\tau^{2}\right)}{\mathbb{E}\left[\gamma^{1}\right]}\left(y^{2}\left(\omega\right)-\mathbb{E}\left[y^{2}\left(\omega\right)\right]+\frac{\tau^{2}}{1-\tau^{2}}e\mathbb{E}\left[y^{2}\left(\omega\right)\right]\right)\right]<0$$

or equivalently

$$\frac{\tau^{2}}{1-\tau^{2}} \quad < \quad -\frac{1}{e}\mathbb{C}\mathrm{ov}\left(\frac{\gamma^{1}\left(\omega\mid\tau^{1},\tau^{2}\right)}{\mathbb{E}\left[\gamma^{1}\right]},\frac{y^{2}\left(\omega\right)}{\mathbb{E}y^{2}}\right),$$

where

$$\gamma^{1}\left(\omega\mid\tau^{1},\tau^{2}\right) \quad = \quad \frac{1}{1+\psi^{1}(\tau^{1},\tau^{2})}B\left(\Delta u\left(\cdot\right)\mid\omega\right)+\frac{\psi^{1}(\tau^{1},\tau^{2})}{1+\psi^{1}(\tau^{1},\tau^{2})}\left(1-B\left(\Delta u\left(\cdot\right)\mid\omega\right)\right).$$

Since the previous expectation is negative, the tax rate τ^2 is not Pareto dominated, i.e., $\frac{\tau^2}{1-\tau^2} < 1/e$. Hence the term $Y^2(\omega)$ (defined above) is negative for low values of ω and monotonically increasing in ω. Now, notice that $\tilde{\gamma}(\omega \mid \tau^1, \tau^2)/\mathbb{E}[\tilde{\gamma}]$ puts relatively more weight on low incomes (hence on negative values of $Y(\omega)$), and less on high incomes, than $\gamma^1(\omega \mid \tau^1, \tau^2)/\mathbb{E}[\gamma^1]$. Indeed, $\tilde{\gamma}(\omega \mid \tau^1, \tau^2)$ puts a larger (resp., smaller) relative weight than $\gamma^1(\omega \mid \tau^1, \tau^2)$ on $B(\Delta u(\cdot) \mid \omega)$ (respectively, on $(1 - B(\Delta u(\cdot) \mid \omega))$), which is decreasing (resp., increasing) in ω and has therefore a negative (resp., positive) covariance with incomes $y^2(\omega)$. As a result, we have

$$
\begin{aligned}
& \mathbb{E}\left[\frac{\tilde{\gamma}(\omega \mid \tau^1, \tau^2)}{\mathbb{E}[\tilde{\gamma}]}\left(y^2(\omega) - \mathbb{E}y^2 + \frac{\tau^2}{1-\tau^2} e\mathbb{E}\left[y^2(\omega)\right]\right)\right] \\
< \; & \mathbb{E}\left[\frac{\gamma^1(\omega \mid \tau^1, \tau^2)}{\mathbb{E}[\gamma^1]}\left(y^2(\omega) - \mathbb{E}\left[y^2(\omega)\right] + \frac{\tau^2}{1-\tau^2} e\mathbb{E}\left[y^2(\omega)\right]\right)\right] < 0.
\end{aligned}
$$

Therefore, overall turnout Σ decreases in response to party 2's deviation.

Relative turnout rates. Finally, we show that the relative turnout σ^{1*}/σ^{2*} also decreases. We have

$$
\frac{\sigma^{1*}(\tau^1, \tau^2)}{\sigma^{2*}(\tau^1, \tau^2)} = \psi^1(\tau^1, \tau^2)\frac{\mathbb{E}\left[\bar{q}(\omega)\right] - \mathbf{B}^1(\tau^1, \tau^2)}{\mathbf{B}^1(\tau^1, \tau^2)}.
$$

In response to party 2's deviation, we have

$$
\begin{aligned}
\frac{\partial \mathbf{B}^1(\tau^1, \tau^2)}{\partial \tau^2} \; &= \; -\mathbb{E}\left[\bar{q}(\omega)\, b\left(u\left(\tau^1, \omega\right) - u\left(\tau^2, \omega\right) \mid \omega\right) \frac{\partial u(\tau^2, \omega)}{\partial \tau^2}\right] \\
= \; & \mathbb{E}\left[\bar{q}(\omega)\, b\left(u\left(\tau^1, \omega\right) - u\left(\tau^2, \omega\right) \mid \omega\right)\left(y^2(\omega) - \mathbb{E}y^2 + \frac{\tau^2}{1-\tau^2} e\mathbb{E}\left[y^2(\omega)\right]\right)\right].
\end{aligned}
$$

We thus have

$$
\begin{aligned}
& \frac{\partial}{\partial \tau^2}\frac{\sigma^{1*}(\tau^1, \tau^2)}{\sigma^{2*}(\tau^1, \tau^2)} \\
= \; & \frac{\mathbb{E}\bar{q} - \mathbf{B}^1(\tau^1, \tau^2)}{\mathbf{B}^1(\tau^1, \tau^2)}\frac{\partial \psi^1(\tau^1, \tau^2)}{\partial \tau^2} \\
& - \left[\frac{\psi^1(\tau^1, \tau^2)}{\mathbf{B}^1(\tau^1, \tau^2)} + \psi^1(\tau^1, \tau^2)\frac{\mathbb{E}\bar{q} - \mathbf{B}^1(\tau^1, \tau^2)}{(\mathbf{B}^1(\tau^1, \tau^2))^2}\right]\frac{\partial \mathbf{B}^1(\tau^1, \tau^2)}{\partial \tau^2} \\
= \; & \mathbb{E}\left[\tilde{\tilde{\gamma}}\left(\omega \mid \tau^1, \tau^2\right)\left(y^2(\omega) - \mathbb{E}y^2 + \frac{\tau^2}{1-\tau^2} e\mathbb{E}\left[y^2(\omega)\right]\right)\right]
\end{aligned}
$$

where

$$\begin{aligned}\tilde{\tilde{\gamma}}\left(\omega \mid \tau^1,\tau^2\right) &= B\left(\Delta u\left(\cdot\right) \mid \omega\right)+\psi^1(\tau^1,\tau^2)\left(1-B\left(\Delta u\left(\cdot\right) \mid \omega\right)\right) \\ & \quad -\frac{W^1\left(\tau^1,\tau^2\right)}{\mathbf{B}^1(\tau^1,\tau^2)}\left(1+\frac{\mathbf{B}^1(\tau^1,\tau^2)}{\mathbf{B}^2(\tau^1,\tau^2)}\right)b\left(\Delta u\left(\cdot\right) \mid \omega\right).\end{aligned}$$

Hence $\frac{\partial}{\partial\tau^2}\frac{\sigma^{1*}(\tau^1,\tau^2)}{\sigma^{2*}(\tau^1,\tau^2)}<0$ if

$$\mathbb{E}\left[\frac{\tilde{\tilde{\gamma}}\left(\omega \mid \tau^1,\tau^2\right)}{\mathbb{E}[\tilde{\tilde{\gamma}}]}\left(y^2\left(\omega\right)-\mathbb{E}y^2+\frac{\tau^2}{1-\tau^2}e\mathbb{E}\left[y^2\left(\omega\right)\right]\right)\right]<0.$$

Now, Assumption 2 implies that $b\left(\Delta u\left(\cdot\right) \mid \omega\right)$ is a constant, so that

$$\tilde{\tilde{\gamma}}\left(\omega \mid \tau^1,\tau^2\right)=\gamma^1\left(\omega \mid \tau^1,\tau^2\right)-A,$$

where $1>0$ is a constant. Throughout we assume that the weights $\tilde{\tilde{\gamma}}\left(\tau^1,\tau^2 \mid \omega\right)$ are positive, which is ensured by assuming that for all ω we have $\mathfrak{b}\left(\omega\right)\leq\beta$, for some $\beta>0$ small enough, as we also assume in Lemma 4. Moreover, note that the weights $\gamma^1\left(\tau^1,\tau^2 \mid \omega\right)$ are decreasing in ω by Condition (i) in Corollary 2 and the fact, proved above, that $\psi^1(\tau^1,\tau^2)<1$. As a consequence, for any $\omega'>\omega$, we have

$$\frac{\tilde{\tilde{\gamma}}\left(\omega \mid \tau^1,\tau^2\right)}{\tilde{\tilde{\gamma}}\left(\omega' \mid \tau^1,\tau^2\right)}=\frac{\gamma^1\left(\omega \mid \tau^1,\tau^2\right)-A}{\gamma^1\left(\omega' \mid \tau^1,\tau^2\right)-A}>\frac{\gamma^1\left(\omega \mid \tau^1,\tau^2\right)}{\gamma^1\left(\omega' \mid \tau^1,\tau^2\right)}$$

where the inequality follows from $\gamma^1\left(\omega \mid \tau^1,\tau^2\right)>\gamma^1\left(\omega' \mid \tau^1,\tau^2\right)$. Therefore, the relative weight that $\tilde{\tilde{\gamma}}\left(\omega \mid \tau^1,\tau^2\right)$ puts on low (resp., high) incomes is thus larger (resp., smaller) than the relative weight that $\gamma^1\left(\omega \mid \tau^1,\tau^2\right)$ puts on them. Since $Y^2\left(\omega\right)$ is negative for low values of ω and monotonically increasing in ω, we have

$$\begin{aligned}& \mathbb{E}\left[\frac{\tilde{\tilde{\gamma}}\left(\omega \mid \tau^1,\tau^2\right)}{\mathbb{E}[\tilde{\tilde{\gamma}}]}\left(y^2\left(\omega\right)-\mathbb{E}y^2+\frac{\tau^2}{1-\tau^2}e\mathbb{E}\left[y^2\left(\omega\right)\right]\right)\right] \\ < \; & \mathbb{E}\left[\frac{\gamma^1\left(\omega \mid \tau^1,\tau^2\right)}{\mathbb{E}\left[\gamma^1\right]}\left(y^2\left(\omega\right)-\mathbb{E}y^2+\frac{\tau^2}{1-\tau^2}e\mathbb{E}\left[y^2\left(\omega\right)\right]\right)\right]<0\end{aligned}$$

where the second inequality follows from the fact that the deviation raises party 2's probability of winning. This concludes the proof that the relative turnout decreases in τ^2. $\square$

D Proofs of Section 4

In this section, we provide technical elements regarding the environment of Section 4 and prove Proposition 3. We postpone the analysis of equilibrium existence to Section E below.

D.1 Preliminaries: On the Pareto-frontier

In this preliminary section we set up the mechanism design approach to obtain a characterization of admissible tax systems. Suppose that the preferences of a type ω individual over (c, y)-pairs are represented by a utility function $c - v(y, \omega)$, with $v_1 > 0$, $v_{11} > 0$, $v_2 < 0$ and $v_{12} < 0$. An admissible tax system T can be represented by a non-decreasing earnings function $y : \Omega \to \mathbb{R}_+$. Specifically, by the taxation principle, see e.g. Hammond (1979) or Guesnerie (1995), an allocation (c, y) consisting of a consumption schedule $c : \Omega \to \mathbb{R}_+$ and an earnings schedule $y : \Omega \to \mathbb{R}_+$ can be induced by an income tax if and only if it satisfies the resource constraint,

$$\mathbb{E}[y(\omega)] \geq \mathbb{E}[c(\omega)] \tag{41}$$

and incentive compatibility constraints: For all ω and ω',

$$u(\omega) \geq c(\omega') - v(y(\omega'), \omega) \ . \tag{42}$$

where

$$u(\omega) := c(\omega) - v(y(\omega), \omega) \ . \tag{43}$$

gives the utility that a type ω individual realizes under allocation (c, y).

It is also well known how to obtain a characterization of incentive compatible allocations in models with quasilinear preferences, see e.g. Myerson (1981). The utility realized by any one type-ω individual under earnings function y can be written as a sum of two terms, the minimal level of utility that is realized by the "poorest" type and the extra utility realized by higher types. More formally, an application of the envelope theorem makes it possible to show that incentive compatibility holds if and only if two conditions are satisfied. First, for all ω,

$$u(\omega) = \underline{u} - \int_{\underline{\omega}}^{\omega} v_2(y(z), z) \, dz \ , \tag{44}$$

where $\underline{u} := u(\underline{\omega})$ is a shorthand for the lowest type's utility and $-\int_{\underline{\omega}}^{\omega} v_2(y(z), z)\ dz$ is the *information rent* realized by a higher type $\omega > \underline{\omega}$ in the presence of incentive compatibility constraints. This terminology reflects that private information on types is the impediment to first-best redistribution. Second, y is a non-decreasing function, i.e., individuals with higher productive abilities must not earn less than individuals with lower productive abilities.

We can use these insights to derive a representation of preferences over tax polices in a reduced form that only depends on the income function y and no longer involves a reference to the consumption function c. This will enable us to represent a party's problem of choosing a tax policy as an optimization problems that no longer involve resource and incentive constraints. Suppose that (c, y) is incentive compatible, then using (43), (44) and an integration by parts we obtain

$$\mathbb{E}[c(\omega)] = \underline{u} + \mathbb{E}\left[v(y(\omega), \omega) - \frac{1 - F(\omega)}{f(\omega)}\ v_2(y(\omega), \omega)\right] .$$

Plugging this expression into the public sector budget constraint $\mathbb{E}[y(\omega)] - \mathbb{E}[c(\omega)] = 0$ yields an expression for $\underline{u}$; it is equal to the *virtual surplus* that is associated with an earnings function y:

$$\underline{u} := s_v(y) := \mathbb{E}\left[y(\omega) - v(y(\omega), \omega) + \frac{1 - F(\omega)}{f(\omega)}\ v_2(y(\omega), \omega)\right] . \tag{45}$$

The virtual surplus is a surplus measure that takes account of the information rents that tax-payers realize and which reduces what is available for the lowest type. To arrive at the virtual surplus, the surplus of aggregate output over costs of effort

$$s_v(y) := \mathbb{E}\left[y(\omega) - v(y(\omega), \omega)\right]$$

is reduced by the aggregate information rent

$$-\mathbb{E}\left[\int_{\underline{\omega}}^{\omega} v_2(y(z), z)\ dz\right] = -\mathbb{E}\left[\frac{1 - F(\omega)}{f(\omega)}\ v_2(y(\omega), \omega)\right] ,$$

where the equality follows from an integration by parts. Thus,

$$\underline{u} = s_v(y) = \mathbb{E}\left[y(\omega) - v(y(\omega), \omega) + \frac{1 - F(\omega)}{f(\omega)}\ v_2(y(\omega), \omega)\right] . \tag{46}$$

Indirect utility induced by an incentive compatible allocation can now be written as a sum of virtual surplus and information rents

$$u(\omega) := s_v(y) - \int_{\underline{\omega}}^{\omega} v_2(y(z), z)\, dz\ . \tag{47}$$

With this characterization, the utility realized by a type ω individual depends on the whole earnings schedule $y : \Omega \to \mathbb{R}_+$ but no longer on the consumption schedule $c : \Omega \to \mathbb{R}_+$. When we study political competition over income tax schedules, we will denote party 1's proposal by $y^1 : \Omega \to \mathbb{R}_+$ and party 2's proposal by $y^2 : \Omega \to \mathbb{R}_+$ and we denote by $u^1(\omega)$ and $u^2(\omega)$ the associated utility levels for a type ω individual.

To summarize, for the application of non-linear income taxation, the policy space is the set of all non-decreasing earnings functions. Any such function generates a payoff profile that is characterized by equations (44) and (46). We are particularly interested in the marginal tax rates that are associated with the tax systems that the parties propose. To get from an incentive compatible allocation to the associated tax schedule T we use the first order condition of the utility-maximization problem that individuals face in the presence of this tax system. If tax system T induces an incentive compatible allocation (c, y), then

$$1 - T'(y(\omega)) = v_1(y(\omega), \omega)\ .$$

Hence, $1 - v_1(y(\omega), \omega)$ is interpreted as the marginal tax rates that type ω agents face.

For the government problem, we consider a class of additive social welfare functions. Members of this class differ with respect to the specification of welfare weights $\{\mathbf{B}(\omega)\}_{\omega\in\Omega}$. The social welfare functions take the form

$$\mathcal{SW}(y) = \mathbb{E}\left[\mathbf{B}(\omega)\, u(\omega)\right]\ .$$

It is wlog to adopt the normalization $\mathbb{E}[\mathbf{B}(\omega)] = 1$. Using equations (44) and (46) and another integration by parts we can rewrite social welfare as

$$\mathcal{SW}(y) = \mathbb{E}\left[y(\omega) - v(y(\omega), \omega) + \frac{1 - F(\omega)}{f(\omega)}\left(1 - \mathcal{G}_{\mathbf{B}}(\omega)\right) v_2(y(\omega), \omega)\right]\ ,$$

where $\mathcal{G}_{\mathbf{B}}(\omega) := \mathbb{E}[\mathbf{B}(z) \mid z \geq \omega]$. The problem of welfare-maximization is to choose the earnings function y that maximizes this welfare objective over the set of non-decreasing functions. The relaxed problem of welfare-maximization is to choose the earnings function that maximizes this welfare objective over the set of all functions.

Assumption 4. *Suppose that the following conditions hold:*

1. *For any weighting function $\mathbf{B} : \Omega \to \mathbb{R}_+$ with $\mathbb{E}[\mathbf{B}(\omega)] = 1$, the above problem of welfare-maximization has a unique solution. Let $y_{\mathbf{B}}$ be the earnings function that solves this problem.*

2. *For any weighting function $\mathbf{B} : \Omega \to \mathbb{R}_+$ with $\mathbb{E}[\mathbf{B}(\omega)] = 1$, the relaxed problem of welfare-maximization has a unique solution. Let $y^r_{\mathbf{B}}$ be the earnings function that solves the this relaxed problem.*

3. *The function $y^r_{\mathbf{B}}$ satisfies Diamond's formula; i.e. for all ω,*

$$1 - v_1(y(\omega), \omega) = -\frac{1 - F(\omega)}{f(\omega)} \; (1 - \mathcal{G}_{\mathbf{B}}(\omega)) \; v_{21}(y(\omega), \omega) \; .$$

   Moreover, $y^r_{\mathbf{B}}$ is the only function that satisfies Diamond's formula.

4. *For all ω so that the monotonicity constraint on $y_{\mathbf{B}}$ is not binding, $y^r_{\mathbf{B}}(\omega) = y_{\mathbf{B}}(\omega)$.*

Assumption 4 is routinely invoked in models of optimal income taxation. The assumption can be justified. With an appropriate choice of the primitives, the solutions to the relaxed and the full problem of welfare-maximization can be shown to satisfy properties 1.– 4. We simply impose Assumption 4 as a shortcut.

D.2 Proof of Proposition 3

We start by analyzing the problem to choose y^1 so as to maximize $W^1(y^1, y^2)$.[24] The following notation will prove helpful. Remember that $W^1(y^1, y^2) = \mathbb{E}[G^1_W(u^1(\omega) -$

[24] We thereby provide a characterization of the solution to a relaxed problem, as opposed to the full problem to maximize $W^1(y^1, y^2)$ subject to the constraint that y^1 is a non-decreasing function. Obviously, if the solution to the relaxed problem is non-decreasing then it is also a solution to the full problem. Otherwise, the solution of the full problem will give rise to bunching. While it is well known how the analysis would have to be modified if bunching is an issue, see e.g. Hellwig (2007) or Brett and Weymark (2017), the trade-offs that shape best responses are, however, more easily exposed when focusing on the relaxed problem.

$u^2(\omega) \mid \omega)]$. We denote the derivative of the function $G^1_W(\cdot \mid \omega)$ by

$$g^1_W(\cdot \mid \omega) := B\left(u^1(\omega) - u^2(\omega) \mid \omega\right)$$

(this expression has been derived in Section C) and write

$$\begin{array}{rcl} \bar{g}^1_W(\omega \mid y^1, y^2) & := & \mathbb{E}[B(u^1(\omega') - u^2(\omega') \mid \omega') \mid \omega' \geq \omega] \\ & = & \int_\omega^{\bar{\omega}} B(u^1(\omega') - u^2(\omega') \mid \omega') \frac{f(\omega')}{1-F(\omega)} d\omega' \end{array}$$

for the average value of $B(u^1(\omega') - u^2(\omega') \mid \omega')$ among individuals with a type ω' above some cutoff ω. To interpret these expressions, suppose that party 1 offers slightly more utility to type ω' individuals. Then $B(u^1(\omega') - u^2(\omega') \mid \omega')$ measures the extra gain that type ω'-supporters of party 1 realize in the event that party 1 wins rather party 2. Therefore, $\bar{g}^1_W(\omega \mid T^1, T^2)$ is the gain that party 1 can generate by offering all agents with types above ω slightly more utility. The gain that party 1 can generate by slightly raising everybody's utility is given by $\bar{g}^1_W(\underline{\omega} \mid y^1, y^2)$ and the ratio

$$\mathcal{G}^1_W(\omega \mid y^1, y^2) := \frac{\bar{g}^1_W(\omega \mid y^1, y^2)}{\bar{g}^1_W(\underline{\omega} \mid y^1, y^2)}$$

relates the gain from making everybody with a type above ω better off to the gain from making everybody better off.

Lemma 9. *Given y^2, the solution to $\max_{y^1}\ W^1(y^1, y^2)$ is such that, for all ω,*

$$\frac{T^{1\prime}(y^1(\omega))}{1 - T^{1\prime}(y^1(\omega))} \quad = \quad -\frac{1 - F(\omega)}{f(\omega)} \ \left(1 - \mathcal{G}^1_W(\omega \mid y^1, y^2)\right) \ \frac{v_{21}(y^1(\omega), \omega))}{v_1(y^1(\omega), \omega)} \ . \tag{48}$$

__**Proof of Lemma 9**__. We begin by stating party 1's best response problem in a way that enables an analysis using a Gateaux differential. Let $y^1 = y^{1*} + \epsilon\ h^1$, be a perturbed version of party 1's best response y^{1*}, in which ϵ is a scalar and $h^1 : \Omega \to \mathbb{R}$ is a function. If y^{1*} is a best response, then, for any perturbation (ϵ, h^1),

$$\begin{array}{l} \mathbb{E}\left[G^1_W\left(s_v(y^{1*}) - \int_{\underline{\omega}}^{\omega} v_2(y^{1*}(z), z)\ dz\ - u^2(\omega) \mid \omega\right)\right] \\ \geq \mathbb{E}\left[G^1_W\left(s_v(y^{1*} + \epsilon\ h^1) - \int_{\underline{\omega}}^{\omega} v_2(y^{1*}(z) + \epsilon\ h^1(z), z)\ dz\ - u^2(\omega) \mid \omega\right)\right] \ . \end{array} \tag{49}$$

Equivalently, for any function h^1, $\epsilon = 0$ must be a maximizer of the auxiliary function

$$A(\epsilon \mid y^{1*}, y^2) = \mathbb{E}\left[G_W^1\left(s_v(y^{1*} + \epsilon\, h^1) - \int_{\underline{\omega}}^{\omega} v_2(y^{1*}(z) + \epsilon\, h^1(z), z)\, dz\ \ - u^2(\omega) \mid \omega\right)\right] .$$

In the following, we will characterize y^{1*} by analyzing the implications of the requirement that the derivative of this expression with respect to ϵ, evaluated at $\epsilon = 0$, is equal to zero, i.e.,

$$A_{h^1}(y^{1*}, y^2) = 0\ , \tag{50}$$

for all functions h^1. Formally, $A_{h^1}(y^{1*}, y^2)$ is the Gateaux differential of

$$\mathbb{E}\left[G_W^1\left(s_v(y^1) - \int_{\underline{\omega}}^{\omega} v_2(y^1(z), z)\, dz\ \ - u^2(\omega) \mid \omega\right)\right]$$

in direction h^1 evaluated at $y^1 = y^{1*}$.

We note that

$$A_{h^1}(y^{1*}, y^2) = \mathbb{E}\left[g_W^1(\omega \mid y^{1*}, y^2)\ \left(s_{v,\tau^1}(y^{1*}) - \int_{\underline{\omega}}^{\omega} h^1(z)\ v_{21}(y^{1*}(z), z)\, dz\right)\right]\ , \tag{51}$$

where

$$g_W^1(\omega \mid y^{1*}, y^2) := b\left(s_v(y^{1*}) - \int_{\underline{\omega}}^{\omega} v_2(y^{1*}(z), z)\, dz - u^2(\omega) \mid \omega\right)$$

and

$$s_{v,h^1}(y^{1*}) \ \ = \ \ \mathbb{E}\left[h^1(\omega)\left(1 - v_1(y^{1*}(\omega), \omega) + \tfrac{1-F(\omega)}{f(\omega)} v_{21}(y^{1*}(\omega), \omega)\right)\right]$$

is the Gateaux differential of the virtual surplus $s_v(y^1)$ in direction h^1 evaluated at $y^1 = y^{1*}$.

Equation (51) can now be rewritten as

$$\begin{aligned} A_{h^1}(y^{1*}, y^2) \ \ = \ \ & \bar{g}_W^1(\underline{\omega} \mid y^{1*}, y^2)\mathbb{E}\left[h^1(\omega)\left(1 - v_1(y^{1*}(\omega), \omega) + \tfrac{1-F(\omega)}{f(\omega)} v_{21}(y^{1*}(\omega), \omega)\right)\right] \\ & - \mathbb{E}\left[g_W^1(\omega \mid y^{1*}, y^2)\ \textstyle\int_{\underline{\omega}}^{\omega} h^1(z)\ v_{21}(y^{1*}(z), z)\, dz\right] \end{aligned} \tag{52}$$

where, for any $\omega \in \Omega$, $\bar{g}_W^1(\omega \mid y^{1*}, y^2) := \mathbb{E}[g_W^1(\omega' \mid y^{1*}, y^2) \mid \omega' \geq \omega]$. Moreover, an integration by parts shows that

$$\mathbb{E}\left[g_W^1(\omega \mid y^{1*}, y^2) \int_{\underline{\omega}}^{\omega} h^1(z)\, v_{21}(y^{1*}(z), z)\, dz\right] \tag{53}$$
$$= \mathbb{E}\left[h^1(\omega)\, \bar{g}_W^1(\omega \mid y^{1*}, y^2)\, \frac{1-F(\omega)}{f(\omega)} v_{21}(y^{1*}(\omega), \omega)\right]$$

so that condition (50) can equivalently be written as the requirement that, for all functions h^1,

$$\mathbb{E}\left[h^1(\omega)\left(1 - v_1(y^{1*}(\omega), \omega) + (1 - \mathcal{G}_W^1(\omega \mid y^{1*}, y^2))\frac{1-F(\omega)}{f(\omega)} v_{21}(y^{1*}(\omega), \omega)\right)\right] = 0 \tag{54}$$

where $\mathcal{G}_W^1(\omega \mid y^{1*}, y^2) = \frac{\bar{g}_W^1(\omega | y^{1*}, y^2)}{\bar{g}_W^1(\underline{\omega} | y^{1*}, y^2)}$. Condition (54) can hold only if, for all ω,

$$1 - v_1(y^{1*}(\omega), \omega) + (1 - \mathcal{G}_W^1(\omega \mid y^{1*}, y^2))\frac{1-F(\omega)}{f(\omega)} v_{21}(y^{1*}(\omega), \omega) = 0\ , \tag{55}$$

or, equivalently, if

$$\frac{1 - v_1(y^{1*}(\omega), \omega)}{v_1(y^{1*}(\omega), \omega)} = -(1 - \mathcal{G}_W^1(\omega \mid y^{1*}, y^2))\ \frac{1-F(\omega)}{f(\omega)}\ \frac{v_{21}(y^{1*}(\omega), \omega)}{v_1(y^{1*}(\omega), \omega)}\ . \tag{56}$$

Using $T^{1\prime}(y^{1*}(\omega)) = 1 - v_1(y^{1*}(\omega), \omega)$ we can rewrite this equation as

$$\frac{T^{1\prime}(y^{1*}(\omega))}{1 - T^{1\prime}(y^{1*}(\omega))} = -(1 - \mathcal{G}_W^1(\omega \mid y^{1*}, y^2))\ \frac{1-F(\omega)}{f(\omega)}\ \frac{v_{21}(y^{1*}(\omega), \omega)}{v_1(y^{1*}(\omega), \omega)}\ , \tag{57}$$

which is what had to be shown. □

The optimality condition in (48) could also be derived on the assumption that party 1 seeks to maximize $s_v(y^1) - \mathbb{E}\left[\mathcal{G}(\omega)\ v_2(y^1(\omega), \omega)\right]$ for $\mathcal{G}(\omega) = \mathcal{G}_W^1(\omega \mid y^1, y^2)$, i.e., a utilitarian welfare objective with particular weights on the rents that are realized by types $\omega > \underline{\omega}$. In this maximization problem the weighting function $\mathcal{G}$ is exogenously fixed, albeit at the level that is induced by y^2 and party 1's best response. Thus, as discussed in the main body, equation (48) is akin to Diamond (1998)'s formula.

The following Lemma describes the solution to another auxiliary problem for party 1, namely the problem to choose policy with the objective to minimize what is at stake for the supporters of party 2. We omit a proof and discussion of the Lemma as it would involve only a straightforward adjustment of those of Lemma 9. The Lemma involves a weighting function $\mathcal{G}_W^2$ for information rents that is derived from $W^2(y^1, y^2) = \mathbb{E}[G_W^2(u^2(\omega) - u^1(\omega) \mid \omega)]$ along the same lines as $\mathcal{G}_W^1$ is derived from $W^1(y^1, y^2)$, where we now have:

$$g_W^2(\cdot \mid \omega) \equiv 1 - B\left(u^1(\omega) - u^2(\omega) \mid \omega\right).$$

Lemma 10. *Given y^2, the solution to $\min_{y^1}\ W^2(y^1, y^2)$ is such that, for all ω,*

$$\frac{T^{1\prime}(y^1(\omega))}{1 - T^{1\prime}(y^1(\omega))} \quad = \quad -\frac{1 - F(\omega)}{f(\omega)}\ \left(1 - \mathcal{G}_W^2(\omega \mid y^1, y^2)\right)\ \frac{v_{21}(y^1(\omega), \omega))}{v_1(y^1(\omega), \omega)}. \tag{58}$$

Main result. We finally turn to party 1's best response problem of interest and then provide an equilibrium characterization. We proceed by introducing notation for a weighted average of $\mathcal{G}_W^1$ and $\mathcal{G}_W^2$. Let

$$\begin{aligned}
\gamma^1(\omega \mid y^1, y^2) \quad &:= \quad \frac{1}{1 + \psi^1(y^1, y^2)} g_W^1(u^1(\omega) - u^2(\omega) \mid y^1, y^2) \\
&\quad\quad + \frac{\psi^1(y^1, y^2)}{1 + \psi^1(y^1, y^2)} g_W^2(u^1(\omega) - u^2(\omega) \mid y^1, y^2) \\
&= \quad \frac{1}{1 + \psi^1(y^1, y^2)} B\left(u^1(\omega) - u^2(\omega) \mid \omega\right) \\
&\quad\quad + \frac{\psi^1(y^1, y^2)}{1 + \psi^1(y^1, y^2)} \left(1 - B\left(u^1(\omega) - u^2(\omega) \mid \omega\right)\right)
\end{aligned}$$

and denote by

$$\bar{\gamma}^1(\omega \mid y^1, y^2) \quad := \quad \mathbb{E}[\gamma^1(\omega' \mid y^1, y^2) \mid \omega' \geq \omega]$$

and

$$\mathcal{G}_{SP}^1(\omega \mid y^1, y^2) := \frac{\bar{\gamma}^1(\omega \mid y^1, y^2)}{\bar{\gamma}^1(\underline{\omega} \mid y^1, y^2)}.$$

When both parties propose the same policies $y^1 = y^2 =: y$, we have $u^1(\omega) - u^2(\omega) = 0$ and we can suppress the dependence of these expressions on y^1 and y^2 and write

$$\bar{\gamma}(\omega) := \bar{\gamma}^1(\omega \mid y, y), \quad \text{and} \quad \mathcal{G}_{SP}(\omega) := \frac{\bar{\gamma}(\omega)}{\bar{\gamma}(\underline{\omega})}.$$

For convenience, we restate our main result, Proposition 3, using the notations of this section.

Proposition (Political equilibrium nonlinear tax policy). *Given y^2, if y^1 is a maximizer of $\psi^1(y^1, y^2)$ then, for all ω,*

$$\frac{{T^1}'(y^1(\omega))}{1 - {T^1}'(y^1(\omega))} \quad = \quad -\frac{1 - F(\omega)}{f(\omega)} \left(1 - \mathcal{G}^1_{SP}(\omega \mid y^1, y^2)\right) \, \frac{v_{21}(y^1(\omega), \omega))}{v_1(y^1(\omega), \omega)} \, . \tag{59}$$

If (y^1, y^2) is a saddle point of ψ^1, then $y^1 = y^2$, where y^1 is such that, for all ω,

$$\frac{{T^1}'(y^1(\omega))}{1 - {T^1}'(y^1(\omega))} \quad = \quad -\frac{1 - F(\omega)}{f(\omega)} \left(1 - \mathcal{G}_{SP}(\omega)\right) \, \frac{v_{21}(y^1(\omega), \omega))}{v_1(y^1(\omega), \omega)} \, . \tag{60}$$

Proof of Proposition 3. Let y^2 be an arbitrary, but given function, possibly equal to y^{2*}. Given y^2 we look at the problem to choose y^1 with the objective to maximize

$$\psi^1(y^1, y^2) = \frac{W^1(y^1, y^2)}{W^2(y^1, y^2)} \, .$$

Suppose that y^{1*} is a solution to that problem. Then, it must also be that case that $\epsilon = 0$ solves the problem to choose a scalar ϵ with the objective to maximize

$$\psi^1(y^{1*} + \epsilon \, h^1, y^2) = \frac{W^1(y^{1*} + \epsilon \, h^1, y^2)}{W^2(y^{1*} + \epsilon \, h^1, y^2)} \, .$$

for any given but arbitrary function h^1. That is, we can characterize y^{1*} be the requirement that

$$\left. \frac{\partial \psi^1(y^{1*} + \epsilon \, h^1, y^2)}{\partial \epsilon} \right|_{\epsilon = 0} = 0 \, ,$$

or, equivalently, that

$$W^1_{h^1}(y^{1*}, y^2) \, W^2(y^{1*}, y^2) - W^1(y^{1*}, y^2) \, W^2_{h^1}(y^{1*}, y^2) = 0 \, ,$$

where $W^j_{h^1}$ is the Gateaux differential of W^1 is the direction h^1.

Best responses. The following equations provide a characterization of $W^1_{h^1}$ and $W^2_{h^1}$ and of the analogous expressions $W^1_{h^2}$ and $W^2_{h^2}$ that are relevant for party 2's best response problem. We omit a proof as it would require only an easy adaptation of the arguments in the proof of Lemma 9. The Gateaux differential of W^1 in the direction h^1 evaluated at (y^{1*}, y^2) equals

$$\begin{aligned} & W^1_{h^1}(y^{1*}, y^2) = \bar{g}^1_W(\underline{\omega} \mid y^{1*}, y^2) \times \\ & \mathbb{E}\left[h^1(\omega)\left\{1 - v_1(y^{1*}(\omega), \omega) + \tfrac{1-F(\omega)}{f(\omega)}\left(1 - \mathcal{G}^1_W(\omega \mid y^{1*}, y^2)\right) v_{21}(y^{1*}(\omega), \omega)\right\}\right]. \end{aligned} \tag{61}$$

The Gateaux differential of W^2 in the direction h^1 evaluated at (y^{1*}, y^2) equals

$$\begin{aligned} & W^2_{h^1}(y^{1*}, y^2) = -\bar{g}^2_W(\underline{\omega} \mid y^{1*}, y^2) \times \\ & \mathbb{E}\left[h^1(\omega)\left\{1 - v_1(y^{1*}(\omega), \omega) + \tfrac{1-F(\omega)}{f(\omega)}\left(1 - \mathcal{G}^2_W(\omega \mid y^{1*}, y^2)\right) v_{21}(y^{1*}(\omega), \omega)\right\}\right]. \end{aligned} \tag{62}$$

The Gateaux differential of W^1 in the direction h^2 evaluated at (y^1, y^{2*}) equals

$$\begin{aligned} & W^1_{h^2}(y^1, y^{2*}) = -\bar{g}^1_W(\underline{\omega} \mid y^1, y^{2*}) \times \\ & \mathbb{E}\left[h^2(\omega)\left\{1 - v_1(y^{2*}(\omega), \omega) + \tfrac{1-F(\omega)}{f(\omega)}\left(1 - \mathcal{G}^1_W(\omega \mid y^1, y^{2*})\right) v_{21}(y^{2*}(\omega), \omega)\right\}\right] \ . \end{aligned} \tag{63}$$

The Gateaux differential of W^2 in the direction h^2 evaluated at (y^1, y^{2*}) equals

$$\begin{aligned} & W^2_{h^2}(y^1, y^{2*}) = \bar{g}^2_W(\underline{\omega} \mid y^1, y^{2*}) \times \\ & \mathbb{E}\left[h^2(\omega)\left\{1 - v_1(y^{2*}(\omega), \omega) + \tfrac{1-F(\omega)}{f(\omega)}\left(1 - \mathcal{G}^2_W(\omega \mid y^1, y^{2*})\right) v_{21}(y^{2*}(\omega), \omega)\right\}\right] \ . \end{aligned} \tag{64}$$

Now, the Gateaux differential of $\psi^1 = W^1/W^2$ in the direction h^1 evaluated at (y^{1*}, y^2) has the same sign as

$$\begin{aligned} & \frac{1}{\bar{\gamma}^1(\underline{\omega} \mid y^{1*}, y^2)} \frac{1}{1 + \psi^1(y^{1*}, y^2)} \frac{1}{W^2(y^{1*}, y^2)} \times \\ & \left\{W^1_{h^1}(y^{1*}, y^2)\ W^2(y^{1*}, y^2) - W^1(y^{1*}, y^2)\ W^2_{h^1}(y^{1*}, y^2)\right\} \\ = \ & \frac{1}{\bar{\gamma}^1(\underline{\omega} \mid y^{1*}, y^2)} \left\{\frac{1}{1 + \psi^1(y^{1*}, y^2)} W^1_{h^1}(y^{1*}, y^2) - \frac{\psi^1(y^{1*}, y^2)}{1 + \psi^1(y^{1*}, y^2)} W^2_{h^1}(y^{1*}, y^2)\right\} \end{aligned}$$

Now notice that

$$
\begin{aligned}
& \frac{1}{1+\psi^1(y^{1*},y^2)} \bar{g}^1_W(\underline{\omega} \mid y^{1*},y^2) \left(1-\mathcal{G}^1_W(\omega \mid y^{1*},y^2)\right) \\
& + \frac{\psi^1(y^{1*},y^2)}{1+\psi^1(y^{1*},y^2)} \bar{g}^2_W(\underline{\omega} \mid y^{1*},y^2) \left(1-\mathcal{G}^2_W(\omega \mid y^{1*},y^2)\right) \\
= \;& \left\{ \frac{1}{1+\psi^1(y^{1*},y^2)} \bar{g}^1_W(\underline{\omega} \mid y^{1*},y^2) + \frac{\psi^1(y^{1*},y^2)}{1+\psi^1(y^{1*},y^2)} \bar{g}^2_W(\underline{\omega} \mid y^{1*},y^2) \right\} \\
& - \left\{ \frac{1}{1+\psi^1(y^{1*},y^2)} \bar{g}^1_W(\omega \mid y^{1*},y^2) + \frac{\psi^1(y^{1*},y^2)}{1+\psi^1(y^{1*},y^2)} \bar{g}^2_W(\omega \mid y^{1*},y^2) \right\} \\
= \;& \bar{\gamma}^1(\underline{\omega} \mid y^{1*},y^2) \left(1-\mathcal{G}^1_{SP}(\omega \mid y^1,y^2)\right) .
\end{aligned}
$$

Therefore, expressions (61) and (62) imply that the Gateaux differential of ψ^1 is given by

$$
\mathbb{E}\left[h^1(\omega) \left\{ 1 - v_1(y^{1*}(\omega),\omega) + \tfrac{1-F(\omega)}{f(\omega)} \left(1-\mathcal{G}^1_{SP}(\omega \mid y^{1*},y^2)\right) v_{21}(y^{1*}(\omega),\omega) \right\} \right] . \tag{65}
$$

Analogously, the Gateaux differential of ψ^1 in the direction h^2 evaluated at (y^1, y^{2*}) has the same sign as

$$
\begin{aligned}
& \tfrac{1}{\bar{\gamma}^1(\underline{\omega} \mid y^1,y^{2*})} \tfrac{1}{1+\psi^1(y^1,y^{2*})} \tfrac{1}{W^2(y^1,y^{2*})} \left\{ W^1_{h^2}(y^1,y^{2*})\, W^2(y^1,y^{2*}) - W^1(y^1,y^{2*})\, W^2_{h^2}(y^1,y^{2*}) \right\} \\
& = -\mathbb{E}\left[h^2(\omega) \left\{ 1 - v_1(y^{2*}(\omega),\omega) + \tfrac{1-F(\omega)}{f(\omega)} \left(1-\mathcal{G}^1_{SP}(\omega \mid y^1,y^{2*})\right) v_{21}(y^{2*}(\omega),\omega) \right\} \right] .
\end{aligned} \tag{66}
$$

Subgame perfect equilibrium. Now, if y^{1*} and y^{2*} are mutually best responses, then it must be the case that the Gateaux differentials

$$
W^1_{h^1}(y^{1*},y^{2*})\, W^2(y^{1*},y^{2*}) - W^1(y^{1*},y^{2*})\, W^2_{h^1}(y^{1*},y^{2*}) = 0 ,
$$

for all functions h^1 and

$$
W^1_{h^2}(y^{1*},y^{2*})\, W^2(y^{1*},y^{2*}) - W^1(y^{1*},y^{2*})\, W^2_{h^2}(y^{1*},y^{2*}) = 0 ,
$$

for all functions h^2. The conditions above require that, for all ω,

$$1 - v_1(y^{1*}(\omega), \omega) + \tfrac{1-F(\omega)}{f(\omega)} \left(1 - \mathcal{G}^1_{SP}(\omega \mid y^{1*}, y^{2*})\right) v_{21}(y^{1*}(\omega), \omega) = 0 \ , \tag{67}$$

and

$$1 - v_1(y^{2*}(\omega), \omega) + \tfrac{1-F(\omega)}{f(\omega)} \left(1 - \mathcal{G}^1_{SP}(\omega \mid y^{1*}, y^{2*})\right) v_{21}(y^{2*}(\omega), \omega) = 0 \ . \tag{68}$$

An inspection of (67) and (68) implies in particular that in any interior subgame perfect equilibrium, $y^{1*} = y^{2*}$. To complete the proof of Proposition 3, we note that rearranging the terms in (67) and using that $T'(y(\omega)) = v_1(y(\omega), \omega)$ yields equation (60). □

E Proof of Lemma 4: Equilibrium existence for non-linear taxation

Party 1 chooses an earnings function y^1 and party 2 chooses an earnings function y^2. The objective of party 1 is to maximize

$$\psi^1(y^1, y^2) = \frac{W^1(y^1, y^2)}{W^2(y^1, y^2)}$$

and the objective of party 2 is to minimize this expression. As we clarify below, focusing on this case $\mu \to \infty$ simplifies the exposition and the argument does not depend on it. For ease of exposition, we ignore the monotonicity constraints on y^1 and y^2 in what follows. Thus, we characterize the policies that are mutual best responses in the sense of solving relaxed best response problems.

Saddle point. We have shown above that, under Assumption 4, if an equilibrium exists, it is symmetric, i.e. such that $y^1 = y^2$. Assumption 4 ensures that there is a unique solution to the first order conditions (67) and (68). We now show that, under Assumption 2, this hypothetical equilibrium is a local saddle point of the function ψ^1.

Lemma 11. *Suppose that the Assumptions of Lemma 4 are satisfied. Then, a pair of policies that satisfies (60) is a saddle point of the function ψ^1.*

Proof of Lemma 11. Suppose that a pair of policies (y^{1*}, y^{2*}) satisfies the first order conditions (74) and (74). We seek to show that (y^{1*}, y^{2*}) is a saddle point of the function

$$\psi^1(y^1, y^2) = \frac{W^1(y^1, y^2)}{W^2(y^1, y^2)} \ .$$

We now state this saddle point condition in a way that enables an analysis using functional derivatives. Let $y^1 = y^{1*} + \epsilon^1 \ h^1$, be a perturbed version of y^{1*}, in which ϵ^1 is a scalar and $h^1 : \Omega \to \mathbb{R}$ is a function. Analogously, let $y^2 = y^{2*} + \epsilon^2 \ h^2$, be a perturbed version of y^2. The saddle point condition according to which, for all (y^1, y^2),

$$\psi^1(y^1, y^{2*}) \leq \psi^1(y^{1*}, y^{2*}) \leq \psi^1(y^{1*}, y^2)$$

can therefore be written as: for any pair of perturbations (ϵ^1, h^1) and (ϵ^2, h^2),

$$\psi^1(y^{1*} + \epsilon^1 \ h^1, y^{2*}) \leq \psi^1(y^{1*}, y^{2*}) \leq \psi^1(y^{1*}, y^{*2} + \epsilon^2 \ h^2) \ . \tag{69}$$

Equivalently, for all functions (h^1, h^2), the point $(\epsilon^1, \epsilon^2) = (0, 0)$ must be a saddle-point of

$$\psi^1(y^{1*} + \epsilon^1 \ h^1, y^{2*} + \epsilon^1 \ h^2) = \frac{W^1(y^{1*} + \epsilon^1 \ h^1, y^{*2} + \epsilon^2 \ h^2)}{W^2(y^{1*} + \epsilon^1 \ h^1, y^{*2} + \epsilon^2 \ h^2)} \ .$$

Having a saddle point requires that all entries of the Jacobi-matrix

$$J_\psi(y^{1*}, y^{2*}) = \begin{pmatrix} \psi^1_{\epsilon^1}(y^{1*}, y^{2*}) \\ \psi^1_{\epsilon^2}(y^{1*}, y^{2*}) \end{pmatrix}$$

are equal to zero and that the Hessian

$$H_\psi(y^{1*}, y^{2*}) = \begin{pmatrix} \psi^1_{\epsilon^1,\epsilon^1}(y^{1*}, y^{2*}) & \psi^1_{\epsilon^1,\epsilon^2}(y^{1*}, y^{2*}) \\ \psi^1_{\epsilon^1,\epsilon^2}(y^{1*}, y^{2*}) & \psi^1_{\epsilon^2,\epsilon^2}(y^{1*}, y^{2*}) \end{pmatrix}$$

is indefinite. The policies (y^{1*}, y^{2*}) satisfying the first order conditions (74) and (74) is equivalent to all entries of the Jacobi-matrix being equal to zero. Hence, what remains to be shown is that $H_\psi(y^{1*}, y^{2*})$ is indefinite. To this end, it suffices to show that $\psi^1_{\epsilon^1,\epsilon^1}(y^{1*}, y^{2*}) < 0$, and $\psi^1_{\epsilon^2,\epsilon^2}(y^{1*}, y^{2*}) > 0$. These two inequalities can be shown to hold provided that

$$\frac{\partial}{\partial \epsilon^1} \left\{ W^1_{\epsilon^1}(y^{1*}, y^{2*}) \ W^2(y^{1*}, y^{2*}) - W^1(y^{1*}, y^{2*}) \ W^2_{\epsilon^1}(y^{1*}, y^{2*}) \right\} < 0 \ ,$$

and

$$\frac{\partial}{\partial \epsilon^2}\left\{W^1_{\epsilon^2}(y^{1*},y^{2*})\, W^2(y^{1*},y^{2*}) - W^1(y^{1*},y^{2*})\, W^2_{\epsilon^2}(y^{1*},y^{2*})\right\} > 0\ ,$$

or, equivalently, if

$$W^1_{\epsilon^1,\epsilon^1}(y^{1*},y^{2*})\, W^2(y^{1*},y^{2*}) - W^1(y^{1*},y^{2*})\, W^2_{\epsilon^1,\epsilon^1}(y^{1*},y^{2*}) < 0\ , \tag{70}$$

and

$$W^1_{\epsilon^2,\epsilon^2}(y^{1*},y^{2*})\, W^2(y^{1*},y^{2*}) - W^1(y^{1*},y^{2*})\, W^2_{\epsilon^2,\epsilon^2}(y^{1*},y^{2*}) > 0\ . \tag{71}$$

Now, under Assumption 2,

$$\begin{aligned} & W^1(y^1,y^2) \\ = \;& \mathbb{E}\left[\mathfrak{B}(\omega)\left(u^1(\omega)-u^2(\omega)\right) + \frac{1}{2}\,\mathfrak{b}(\omega)\left(u^1(\omega)-u^2(\omega)\right)^2 + \frac{1}{2}\left(\frac{\mathfrak{B}(\omega)^2}{\mathfrak{b}(\omega)}\right)\right] \end{aligned}$$

and

$$\begin{aligned} & W^2(y^1,y^2) \\ = \;& \mathbb{E}\left[(1-\mathfrak{B}(\omega))\left(u^2(\omega)-u^1(\omega)\right) + \frac{1}{2}\,\mathfrak{b}(\omega)\left(u^2(\omega)-u^1(\omega)\right)^2 + \frac{1}{2}\left(\frac{1-\mathfrak{B}(\omega)^2}{\mathfrak{b}(\omega)}\right)\right]. \end{aligned}$$

It follows from Proposition 3 that $y^{1*}=y^{2*}$, hence $W^1(y^{1*},y^{2*})>0$ and $W^2(y^{1*},y^{2*})>0$. Sufficient conditions for the validity of (70) and (71) are therefore that

$$W^1_{\epsilon^1,\epsilon^1}(y^{1*},y^{2*}) < 0 \quad \text{and} \quad W^2_{\epsilon^1,\epsilon^1}(y^{1*},y^{2*}) > 0\ , \tag{72}$$

and

$$W^1_{\epsilon^2,\epsilon^2}(y^{1*},y^{2*}) > 0 \quad \text{and} \quad W^2_{\epsilon^2,\epsilon^2}(y^{1*},y^{2*}) < 0\ . \tag{73}$$

We can now use the expressions for $W^1_{\epsilon^1}$, $W^2_{\epsilon^1}$, $W^1_{\epsilon^2}$ and $W^2_{\epsilon^2}$ (or, equivalently, the Gateaux differentials $W^1_{h^1}$, $W^2_{h^1}$, $W^1_{h^2}$ and $W^2_{h^2}$) derived above to compute $W^1_{\epsilon^1,\epsilon^1}$, $W^2_{\epsilon^1,\epsilon^1}$, $W^1_{\epsilon^2,\epsilon^2}$ and $W^2_{\epsilon^2,\epsilon^2}$. If we evaluate the resulting expressions in the limit case as $\mathfrak{b}(\omega)$ arbitrarily close to zero for all ω, we can verify that (72) and (73) indeed hold. For instance, we then find that

$$\begin{aligned} & W^1_{\epsilon^1,\epsilon^1}(y^{1*},y^{2*}) \\ = \;& \bar{\mathfrak{B}}(\underline{\omega})\mathbb{E}\left[h^1(\omega)^2\left(-v_{11}(y^*(\omega),\omega) + \frac{1-F(\omega)}{f(\omega)}\left(1-\frac{\bar{\mathfrak{B}}(\omega)}{\bar{\mathfrak{B}}(\underline{\omega})}\right)v_{211}(y^*(\omega),\omega)\right)\right], \end{aligned}$$

where $\bar{\mathfrak{B}}(\omega) := \mathbb{E}\left[\mathfrak{B}(\omega') \mid \omega' \geq \omega\right]$. With $\mathfrak{B}(\omega) = \frac{1}{2}$, for all ω,

$$1 - \frac{\bar{\mathfrak{B}}(\omega)}{\bar{\mathfrak{B}}(\underline{\omega})} = 0 \ ,$$

for all ω, so that

$$W^1_{\epsilon^1,\epsilon^1}(y^{1*}, y^{2*}) = \bar{\mathfrak{B}}(\underline{\omega})\mathbb{E}\left[h^1(\omega)^2\left(-v_{11}(y^*(\omega),\omega)\right)\right] < 0 \ .$$

□

Equilibrium. We now show that the local saddle point characterized in the previous proof is indeed an equilibrium point. To this end, we need to show that it is a best response for party 1 to play the hypothetical equilibrium strategy – on the assumption that party 2 also plays this strategy. The results stated so far only imply that playing the the hypothetical equilibrium strategy is a local best response for party 1. What remains to be shown is that this local best response is also the global best response and that there is no other global best respone. In the following section we will use the contraction mapping theorem to show that this is indeed the case. A symmetric argument then implies that it is a best response for party 2 to play the hypothetical equilibrium strategy provided that party 1 plays accordingly. To simplify the exposition, suppose moreover that the utility function takes the form (11).

Proof of Lemma 4. Let y^{eq} be the hypothetical equilibrium strategy and suppose that $y^2 = y^{eq}$. We have shown above that a best response y^{1*} for party 1 satisfies:

$$\frac{1 - v_1(y^{1*}(\omega),\omega)}{v_1(y^{1*}(\omega),\omega)} = \left(1 + \frac{1}{e}\right)\frac{1 - F(\omega)}{f(\omega)\,\omega}\left(1 - \mathcal{G}^1_{SP}(\omega \mid y^{1*}, y^{eq})\right) \ . \tag{74}$$

or, equivalently,

$$\omega^{1+\frac{1}{e}} y^{1*}(\omega)^{-\frac{1}{e}} - 1 = \left(1 + \frac{1}{e}\right)\frac{1 - F(\omega)}{f(\omega)\,\omega}\left(1 - \mathcal{G}^1_{SP}(\omega \mid y^{1*}, y^{eq})\right) \ .$$

For an arbitrary earnings function y define $A(y) = \{A(\omega, y)\}_{\omega\in\Omega}$ with

$$A(\omega, y) = \left(1 + \left(1 + \frac{1}{e}\right)\frac{1 - F(\omega)}{f(\omega)\,\omega}\left(1 - \mathcal{G}^1_{SP}(\omega \mid y, y^{eq})\right)\right)^{-e}\omega^{1+e} \ .$$

Armed with this notation, we rewrite the previous equation one more time as

$$y^{1*}(\omega) = A(\omega, y^{1*}) \ ,$$

for all ω. We also know from the previous arguments that this equation is satisfied for $y^1 = y^{eq}$.

It proves useful to interpret this equation as characterizing a fixed point in a functional space. Therefore, given an arbitrary earnings function y, first interpret $A(\cdot)$ as a function of the earnings function y and then define by $y^*(A(y))$ the earnings function that satisfies,

$$y^*(\omega, A(y)) = A(\omega, y) \ ,$$

for all ω. By interpreting y^* also as a function of y, we can say that a fixed point of y^* is an earnings function y^{fix} with the property that $y^*(A(y^{fix})) = y^{fix}$. By the previous arguments, we also know that y^{eq} is such a fixed point.

Now, if y^{eq} is not the best response of party 1, this implies that there must be another solution $y^{fix} \neq y^{eq}$ to this fixed point equation. In the following we will rule this possibility out, by showing that, under the conditions of Lemma 4, y^* is a contraction mapping and therefore has one and only one fixed point.

Consider a metric space of earnings functions equipped with the *sup* metric, i.e. for two earnings functions y_a and y_b,

$$d\left(y_a, y_b\right) := \sup_{\omega \in \Omega} \ \mid y_a(\omega) - y_b(\omega) \mid \ .$$

To establish that $y^*(\cdot)$ is a contraction mapping, we need to show that, for any pair (y_a, y_b),

$$d\left(y^*(A(y_a)), y^*(A(y_b))\right) \leq k \ d\left(y_a, y_b\right) \ , \tag{75}$$

for $k \in (0, 1)$.

Remember that the analysis proceeds under the assumption that $\mathfrak{B}(\omega) \in [\frac{1}{2} - \alpha, \frac{1}{2} + \alpha]$ and $\mathfrak{b}(\omega) \leq \beta$, for all ω. In the following, we show that an appropriate choice of α and β ensures that, for any ω, $\mid y^*(\omega, A(y_a)) - y^*(\omega, A(y_b)) \mid$ becomes arbitrarily small. Note that

$$
\begin{aligned}
&\mid y^*(\omega, A(y_a)) - y^*(\omega, A(y_b)) \mid \\
&= \omega^{1+e} \mid \left(1 + \tfrac{1-F(\omega)}{f(\omega)\,\omega}\,(1 - \mathcal{G}^1_{SP}(\omega \mid y_a, y^{eq}))\,\tfrac{1}{e}\right)^{-e} \\
&\qquad\qquad - \left(1 + \tfrac{1-F(\omega)}{f(\omega)\,\omega}\,(1 - \mathcal{G}^1_{SP}(\omega \mid y_b, y^{eq}))\,\tfrac{1}{e}\right)^{-e} \mid \ .
\end{aligned}
$$

Moreover, by continuity,

$$
\mathcal{G}^1_{SP}(\omega \mid y_a, y^{eq}) \to \mathcal{G}^1_{SP}(\omega \mid y_b, y^{eq})
$$

implies

$$
\begin{aligned}
&\mid \left(1 + \tfrac{1-F(\omega)}{f(\omega)\,\omega}\,(1 - \mathcal{G}^1_{SP}(\omega \mid y_a, y^{eq}))\,\tfrac{1}{e}\right)^{-e} \\
&\quad - \left(1 + \tfrac{1-F(\omega)}{f(\omega)\,\omega}\,(1 - \mathcal{G}^1_{SP}(\omega \mid y_b, y^{eq}))\,\tfrac{1}{e}\right)^{-e} \mid \ \to \ 0\ .
\end{aligned}
$$

Thus, it suffices to show that $\mathcal{G}^1_{SP}(\omega \mid y_a, y^{eq})$ is arbitrarily close to $\mathcal{G}^1_{SP}(\omega \mid y_b, y^{eq})$ for an appropriate choice of α and β.

Let $\Delta u(y_a, y^{eq}, \omega) = u(y_a, \omega) - u(y^{eq}, \omega)$. Also let

$$
\overline{\Delta u}(y_a, y^{eq}) = \max_{\omega \in \Omega}\ \Delta u(y_a, y^{eq}, \omega)
$$

and

$$
\underline{\Delta u}(y_a, y^{eq}) = \min_{\omega \in \Omega}\ \Delta u(y_a, y^{eq}, \omega)\ .
$$

It is wlog to assume that y_a is a Pareto-efficient earnings function which implies that

$$
\overline{\Delta u}(y_a, y^{eq}) > 0 > \underline{\Delta u}(y_a, y^{eq})\ .
$$

Recall that (see Section D)

$$
\mathcal{G}^1_{SP}(\omega \mid y_a, y^{eq}) \;=\; \lambda^1(y_a, y^{eq})\,\mathcal{G}^1_W(\omega \mid y_a, y^{eq}) + \lambda^2(y_a, y^{eq}))\,\mathcal{G}^2_W(\omega \mid y_a, y^{eq})\ , \tag{76}
$$

where

$$
\mathcal{G}^1_W(\omega \mid y_a, y^{eq}) = \frac{\int_\omega^{\overline{\omega}} B(\Delta u(y_a, y^{eq}, \omega) \mid \omega)\, d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} B(\Delta u(y_a, y^{eq}, \omega) \mid \omega)\, d\omega} = \frac{\int_\omega^{\overline{\omega}} \{\mathfrak{B}(\omega) + \mathfrak{b}(\omega)\Delta u(y_a, y^{eq}, \omega)\} d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \{\mathfrak{B}(\omega) + \mathfrak{b}(\omega)\Delta u(y_a, y^{eq}, \omega)\} d\omega}
$$

and

$$
\begin{aligned}
\mathcal{G}^2_W(\omega \mid y_a, y^{eq}) &= \frac{\int_\omega^{\overline{\omega}} (1 - B(\Delta u(y_a, y^{eq}, \omega) \mid \omega))\, d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} (1 - B(\Delta u(y_a, y^{eq}, \omega) \mid \omega))\, d\omega} \\
&= \frac{\int_\omega^{\overline{\omega}} \{1 - \mathfrak{B}(\omega) - \mathfrak{b}(\omega)\Delta u(y_a, y^{eq}, \omega)\} d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \{1 - \mathfrak{B}(\omega) - \mathfrak{b}(\omega)\Delta u(y_a, y^{eq}, \omega)\} d\omega}
\end{aligned}
$$

and

$$
\begin{aligned}
\lambda^1(y^1, y^2) &:= \frac{1}{1 + \psi^1(y^1, y^2)} \frac{\bar{g}^1_W(\underline{\omega} \mid y^1, y^2)}{\bar{\gamma}^1(\underline{\omega} \mid y^1, y^2)} \\
\lambda^2(y^1, y^2) &:= \frac{\psi^1(y_a, y^{eq})}{1 + \psi^1(y_a, y^{eq})} \frac{\bar{g}^2_W(\underline{\omega} \mid y_a, y^{eq})}{\bar{\gamma}^1(\underline{\omega} \mid y_a, y^{eq})}
\end{aligned}
$$

so that

$$
\lambda^1(y^1, y^2) + \lambda^2(y^1, y^2) = 1.
$$

The assumptions that $\mathfrak{B}(\omega) \in [\frac{1}{2} - \alpha, \frac{1}{2} + \alpha]$ and $\mathfrak{b}(\omega) \leq \beta$, for all ω, imply that, for all ω,

$$
\frac{1}{2} + \alpha + \beta\, \overline{\Delta u}(y_a, y^{eq}) \quad \geq \quad \mathfrak{B}(\omega) + \mathfrak{b}(\omega)\Delta u(y_a, y^{eq}, \omega) \quad \geq \quad \frac{1}{2} - \alpha + \beta\, \underline{\Delta u}(y_a, y^{eq})\ ,
$$

and

$$
\frac{1}{2} + \alpha - \beta\, \underline{\Delta u}(y_a, y^{eq}) \quad \geq \quad 1 - \mathfrak{B}(\omega) - \mathfrak{b}(\omega)\Delta u(y_a, y^{eq}, \omega) \quad \geq \quad \frac{1}{2} - \alpha - \beta\, \overline{\Delta u}(y_a, y^{eq})\ .
$$

Thus,

$$
\begin{aligned}
\frac{(1 - F(\omega))(\frac{1}{2} + \alpha + \beta\, \overline{\Delta u}(y_a, y^{eq}))}{\frac{1}{2} - \alpha + \beta\, \underline{\Delta u}(y_a, y^{eq})} &\geq \mathcal{G}^1_W(\omega \mid y_a, y^{eq}) \\
&\geq \frac{(1 - F(\omega))(\frac{1}{2} - \alpha + \beta\, \underline{\Delta u}(y_a, y^{eq}))}{\frac{1}{2} + \alpha + \beta\, \overline{\Delta u}(y_a, y^{eq})}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{(1 - F(\omega))(\frac{1}{2} + \alpha - \beta\, \underline{\Delta u}(y_a, y^{eq}))}{\frac{1}{2} - \alpha - \beta\, \overline{\Delta u}(y_a, y^{eq})} &\geq \mathcal{G}^2_W(\omega \mid y_a, y^{eq}) \\
&\geq \frac{(1 - F(\omega))(\frac{1}{2} - \alpha - \beta\, \overline{\Delta u}(y_a, y^{eq}))}{\frac{1}{2} - \alpha - \beta\, \underline{\Delta u}(y_a, y^{eq})}\ ,
\end{aligned}
$$

which implies that

$$\lim_{\alpha,\beta\to 0}\ \mathcal{G}^1_W(\omega \mid y_a, y^{eq}) = \lim_{\alpha,\beta\to 0}\ \mathcal{G}^2_W(\omega \mid y_a, y^{eq}) = 1 - F(\omega)\ . \tag{77}$$

A symmetric argument implies that

$$\mathcal{G}^1_{SP}(\omega \mid y_b, y^{eq}) = \lambda^1(y_b, y^{eq})\ \mathcal{G}^1_W(\omega \mid y_b, y^{eq}) + (1 - \lambda^1(y_b, y^{eq}))\ \mathcal{G}^2_W(\omega \mid y_b, y^{eq})\ , \tag{78}$$

and

$$\lim_{\alpha,\beta\to 0}\ \mathcal{G}^1_W(\omega \mid y_b, y^{eq}) = \lim_{\alpha,\beta\to 0}\ \mathcal{G}^2_W(\omega \mid y_b, y^{eq}) = 1 - F(\omega)\ . \tag{79}$$

Equations (76)-(79) imply that

$$\begin{aligned}
&\lim\nolimits_{\alpha,\beta\to 0}\ \mathcal{G}^1_{SP}(\omega \mid y_a, y^{eq}) - \mathcal{G}^1_{SP}(\omega \mid y_b, y^{eq}) \\
&= (1 - F(\omega))\,(\lambda^1(y_a, y^{eq}) + (1 - \lambda^1(y_a, y^{eq})) - \lambda^1(y_b, y^{eq}) - (1 - \lambda^1(y_b, y^{eq})) \\
&= 0\ .
\end{aligned}$$

$\square$

Interpretation. The assumption that $\mathfrak{b}(\omega) \leq \beta$ for β close to zero implies that strategic substitutes and complements play a limited role. In the limit, i.e. for $\beta = 0$, equilibria are in dominant strategies, and best responses no longer depend on the tax policy propsed by the other party. The assumption that $\mathfrak{B}(\omega) \in [\frac{1}{2} - \alpha, \frac{1}{2} + \alpha]$ for α close to zero implies that the race between the two parties is close. In a neighborhood of such a symmetric, dominant strategy equilibrium we can be assured that an equilibrium exists, is unique, and characterized by the optimality conditions referred to in Proposition 1. For the case $\mu = 0$ there is also a dominant strategy equilibrium under Assumption 2. Looking at an interior μ therefore yields the same conclusion as $\mu = \infty$. Recall that the existence of a pure strategy equilibrium for generic $B(\cdot)$-functions cannot be expected because of the multi-dimensionality of the policy space. The conditions that we have for equilibrium existence, equilibrium uniqness and equilibrium characterization are to be evaluated against this background.

F Alternative Settings

F.1 Amodel that includes ethical voters who always vote

We now assume that the base of each party is split into three groups: a group that always votes, a group that always abstains, and a group of voters whose voting decision follows from a rule-utilitarian calculation. We denote by $\tilde{q}^{jv}(\omega)$ the fraction of definite voters among the type ω supporters of party j, by $\tilde{q}^{ja}(\omega)$ the fraction of definite abstainers and by $\tilde{q}^{ju}(\omega)$ the fraction of rule-utilitarian or ethical supporters, with $\tilde{q}^{jv}(\omega) + \tilde{q}^{ja}(\omega) + \tilde{q}^{ju}(\omega) = 1$. We assume that these are random quantities both from the perspective of parties when choosing platforms and from the perspective of voters when choosing whether or not to vote. We write $\tilde{q}^j = \{\tilde{q}^{jv}(\omega), \tilde{q}^{ja}(\omega), \tilde{q}^{ju}(\omega)\}_{\omega\in\Omega}$ for the collection of random variables that refer to party j. We denote the expected value of the random variable $\tilde{q}^{ju}(\omega)$ by $\bar{q}^{ju}(\omega)$. The total number of votes for party 1 is then a random variable equal to

$$\tilde{V}^1(p^1, p^2, \sigma^1, \tilde{q}^1) = \mathbb{E}[(\tilde{q}^{1v}(\omega) + \sigma^1 \ \tilde{q}^{1u}(\omega))B(u(p^1, \omega) - u(p^2, \omega) \mid \omega)] \ .$$

Analogously, the total number of votes for party 2 equals

$$\tilde{V}^2(p^1, p^2, \sigma^2, \tilde{q}^2) = \mathbb{E}[(\tilde{q}^{2v}(\omega) + \sigma^2 \ \tilde{q}^{2u}(\omega))(1 - B(u(p^1, \omega) \ - u(p^2, \omega) \mid \omega))] \ .$$

We assume throughout that $\mu \to \infty$, so that the per capital cost of voting is equal to $\kappa\sigma^j$.

Assume furthermore that the random variables $\tilde{q}^1$ and $\tilde{q}^2$ are driven by aggregate shocks that affect the shares of definite and rule-utilitarian voters one the one hand and of definite abstainers on the other so that the following two properties are satisfied. First, the ratio of definite and rule-utilitarian voters is not subject to randomness; i.e., shocks affect the ratio of potential voters to definite abstainers without affecting the internal composition of the set of potential voters. Second, among the supporters of party j, the ratio of definite to rule-utilitarian voters is the same for all types.

Assumption 5. *There is a pair of independent random variables, η_1 and η_2, so that, for all ω,*

$$\tilde{q}^{1v}(\omega) = \bar{q}^{1v}(\omega) \ \eta_1 \quad \textit{and} \quad \tilde{q}^{1u}(\omega) = \bar{q}^{1u}(\omega) \ \eta_1$$

and

$$\tilde{q}^{2v}(\omega) = \bar{q}^{2v}(\omega)\ \eta_2 \quad \textit{and} \quad \tilde{q}^{2u}(\omega) = \bar{q}^{2u}(\omega)\ \eta_2.$$

In addition, there are numbers q^{1v}, q^{1u}, q^{2v} and q^{2u} so that, for all ω

$$\bar{q}^{1v}(\omega) = q^{1v} \quad \textit{and} \quad \bar{q}^{1u}(\omega) = q^{1u}$$

and

$$\bar{q}^{2v}(\omega) = q^{2v} \quad \textit{and} \quad \bar{q}^{2u}(\omega) = q^{2u}\ .$$

Under Assumption 5 the total number of votes for party 1 can be written as

$$\tilde{V}^1(p^1, p^2, \sigma^1, \tilde{q}^1) = \eta^1\ V^1(p^1, p^2, \sigma^1)$$

where $V^1(p^1, p^2, \sigma^1) := m^1(\sigma^1)\ \mathbf{B}^1(p^1, p^2)$ and $m^1(\sigma^1) := q^{1v} + \sigma^1\ q^{1u}$ is a multiplier that determines how party 1's base $\mathbf{B}^1(p^1, p^2)$ is transformed into actual votes. Analogously, the votes for party 2 are given by $\tilde{V}^2(p^1, p^2, \sigma^2, \tilde{q}^2) = \eta^2\ V^2(p^1, p^2, \sigma^2)$, where $V^2(p^1, p^2, \sigma^2) := m^2(\sigma^2)\ \mathbf{B}^2(p^1, p^2)$ and $m^2(\sigma^2) = q^{2v} + \sigma^2\ q^{2u}$. Armed with this notation, we can express the probability that party 1 wins as

$$\pi^1(p^1, p^2, \sigma^1, \sigma^2) = P\left(\frac{V^1(p^1, p^2, \sigma^1)}{V^2(p^1, p^2, \sigma^2)}\right) = P\left(\frac{m^1(\sigma^1)}{m^2(\sigma^2)} \frac{\mathbf{B}^1(p^1, p^2)}{\mathbf{B}^2(p^1, p^2)}\right)\ , \tag{80}$$

where P is the *cdf* of the random variable η^2/η^1. Its density function is denoted by ρ. Note that imposing Assumption 5 implies a multiplicative separability between the term

$$R^\sigma(p^1, p^2) = \frac{m^1(\sigma^1(p^1, p^2))}{m^2(\sigma^2(p^1, p^2))}\ , \tag{81}$$

that is shaped by the rule-utilitarian voter's participation thresholds and the ratio of their bases

$$R^{\mathbf{B}}(p^1, p^2) = \frac{\mathbf{B}^1(p^1, p^2)}{\mathbf{B}^2(p^1, p^2)}$$

so that we can write

$$\pi^1(p^1, p^2, \sigma^1, \sigma^2) = P\left(R^\sigma(p^1, p^2)\ R^{\mathbf{B}}(p^1, p^2)\right)\ . \tag{82}$$

Turnout. For now, we take the party platforms p^1 and p^2 as given and characterize the parties' equilibrium turnout. We say that the turnout game has an interior equilibrium if $0 < \sigma^{1*}(p^1, p^2) < 1$ and $0 < \sigma^{2*}(p^1, p^2) < 1$. If the function P is continuously differentiable then an interior equilibrium is characterized by the first order conditions

$$\pi^1_{\sigma^1}(\cdot)\; W^1 - \kappa\; q^{1u}\; \mathbf{B}^1 = 0\;, \tag{83}$$

and

$$-\;\pi^1_{\sigma^2}(\cdot)\; W^2 - \kappa\; q^{2u}\; \mathbf{B}^2 = 0\;. \tag{84}$$

Using Assumption 5 we can rewrite these conditions as

$$\frac{\rho(\cdot) R^\sigma(p^1,p^2)}{q^{1v} + \sigma^1\; q^{1u}}\; W^1 - \kappa\; \mathbf{B}^1 = 0\;, \tag{85}$$

and

$$\frac{\rho(\cdot) R^\sigma(p^1,p^2)}{q^{2v} + \sigma^2\; q^{2u}}\; W^2 - \kappa\; \mathbf{B}^2 = 0\;. \tag{86}$$

Equations (85) and (86) imply that

$$R^\sigma(p^1,p^2) = \frac{W^1/\kappa\; \mathbf{B}^1}{W^2/\kappa\; \mathbf{B}^2} = \frac{W^1/\mathbf{B}^1}{W^2/\; \mathbf{B}^2}\;, \tag{87}$$

which is the same expression as (7) in the body of the text.

Probability of winning. Suppose first that parties seek to maximize the probability of winning, i.e.,

$$P\left(R^\sigma(p^1,p^2)\; \frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\right)$$

and party 2 seeks to minimize this expression. As P is a non-decreasing function we can as well assume that party 1 maximizes

$$R^\sigma(p^1,p^2)\; \frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}$$

or any monotone transformation of it such as, e.g.,

$$\ln\left(R^\sigma(p^1,p^2)\right) + \ln\left(\mathbf{B}^1(p^1,p^2)\right) - \ln\left(\mathbf{B}^2(p^1,p^2)\right)\;. \tag{88}$$

Remark 1. The "conventional" probabilistic voting model can be viewed as a special case of this that is defined by two properties. First, since turnout is exogenous and universal, $R^\sigma(p^1,p^2) = 1$ for all (p^1,p^2) and hence $\ln\left(R^\sigma(p^1,p^2)\right) = 0$. Second, and again for the reason that turnout is exogenous and universal, $V^1 = \mathbf{B}^1(p^1,p^2)$ and

$V^2 = \mathbf{B}^2(p^1,p^2) = 1 - \mathbf{B}^1(p^1,p^2)$. In the probabilistic voting model, the objective of party 1 can therefore be taken to be $\ln(\mathbf{B}^1(p^1,p^2)) - \ln(1 - \mathbf{B}^1(p^1,p^2))$ or simply $V^1 = \mathbf{B}^1(p^1,p^2)$. I.e., maximizing the probability of winning is the same as maximizing the number of votes.

Remark 2. With Nash equilibrium rather than subgame perfect equilibrium as the solution concept (or, equivalently, with $\mu = 0$), the parties view $R^\sigma(p^1,p^2)$ as exogenously fixed, albeit at the level that is induced by the equilibrium policies. Party 1 then seeks to maximize

$$\ln\left(\mathbf{B}^1(p^1,p^2)\right) - \ln\left(\mathbf{B}^2(p^1,p^2)\right)$$

and party 2 seeks to minimize this expression. Since $\mathbf{B}^2(p^1,p^2) = 1 - \mathbf{B}^1(p^1,p^2)$, party 1's objective can as well simply taken to be $\mathbf{B}^1(p^1,p^2)$ and $\mathbf{B}^2(p^1,p^2)$ can be taken to be the objective of party 2. Nash equilibrium then requires that p^1 solves $\max_{\hat{p}^1 \in P}\ \mathbf{B}^1(\hat{p}^1,p^2)$ and that p^2 solves $\max_{\hat{p}^2 \in P}\ \mathbf{B}^2(p^1,\hat{p}^2)$. Note that these equilibrium are also the equilibrium conditions in the "conventional" probabilistic voting model. Thus, equilibrium existence in the "conventional" probabilistic voting model implies the existence of a Nash equilibrium in the given setup.

If the turnout subgame has an interior equilibrium, then the probability of winning for party 1 can be written in a reduced form that no longer involves an explicit reference to the participation thresholds σ^1 and σ^2. Specifically, equation (87) implies that the winning probability in (82) becomes

$$\bar{\pi}^1(p^1,p^2) = P\left(\psi^1(p^1,p^2)\right) \quad \text{for} \quad \psi^1(p^1,p^2) := \frac{W^1(p^1,p^2)}{W^2(p^1,p^2)}\,. \tag{89}$$

Thus, as in the main body of the text (Proposition 1), under Assumption 5, if (p^1,p^2) is a pair of interior subgame perfect equilibrium policies, then it it is a saddle point of the function ψ.

F.2 Public goods

Our framework for studying endogenous turnout and endogenous platforms in political competition is developed for a generic policy domain. We have emphasized that the set of non-linear income tax systems is a policy domain of particular interest. That said, our framework can also be applied to study the implications of endogenous turnout for political competition over other policy domains. In this section, we briefly summarize the results from such an analysis. Specifically, we report on the implications of our framework for public goods provision.

Individuals have preferences over public goods that are given by $u(\omega, p) = \omega\, p - k(p)$, where $p \in \mathbb{R}_+$ denotes the quantity of the public good, $\omega \in \Omega$ is an individual's public goods preference and the cost function k captures the per capita cost of public goods provision.[25] We begin with a characterization of the public good provision level that party 1 would choose if its sole objective was to mobilize its supporters. In this case, it would choose q^1 with the objective to maximize

$$W^1(p^1, p^2) = \mathbb{E}[G^1_W(\omega\, p^1 - k(p^1) - u(p^2, \omega) \mid \omega)] \; .$$

Given p^2, the first order condition characterizing the optimal choice of p^1 is

$$\mathbb{E}\left[\mathcal{G}^1_W(\omega \mid p^1, p^2)\; \omega\right] = k'(p^1)$$

where

$$\mathcal{G}^1_W(\omega \mid p^1, p^2) = \frac{g^1_W(u(p^1, \omega) - u(p^2, \omega) \mid \omega)}{\mathbb{E}[g^1_W(u(p^1, \omega) - u(p^2, \omega) \mid \omega)]} \; .$$

This first order condition is a political economy analogue to the Samuelson rule for first-best public good provision. For the given setup, the Samuelson rule stipulates that $\mathbb{E}[\omega] = k'(p)$, i.e., it requires equal weights for all public goods preferences. For the purpose of mobilizing its supporters, party 1 does not apply equal weights. Instead the public good preferences of different individuals are weighted according to the function $\mathcal{G}^1_W$. The public good provision level that party 1 would choose if it only wanted only to demobilize the supporters of party 2 is such that

$$\mathbb{E}\left[\mathcal{G}^2_W(\omega \mid p^1, p^2)\; \omega\right] = k'(p^1) \; ,$$

and the policy that maximizes $\frac{W^1(p^1 p^2)}{W^2(p^1, p^2)}$ satisfies

$$\mathbb{E}[\mathcal{G}^1_{SP}(\omega \mid p^1, p^2)\; \omega] = k'(p^1)$$

[25] In an economy with a continuum of individuals and private information on public goods preferences, equal cost sharing is the only way of satisfying robust incentive compatibility, budget balance and anonymity, see Bierbrauer and Hellwig (2016).

where

$$\mathcal{G}^1_{SP}(\omega \mid p^1, p^2) := \lambda^1(p^1, p^2)\, \mathcal{G}^1_W(\omega \mid p^1, p^2) + (1 - \lambda^1(p^1, p^2))\, \mathcal{G}^2_W(\omega \mid p^1, p^2)$$

and $\lambda^1(p^1, p^2)$ are defined as above. Again, the party compromises between mobilizing its own supporters and demobilizing the supporters of the other party – with the weight on the own supporters being smaller if the party is more likely to win.

F.3 Alternative modelling choices for ethical voting

The ethical voter models by Feddersen and Sandroni (2006), on the one hand, and by Coate and Conlin (2004), on the other differ, in some aspects. For instance, Feddersen and Sandroni (2006) assume that the population consists of ethical voters and of non-ethical voters. Moreover, the fraction of ethical voters is a priori uncertain. Uncertainty over election outcomes in Feddersen and Sandroni (2006) is entirely due to this uncertainty about the fraction of ethical voters. Coate and Conlin (2004), by contrast, assume that all voters are ethical voters. Uncertainty over election outcomes in their framework is driven by uncertainty over the policy preferences of these ethical voters.

In this section of the Online-Appendix we show that these modelling choices are not essential for our main results. We could go either way. In the main text, we present an analysis that adopts the framework of Feddersen and Sandroni (2006). We show that we could as well work with the model of Coate and Conlin (2004) in Section F.3.2.

For tractability, our adaptation of Feddersen and Sandroni makes use of an assumption which implies that the parties' bases add up to a constant. An advantage is that it becomes transparent that the standard probabilistic voting model is nested as a special case of our analysis. In Section F.3.3 we present an extension that does not rest on this assumption. The extension shows that the parties' trade-off between attracting swing voters, mobilizing their own core voters and demobilizing the opponent's core voters is at the heart of our analysis, with or without the assumption that the parties' bases add up to a constant.

The bottom line is that what is really driving our results is the combination of probabilistic and ethical voting. We use the probabilistic voting model to determine preferences over policies and parties. We use a model of ethical voting to determine turnout. How exactly we model ethical voting is of secondary importance. Our analysis is robust to alternative specifications of ethical voting.

F.3.1 A general framework

We begin with a general framework for political competition that connects probabilistic and ethical voting. As will become clear, the general framework contains as special cases

- an environment where all voters are ethical voters and with uncertainty about policy preferences as in Coate and Conlin (2004),
- an environment where the population share of ethical voters is a random quantity as in Feddersen and Sandroni (2006).

Party and policy preferences. Consider a pair of policies p^1 and p^2 proposed by parties 1 and 2, respectively. As in the body of the text, a type ω-individual preferes a victory of party 1 if

$$u(p^1, \omega) - u(p^2, \omega) \geq \varepsilon \ ,$$

where the utility function u captures policy preferences and the variable ε captures idiosyncratic party preferences. We assume that, conditional on ω, ε is a random variable with a distribution that can be represented by a cumulative distribution function $\tilde{B}(\cdot \mid \omega, \eta)$. Thus,

$$\tilde{B}(u(p^1, \omega) - u(p^2, \omega) \mid \omega, \eta)$$

is the fraction of type ω-voters who are better off if party 1 wins. The complement

$$1 - \tilde{B}(u(p^1, \omega) - u(p^2, \omega) \mid \omega, \eta)$$

is the fraction of type ω-voters who are better off if party 2 wins.

The formulation differs from the one in the main text in that we allow these distributions to be random objects themselves. The distributions of idiosyncratic party preferences now depend on the realization of a random variable η. Thus, we allow for uncertainty in policy preferences as in Coate and Conlin (2004).

Example: Aggregate uncertainty on preferences. At this stage there is no need to introduce more specific assumptions about η. Still, the following example might be helpful to get an idea of what the randomness in party and policy preferences might look like: For any type ω, there is a set of feasible distributions $\Phi(\omega)$, with generic entry $\tilde{B}(\cdot \mid \omega, \eta)$. Distributions in this set can be ordered according to first order stochastic dominance. Let this order be represented by a mapping from the unit interval to the set of feasible distributions. Also suppose that there is a random variable η_ω taking values in the unit interval, indicating which of these distributions materializes. Finally, let there be one such a random variable for each type ω. Then the random variable η that governs the *state of the system* is a stochastic process that can be written as $\eta = \{\eta_\omega\}_{\omega \in \Omega}$.

In Feddersen and Sadroni (2006), by contrast, party and policy preference are deterministically fixed once the alternatives p^1 and p^2 are given. The following assumption contains a more formal version of this statement.

Assumption 6 (Feddersen and Sandroni: No aggregate uncertainty on preferences). *For every type ω, there exists a cumulative distribution function $B(\cdot \mid \omega)$, so that, for all p^1 and p^2 and all possible realizations of η,*

$$\tilde{B}(u(p^1, \omega) - u(p^2, \omega) \mid \omega, \eta) = B(u(p^1, \omega) - u(p^2, \omega) \mid \omega)\ .$$

Ethical and non-ethical voters. A complete description of the state of the system also requires to specify, for each type, the fraction of ethical voters. Formally, let $\tilde{q}^1(\omega, \eta)$ be the fraction of ethical voters among those type ω-individuals who are better off if party 1 wins. Likewise denote by $\tilde{q}^2(\omega, \eta)$ be the fraction of ethical type ω-individuals who are better off if party 2 wins. In the approach of Feddersen and Sandroni these are random objects. Here, we capture this again, through the dependence on an aggregate shock, or, more specifically, the random variable η. By contrast, in the model of Coate and Conlin, $\tilde{q}^1$ and $\tilde{q}^1$ are set equal to one. For ease of reference, we also highlight this assumption.

Assumption 7 (Coate and Conlin: All voters are ethical voters). *For any ω, $\tilde{q}^1(\omega, \eta)$ and $\tilde{q}^2(\omega, \eta)$ are degenerate random variables so that*

$$\tilde{q}^1(\omega, \eta) = \tilde{q}^2(\omega, \eta) = 1$$

for all realizations of η.

The parties' bases. The potential voters of party 1 are those who vote for party 1 in case of turning out to vote. This mass of these voters is a random variable

$$\tilde{\mathbf{B}}^1(p^1, p^2, \eta) := \mathbb{E}\left[\tilde{q}^1(\omega, \eta)\ \tilde{B}(u(p^1, \omega) - u(p^2, \omega) \mid \omega, \eta)\right]$$

Analogously, the mass party 2's potential voters is given by

$$\tilde{\mathbf{B}}^2(p^1, p^2, \eta) := \mathbb{E}\left[\tilde{q}^2(\omega, \eta)(1 - \tilde{B}(u(p^1, \omega) - u(p^2, \omega) \mid \omega, \eta))\right]$$

We denote, respectively, by

$$\mathbf{B}^1(p^1, p^2) = \int \tilde{\mathbf{B}}^1(p^1, p^2, \eta)\ dP(\eta)$$

and

$$\mathbf{B}^2(p^1, p^2) = \int \tilde{\mathbf{B}}^2(p^1, p^2, \eta)\ dP(\eta)$$

the expected values of $\tilde{\mathbf{B}}^1(p^1, p^2, \eta)$ and $\tilde{\mathbf{B}}^2(p^1, p^2, \eta)$, where P is the cumulative distribution function of the random variable η. For brevity, we also refer to $\mathbf{B}^1(p^1, p^2)$ and $\mathbf{B}^2(p^1, p^2)$ as the parties' bases.

The turnout subgame. As in the main text, the ethical voters of party 1 choose σ^1 to maximize

$$\pi^1(\sigma^1, \sigma^2, p^1, p^2)\ W^1(p^1, p^2) - k(\sigma^1)\ \mathbf{B}^1(p^1, p^2)$$

and the ethical voters of party 2 choose σ^2 to maximize

$$(1 - \pi^1(\sigma^1, \sigma^2, p^1, p^2))\ W^2(p^1, p^2) - k(\sigma^2)\ \mathbf{B}^2(p^1, p^2)\ .$$

We have to adjust, however, our definitions of $W^1(p^1, p^2)$ and $W^2(p^1, p^2)$ so that they are consistent with the more general setup that we are currently exploring. We now let

$$\tilde{W}^1(p^1, p^2, \eta) = \mathbb{E}\left[\int_{\mathbb{R}} \max\left\{\ u(p^1, \omega) - [u(p^2, \omega) + \varepsilon]\ ,\ 0\right\}\ \tilde{b}(\varepsilon \mid \omega, \eta)\, \mathrm{d}\varepsilon\right]. \tag{90}$$

denote the stakes for the ethical voters of party 1 in state η and let $W^1(p^1, p^2)$ be the expectation of $\tilde{W}^1(p^1, p^2, \eta)$, conditional on party 1 winning the election. We denote by $\tilde{b}(\cdot \mid \omega, \eta)$ the derivative of $\tilde{B}(\cdot \mid \omega, \eta)$, i.e. $\tilde{b}(\cdot \mid \omega, \eta)$ is the density of ε, conditional on type ω and state η. We define $\tilde{W}^2(p^1, p^2, \eta)$ and $W^2(p^1, p^2)$ analogously. Party 1 wins the election if

$$\sigma^1 \, \tilde{\mathbf{B}}^1(p^1, p^2, \eta) \geq \sigma^2 \, \tilde{\mathbf{B}}^2(p^1, p^2, \eta) ,$$

where σ^1 and σ^2 are the turnout rates of the potential voters of party 1 and party 2, respectively. Equivalently, party 1 wins if

$$\frac{\sigma^1}{\sigma^2} \times \frac{\tilde{\mathbf{B}}^1(p^1, p^2, \eta)}{\tilde{\mathbf{B}}^2(p^1, p^2, \eta)} \geq 1 .$$

The probability that party 1 wins the election is therefore given by

$$\pi^1(\sigma^1, \sigma^2) = \text{prob}\left(\frac{\sigma^1}{\sigma^2} \times \frac{\tilde{\mathbf{B}}^1(p^1, p^2, \eta)}{\tilde{\mathbf{B}}^2(p^1, p^2, \eta)} \geq 1 \right) .$$

For later reference, note that we can also write this winning probability as an average winning probability over the different states η, i.e. so that

$$\bar{\pi}^1(p^1, p^2) = \int \text{prob}\left(\frac{\sigma^1}{\sigma^2} \times \frac{\tilde{\mathbf{B}}^1(p^1, p^2, \eta)}{\tilde{\mathbf{B}}^2(p^1, p^2, \eta)} \geq 1 \mid \eta \right) dP(\eta) . \tag{91}$$

Note that the turnout rates enter this expression only via the ratio $\frac{\sigma^1}{\sigma^2}$. This implies that our analysis of the turnout subgame – for given policies p^1 ans p^2 – does not depend on wether we adopt the Feddersen-Sandroni or the Coate-Conlin formulation. As a consequence, Lemma 1 in the main text goes through. Thus, irrespectively of whether Assumption 7 or Assumption 6 is imposed, in an equilibrium of the turnout subgame

$$\frac{\sigma^{1*}(p^1, p^2)}{\sigma^{2*}(p^1, p^2)} = \left[\frac{W^1(p^1, p^2) \, / \, \mathbf{B}^1(p^1, p^2)}{W^2(p^1, p^2) \, / \, \mathbf{B}^2(p^1, p^2)} \right]^{\frac{\mu}{1+\mu}} . \tag{92}$$

F.3.2 Adopting the approach of Coate and Conlin: Only ethical voters

We now impose Assumption 7, i.e. the Assumption made by Coate and Conlin (2004) that there are only ethical voters. Thus, to have non-trivial winning probabilities, we must not at the same time impose Assumption 6. Put differently, we suppose that policy preferences are subject to aggregate shocks. We will establish two findings: First, our Proposition 1 in the main text rests on a simplifying assumption on the nature of aggregate uncertainty. The same assumption can be made in the Coate and Conlin version of our model and has the same effect. Proposition 1 is therefore robust to the way in which we model ethical voting. Second, the parties' bases add up to a constant. A model of ethical voting that does not share this property therefore requires to relax Assumption 7.

Recall equations (91) and (92), i.e. that taking the endogeneity of turnout into account, the probability of winning can be written as

$$\bar{\pi}^1(p^1, p^2) = \int \text{prob}\left(\frac{\sigma^1}{\sigma^2} \times \frac{\tilde{\mathbf{B}}^1(p^1, p^2, \eta)}{\tilde{\mathbf{B}}^2(p^1, p^2, \eta)} \geq 1 \mid \eta\right) dP(\eta) \ ,$$

where

$$\frac{\sigma^1}{\sigma^2} = \left[\frac{W^1(p^1, p^2) \,/\, \mathbf{B}^1(p^1, p^2)}{W^2(p^1, p^2) \,/\, \mathbf{B}^2(p^1, p^2)}\right]^{\frac{\mu}{1+\mu}} \ .$$

In principle, there is no problem to working directly with this objective, it gives raise to the same tradeoffs as those highlighted in our manuscript. For a tractable comparative statics analysis, we would, however, have to impose (possibly non-parametric) assumptions on how different realizations of the random variable η shift the distributions $\tilde{B}(\cdot)$. Our Assumption 1 in the main text is one conceivable way of doing this, a way that has the advantage of simplicity. The main text focuses on the Feddersen-Sandroni version of ethical voting and Assumption 1 is imposed in this context. As we will now explain, we can get to same conclusions also with a Coate-Conlin approach. To see this, consider the following assumption.

Assumption 8 (Multiplicative shocks I). *Suppose that $\eta = (\eta^1, \eta^2)$ is a pair of two random variables η^1 and η^2 so that*

$$\tilde{\mathbf{B}}^1(p^1, p^2, \eta) = \eta^1 \ \mathbf{B}^1(p^1, p^2) \tag{93}$$

and

$$\tilde{\mathbf{B}}^2(p^1, p^2, \eta) = \eta^2 \ \mathbf{B}^2(p^1, p^2) \ . \tag{94}$$

Under Assumption 8 the aggregate shocks η^1 and η^2 can be interpreted as percentage deviations of the random variables $\tilde{\mathbf{B}}^1(p^1, p^2, \eta)$ and $\tilde{\mathbf{B}}^2(p^1, p^2, \eta)$ from their respective means. To see this, suppose that the means of both η^1 and η^2 are equal to

1 and rewrite (93) and (94) as

$$\frac{\tilde{\mathbf{B}}^1(p^1,p^2,\eta)-\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^1(p^1,p^2)}=\eta^1-1$$

and

$$\frac{\tilde{\mathbf{B}}^2(p^1,p^2,\eta)-\mathbf{B}^2(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}=\eta^2-1$$

The left hand sides of these equations give the percentage deviation of the random variables $\tilde{\mathbf{B}}^1(p^1,p^2,\eta)$ and $\tilde{\mathbf{B}}^2(p^1,p^2,\eta)$ from their respective means. The right-hand sides give the deviations of η^1 and η^2 from their means.

Under Assumption 8 the expression for the probability of winning in (91) becomes

$$\bar{\pi}^1(p^1,p^2)=\text{prob}\left(\left[\frac{W^1(p^1,p^2)\,/\,\mathbf{B}^1(p^1,p^2)}{W^2(p^1,p^2)\,/\,\mathbf{B}^2(p^1,p^2)}\right]^{\frac{\mu}{1+\mu}}\times\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\geq\frac{\eta^2}{\eta^1}\right) \tag{95}$$

Upon letting $\eta := \frac{\eta^2}{\eta^1}$, we can write this as

$$\bar{\pi}^1(p^1,p^2)=P\left(\left[\frac{W^1(p^1,p^2)}{W^2(p^1,p^2)}\right]^{\frac{\mu}{1+\mu}}\times\left[\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\right]^{\frac{1}{1+\mu}}\right)\,. \tag{96}$$

P is a cumulative distribution function and hence a monotonic function. Maximizing (minimizing) $\bar{\pi}^1(p^1,p^2)$ is therefore equivalent to maximizing (minimizing) the argument of P,

$$\left[\frac{W^1(p^1,p^2)}{W^2(p^1,p^2)}\right]^{\frac{\mu}{1+\mu}}\times\left[\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\right]^{\frac{1}{1+\mu}}$$

or any monotone transformation of it such as

$$\frac{1}{1+\mu}\log\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\;+\;\frac{\mu}{1+\mu}\log\frac{W^1(p^1,p^2)}{W^2(p^1,p^2)}\,.$$

We summarize these observations in the following Lemma.

Lemma 12. *Suppose that Assumptions 7 and 8 hold. Then party 1's objective is to maximize*

$$\psi^1(p^1,p^2)\;:=\;\frac{1}{1+\mu}\log\frac{\mathbf{B}^1(p^1,p^2)}{\mathbf{B}^2(p^1,p^2)}\;+\;\frac{\mu}{1+\mu}\log\frac{W^1(p^1,p^2)}{W^2(p^1,p^2)}, \tag{97}$$

and party 2's objective is to minimize it. Thus, if (p^{1*}, p^{2*}) is a pair of interior subgame perfect equilibrium policies, then it is a saddle point of the function $\psi^1(p^1, p^2)$.

Note that the Lemma gives exactly the same conclusion as Proposition 1 in the body of the text. This shows that – even though Assumptions (93) and (94) may have a more plausible microfoundation in the Feddersen-Sadroni-model – our approach is essentially agnostic on the question how to best model ethical voting. We can work both with the Coate-Conlin formulation and with the Feddersen-Sandroni formulation.

The parties' bases add up to a constant. The following Lemma shows that the Coate-Conlin specification of ethical voting shares one property of the model that we present in the main text, namely that the parties bases add up to a constant. Thus, a change of policies that increases the base for, say, party 1 translates one-for-one into a decrease of the base of party 2.

Lemma 13. *Suppose that Assumption 7 holds. Then*

$$\mathbf{B}^1(p^1, p^2) + \mathbf{B}^2(p^1, p^2) = 1 \; .$$

Proof. If $\tilde{q}^1(\omega, \eta) = \tilde{q}^2(\omega, \eta) = 1$ for all realizations of η, we have

$$\begin{aligned}
\tilde{\mathbf{B}}^1(p^1, p^2, \eta) \;&:=\; \mathbb{E}\left[\tilde{q}^1(\omega, \eta)\; \tilde{B}(u(p^1, \omega) - u(p^2, \omega) \mid \omega, \eta)\right] \\
&=\; \mathbb{E}\left[\tilde{B}(u(p^1, \omega) - u(p^2, \omega) \mid \omega, \eta)\right] \\
&=\; 1 - \mathbb{E}\left[1 - \tilde{B}(u(p^1, \omega) - u(p^2, \omega) \mid \omega, \eta)\right] \\
&=\; 1 - \mathbb{E}\left[\tilde{q}^2(\omega, \eta)(1 - \tilde{B}(u(p^1, \omega) - u(p^2, \omega) \mid \omega, \eta))\right] \\
&=\; 1 - \tilde{\mathbf{B}}^2(p^1, p^2, \eta) \; .
\end{aligned}$$

Hence, also

$$\begin{aligned}\mathbf{B}^1(p^1,p^2) &:= \textstyle\int \tilde{\mathbf{B}}^1(p^1,p^2,\eta)\; dP(\eta) \\ &= 1 - \textstyle\int \tilde{\mathbf{B}}^2(p^1,p^2,\eta)\; dP(\eta) \\ &= 1 - \mathbf{B}^2(p^1,p^2)\;.\end{aligned}$$

□

F.3.3 An alternative version of the Feddersen-Sandroni model in which the parties' bases do not add up to a constant

In the following, we consider an extension of our model in which the parties' bases do not add up to a constant. It follows from Lemma 13 that we cannot employ Assumption 7 according to which the electorate consists, in all states, entirely of ethical voters. For ease of exposition, we impose instead Assumption 6, due to Feddersen and Sandroni, so that there is no aggregate uncertainty in policy preferences. This has the expositional advantage that all the aggregate uncertainty in the model is due to the randomness of the share of ethical voters.

We seek to show that the parties' tradeoffs between attracting swing voters, catering to their own core voters in an attempt to mobilize them and catering to the rival's core voters with the intention to demobilize them does not rest on the assumption that the parties's bases add up to a constant. Recall that, in the main text, this property is implied by the assumption, that, for any type ω, the random variables $\tilde{q}^1(\omega,\eta)$ and $\tilde{q}^2(\omega,\eta)$ have the same mean

$$\bar{q}(\omega) \quad := \quad \int \tilde{q}^1(\omega,\eta)\; dP(\eta) \quad = \quad \int \tilde{q}^1(\omega,\eta)\; dP(\eta)\;.$$

Therefore,

$$\begin{aligned}
\mathbf{B}^1(p^1,p^2) &= \int \tilde{\mathbf{B}}^1(p^1,p^2,\eta)\, dP(\eta) r \\
&= \int \mathbb{E}[\tilde{q}^1(\omega,\eta)\; B(u(p^1,\omega)-u(p^2,\omega) \mid \omega)]\, dP(\eta) \\
&= \mathbb{E}\left[\left(\int \tilde{q}^1(\omega,\eta)\; dP(\eta)\right) B(u(p^1,\omega)-u(p^2,\omega) \mid \omega)\right] \\
&= \mathbb{E}\left[\bar{q}(\omega) B(u(p^1,\omega)-u(p^2,\omega) \mid \omega)\right] \;,
\end{aligned} \tag{98}$$

and, by the same logic,

$$\mathbf{B}^2(p^1,p^2) \;=\; \mathbb{E}\left[\bar{q}(\omega)(1-B(u(p^1,\omega)-u(p^2,\omega) \mid \omega))\right] \;. \tag{99}$$

Obviously, equations (98) and (99) imply that

$$\mathbf{B}^1(p^1,p^2)+\mathbf{B}^2(p^1,p^2) = E[\bar{q}(\omega)]$$

so that the two bases add up to an exogenous constant $E[\bar{q}(\omega)]$, i.e. a term that does not depend on the policies that are proposed.

Example. As a simple case that avoids the property that the parties bases add up to a constant consider the following Assumption.

Assumption 9 (Party specific means). *There are numbers $\bar{q}^1$ and $\bar{q}^2$ so that, for all ω,*

$$\bar{q}^1 \;=\; \int \tilde{q}^1(\omega,\eta)\; dP(\eta) \quad \textit{and} \quad \bar{q}^2 \;=\; \int \tilde{q}^2(\omega,\eta)\; dP(\eta)\;.$$

The assumption says, all supporters of party 1 are, irrespective of their type ω, equally likely to be of the ethical type: For any supporter of party 1, this probability is equal to $\bar{q}^1$. Likewise, all supporters of party 2 are of the ethical type with probability $\bar{q}^2$.

An implication of this Assumption is that

$$\mathbf{B}^1(p^1,p^2) = \bar{q}^1\; \mathbb{E}[B(u(p^1,\omega)-u(p^2,\omega) \mid \omega)]$$

and

$$\mathbf{B}^2(p^1,p^2) = \bar{q}^2\; \mathbb{E}[1-B(u(p^1,\omega)-u(p^2,\omega) \mid \omega)]\;.$$

Hence,

$$\mathbf{B}^1(p^1,p^2) + \mathbf{B}^2(p^1,p^2) = \bar{q}^2 + \left(\bar{q}^1 - \bar{q}^2\right) \mathbb{E}[B(u(p^1,\omega) - u(p^2,\omega) \mid \omega)] \; ,$$

which implies that the bases add up to a quantity that depends on p^1 and p^2. The overall mass of potential voters therefore does depend on the policies that the parties. Also note that

$$\begin{aligned}
\mathbf{B}^2(p^1,p^2) &= \bar{q}^2 - \bar{q}^2\mathbb{E}[B(u(p^1,\omega) - u(p^2,\omega) \mid \omega)] \\
&= \bar{q}^2 - \tfrac{\bar{q}^2}{\bar{q}^1}\, \bar{q}^1\, \mathbb{E}[B(u(p^1,\omega) - u(p^2,\omega) \mid \omega)] \\
&= \bar{q}^2 - \tfrac{\bar{q}^2}{\bar{q}^1}\, \mathbf{B}^1(p^1,p^2).
\end{aligned}$$

Note that it is still the case that an increase of party 1's base implies a decrease of party 2's base – even though no longer one-by-one.

Henceforth and in parallel to our previous analysis we impose an assumption of multiplicative shocks. This assumption of multiplicative shocks is consistent with Assumption 9, i.e. both assumptions can hold simultaneously, but does not require it. That is, we can have multiplicative shocks without party specific means.

Assumption 10 (Multiplicative shocks II). *Let $\bar{q}^1(\omega) := \int \tilde{q}^1(\omega,\eta)dP(\eta)$ be the expected value of the random variable $\tilde{q}^1(\omega,\eta)$ for any ω. Analogously, let $\bar{q}^2(\omega) := \int \tilde{q}^2(\omega,\eta)dP(\eta)$ be the expected value of the random variable $\tilde{q}^2(\omega,\eta)$. Suppose that $\eta = (\eta^1,\eta^2)$ is a pair of two random variables η^1 and η^2 so that, for all ω,*

$$\tilde{q}^1(\omega,\eta) = \eta^1\, \bar{q}^1(\omega) \tag{100}$$

and

$$\tilde{q}^2(\omega,\eta) = \eta^2\, \bar{q}^2(\omega) \; . \tag{101}$$

Note the following implications of this Assumption:

$$\begin{aligned}
\tilde{\mathbf{B}}^1(p^1,p^2,\eta) &= \mathbb{E}[\tilde{q}^1(\omega,\eta)B(u(p^1,\omega) - u(p^2,\omega) \mid \omega)] \\
&= \eta^1\mathbb{E}[\bar{q}^1(\omega)B(u(p^1,\omega) - u(p^2,\omega) \mid \omega)] \\
&= \eta^1\, \mathbf{B}^1(p^1,p^2)
\end{aligned}$$

and, analogously,

$$\tilde{\mathbf{B}}^2(p^1, p^2, \eta) \;\; = \;\; \eta^2 \, \mathbf{B}^2(p^1, p^2) \; .$$

This shows that equations (93) and (94) – imposed previously in our analysis of the Coate and Conlin model – also hold in the given context. An immediate implication is that Lemma 12 also extends to the given setup. This observation yields the following Corollary.

Corollary 4. *Suppose that Assumption 10 holds. Then party 1's objective is to maximize*

$$\psi^1(p^1, p^2) \;\; := \;\; \frac{1}{1+\mu} \log \frac{\mathbf{B}^1(p^1, p^2)}{\mathbf{B}^2(p^1, p^2)} \; + \; \frac{\mu}{1+\mu} \log \frac{W^1(p^1, p^2)}{W^2(p^1, p^2)}, \tag{102}$$

and party 2's objective is to minimize it. Thus, if (p^{1*}, p^{2*}) is a pair of interior subgame perfect equilibrium policies, then it is a saddle point of the function $\psi^1(p^1, p^2)$.

The significance of the Corollary is to show that Proposition 1 in our main text also extends to a model in which the parties' bases do not add up to a constant. Thus, the tradeoffs that we highlight in our main text also extend to the given setup, albeit with some modifications. To understand these modifications, it is again instructive to look first at the polar cases $\mu = \infty$ and $\mu = 0$.

For $\mu = \infty$, the parties' bases do not matter at all for the probability of winning the election. The analysis therefore has exactly the same logic as the one presented in the body of the text: Party 1 focuses on maximizing

$$\frac{W^1(p^1, p^2)}{W^2(p^1, p^2)}$$

and party 2 seeks to minimize this expression. From the perspective of party 1, the numerator $W^1(p^1, p^2)$ points to the political returns from increasing the stakes for its own core voters, the denominator points to the political returns from decreasing the stakes for party 2's core voters. Moreover, how these motives balance depends on the equilibrium value of $W^1(p^1, p^2)/W^2(p^1, p^2)$. The larger this quantity, the larger is party 1's equilibrium probability of winning and the more it has an incentive to focus on the demobilization of the potential voters of party 2.[26]

[26] Recall from the analysis in the main text that any equilibrium is symmetric and that this observation makes it possible to pin down the equilibrium value of $W^1(p^1, p^2)/W^2(p^1, p^2)$.

The case $\mu = 0$ is the exact mirror image. The stakes for the parties' core voters play no role, and all that matters is the ratio of the parties bases. Party 1 now seeks to maximize

$$\frac{\mathbf{B}^1(p^1, p^2)}{\mathbf{B}^2(p^1, p^2)}$$

while party 2 minimizes this expression. This problem of party 1 problem is – contrary to our analysis in the main text – not generally equivalent to maximizing $\mathbf{B}^1(p^1, p^2)$.

Remark 3. This equivalence holds, however, if we impose, in addition, Assumption 9. To see this, note that in this case,

$$\frac{\mathbf{B}^1(p^1, p^2)}{\mathbf{B}^2(p^1, p^2)} = \frac{\mathbf{B}^1(p^1, p^2)}{\bar{q}^2 - \frac{\bar{q}^2}{\bar{q}^1}\,\mathbf{B}^1(p^1, p^2)}\ ,$$

which is an expression that is increasing in $\mathbf{B}^1(p^1, p^2)$.

If the equivalence does not hold, party 1 faces a tradeoff between maximizing $\mathbf{B}^1(p^1, p^2)$ and minimizing $\mathbf{B}^2(p^1, p^2)$. Maximizing $\mathbf{B}^1(p^1, p^2)$ would mean to cater primarily to those voters who are likely to swing into the base of party 1. Minimizing $\mathbf{B}^2(p^1, p^2)$ would give priority to those voters who swing out of the base of party 2 if party 1 offers a better deal. Since the bases do not add up to a constant, those who swing out of the base of party 2 do not automatically swing into the base of party 1. Thus, there is again a tradeoff between doing something that is good for the own vote share and doing something that is bad for the rival's vote share. How this tradeoff is resolved depends, again, on the equilibrium value of $\mathbf{B}^1(p^1, p^2)/\mathbf{B}^2(p^1, p^2)$. The larger this value the larger the weight on the minimization of the rival's base.

Obviously, for values of μ that are interior, $\mu \in (0, \infty)$ both forces are at play, and the parties consider the implications of their platforms choices both for their relative base advantage, as measured by $\mathbf{B}^1(p^1, p^2)/\mathbf{B}^2(p^1, p^2)$, and for their relative stake advantage, as measured by $W^1(p^1, p^2)/W^2(p^1, p^2)$.

To summarize this discussion we highlight two observations: First, a more general model in which the parties bases do not add up to a constant gives rise to the same tradeoffs as our analysis in the main text, but possibly with some modifications in the relevant formulas. Second, if we impose an additional assumption, Assumption 9, then no such modifications are needed and the analysis in the main text literally extends – even though the parties' bases do not add up to a constant.

References

Bierbrauer, Felix J and Martin F Hellwig, "Robustly coalition-proof incentive mechanisms for public good provision are voting mechanisms and vice versa," *The Review of Economic Studies*, 2016, *83* (4), 1440–1464.

Brett, C. and J.A. Weymark, "Voting over selfishly optimal nonlinear income tax schedules," *Games and Economic Behavior*, 2017, *101*, 172–188.

Diamond, P.A., "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 1998, *88*, 83–95.

Guesnerie, R., *A Contribution to the Pure Theory of Taxation*, Cambridge University Press, 1995.

Hammond, P., "Straightforward Individual Incentive Compatibility in Large Economies," *Review of Economic Studies*, 1979, *46*, 263–282.

Hellwig, M.F., "A Contribution to the Theory of Optimal Utilitarian Income Taxation," *Journal of Public Economics*, 2007, *91*, 1449–1477.

Myerson, Roger B., "Optimal Auction Design," *Mathematics of Operations Research*, 1981, *6* (1), 58–73.