Essays in Public Economics

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Emanuel Hansen
aus Frankfurt am Main

Bonn 2014

Dekan:	Prof. Dr. Klaus Sandmann
Erstberichterstatter:	Prof. Martin Hellwig, Ph.D.
Zweitberichterstatter:	Prof. Dr. Felix Bierbrauer

Tag der mündlichen Prüfung: 13.08.2014

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

Acknowledgments

For the past few years, this doctoral thesis has played a significant role in my life and, consequently, the lifes of my family and friends. I massively benefited from all facets of working on this project, and I strongly enjoyed putting effort into this work (most of the time). This endeavor was conducted at three places, Bonn, London, and Cologne, and would never have succeeded without a number of important and impressive people to whom I am deeply grateful.

First, I want to thank my supervisor Martin Hellwig for his unconfined support, his indispensable expertise and his throughout clear and critical feedback. I am also deeply grateful to Felix Bierbrauer, my second advisor and mentor, for guiding and pushing me through this project with an admirable mixture of ambition and patience, for spending so much effort in an uncountable number of discussions at all stages of this project and for providing crucial input on all parts of this thesis, especially on the third chapter. I also benefited from invaluable input and advise by Gilat Levy, who advised me during my time at the London School of Economics, and by Dezsö Szalay. I am also grateful to Urs Schweizer, Silke Kinzig and Pamela Mertens for providing so many opportunities and resources, and for making the Bonn Graduate School a great place for doing a PhD. I benefited very much from the support by Monika Stimpson at MPI. Moreover, I am deeply thankful for the great input and enduring impact by Kai Arzheimer, Jürgen Falter, Martin Kolmar, Edeltraud Roller, Karlhans Sauernheimer, and Harald Schoen a few years earlier at Mainz University.

Doing the PhD would never have been such an enriching and entertaining experience without my friends and colleagues at the BGSE and in Cologne. Special thanks go to my co-authors for the second chapter of this dissertation, Andreas Grunewald and Gert Pönitzsch, who also provided important scientific and personal input on many other matters and became close friends. I learned a lot from these guys, and I always enjoyed the joint struggle for meeting and missing the next internal or external deadline. Rafael Aigner, Mark Le Quement, Sina Litterscheid, Désirée Rückert, Dominik Sachs and Felix Wellschmied provided crucial input to this theses by reading, listening to and commenting on my research projects. I am also deeply grateful to Burcu Düzgün, Mara Ewers, Markus

Fels, Dirk Foremny, Michael Hewer, Ulrich Homm, Matthias Lang, Paul Schempp, Matthias Schön, Philipp Strack, Martin Stürmer, Stefan Terstiege, Volker Tjaden, and Venuga Yokeeswaran for providing me with their company, their thoughts and laughs throughout this long PhD journey.

I am deeply indebted for their unlimited love, support and patience to my parents, Laura, Benny and, most importantly, to Daniela and Anouk, who made all this possible.

Contents

List of Figures

List of Tables

Introduction

This thesis consists of three self-contained chapters. Although these chapters study different questions and contribute to distinct branches of the literature, they are related with respect to the research questions and the applied methods. First, all chapters represent contributions to public economics, reflecting my research interest in a better understanding of the interdependence between the economic and the political spheres. The first two chapters concentrate on the subfield of political economics. They study political competition, i.e., the strategic interaction of politicians and citizen that represents the basis of all political decision-making in democratic systems. The incentives created and the results brought about by the political process determine why, how, and when political decision-makers intervene in the economic sphere. The third chapter contributes to the theory of optimal income taxation. The analysis thus studies and evaluates the economic effects of one of the most visible and controversial types of political interventions, but leaves aside the political decision-making process.

Second, despite many differences, the results of all chapters are derived using theoretical models. More precisely, all chapters apply microeconomic theory, with a particular focus on game theory and information economics. Additionally, the second chapter includes an empirical analysis that allows to confront its theoretic results with real-world observations.

Chapter 1 The first chapter of this thesis contributes to the economic theory of electoral competition. In contrast to most of the previous literature in this field, it studies political competition between endogenously formed parties instead of independent candidates. In the model, party formation allows policy-motivated citizens to nominate one of their fellow party members as their candidate for a general election and to share the cost of running in this election. Thus, like-minded citizens are able to coordinate their political behavior in order to improve the policy outcome. The chapter investigates the properties of stable parties and the policy platforms offered by these parties in equilibrium. It focuses on political equilibria with two active parties, which exist for all levels of membership cost and electoral uncertainty. The equilibrium platforms of both parties can neither be fully convergent as in the median voter model (Downs, 1957) nor extremely polarized as in the citizen candidate model (Besley and Coate, 1997). In the benchmark case of full electoral certainty, a unique political equilibrium with positive platform distance exists. Endogenous party formation thus eliminates a major weakness of the citizen candidate model, the extreme multiplicity of equilibria. The model remains tractable, and the qualitative results are shown to be robust under the assumption of electoral uncertainty, where vote results cannot be perfectly predicted.

Chapter 2 The second chapter of the thesis is a slightly modified version of a joint paper with Andreas Grunewald and Gert Pönitzsch (Grunewald, Hansen, and Pönitzsch, 2013). It contributes to a growing literature on political selection and its failure due to informational asymmetries, i.e., on the capability of choosing qualified political candidates by means of public elections. The chapter investigates whether the quality of political selection can be improved through political institutions and, specifically, through variations in the concentration of political power. In our model, candidates are privately informed about their abilities and driven by office rents as well as welfare considerations. We show that variations in power concentration involve a trade-off. On the one hand, higher concentration of power enables the voters' preferred politician to enforce larger parts of his agenda. On the other hand, higher power concentration increases electoral stakes and thereby induces stronger policy distortions. We identify a negative relation between the optimal level of power concentration and the extent of office motivation. In particular, full concentration of power is only desirable if politicians are prevalently welfare-oriented. The results of an empirical analysis are in line with this prediction.

Chapter 3 The third chapter of this thesis contributes to the theory of optimal income taxation. The classical result in this literature is that optimal marginal taxes are strictly positive everywhere below the top, whenever labor supply responds only at the intensive margin and the social planner holds a utilitarian desire to redistribute resources from the rich to the poor (Mirrlees, 1971). Departing from the classical framework, the third chapter of this thesis studies optimal income taxation in a model with labor supply responses at the intensive and the extensive margin. For this empirically more plausible model, it is shown that a utilitarian desire for redistribution does not pin down the signs of optimal marginal taxes and optimal participation taxes. The chapter also provides sufficient conditions for the optimality of tax schedules with negative marginal taxes and negative participation taxes for the working poor, complying with the main features of the US Earned Income Tax Credit. Furthermore, it uncovers a non-standard tradeoff between efficiency at the intensive margin and efficiency at the extensive margin, which provides the economic intuition behind the ambiguous sign of the optimal marginal tax.

1

Political Competition with Endogenous Party Formation and Citizen Activists

1.1 Introduction

This chapter studies electoral competition between two endogenously formed political parties. The agents in this model are policy-motivated citizens who are not only entitled to vote but can also join political parties. Parties serve as twofold coordination devices. First, the members of each party make monetary contributions to the parties in order to finance an exogenous (campaign) cost of running in the general election. Second, they jointly decide about the party's presidential candidate in primary elections. Parties can commit to policy platforms by nominating one of their party fellows with appropriate policy preferences as their presidential candidate. As party membership is costly, the agents will only become active if the induced policy gains resulting from this activity are sufficiently large to outweigh the cost of political activity. In this model, party platforms can be interpreted as local public goods that have to be provided and agreed on by the party members. Agents make their membership decision on the basis of the same policy preferences that also govern their voting behavior. There are two exogenous parameters, the cost of party membership and the degree of electoral uncertainty.

Most of the existing literature on political competition studies the policy platforms proposed by a set of independent candidates that do not engage in party formation. This chapter instead simultaneously investigates the characteristics and platform choices of stable political parties. In a political equilibrium, no citizen has an incentive to change his party affiliation, taking into account the effect of his deviation on the party platforms. Political equilibria can be characterized by the tuple of policy platforms offered by the parties and a partition of the set of agents into the set of independents and the membership sets of both parties. I concentrate on political equilibria with two active parties, which exist for all combinations of the exogenous parameters.

The focus of this chapter is on the effect of endogenous party formation on the equilibrium policy platforms and the implied degree of policy convergence or polarization, respectively. The main contribution is to show that the equilibrium distance between party platforms is bounded from below as well as from above. This in in contrast to the results of the citizen candidate model by Besley and Coate (1997). Intuitively, parties can only attract citizens that are willing to incur the membership cost if their platforms are sufficiently different. Thus, there can never be too much (or even full) policy convergence. On the other hand, political polarization is limited by the desire to offer competitive platforms and by the coordination enabled by political parties. If both platforms were too polarized, the members of each party would prefer to nominate a more moderate presidential candidate in order to increase the probability of winning the general election. In this situation, independent citizens with moderate policy preferences would indeed benefit from becoming politically active as the achievable policy gains would outweigh the membership cost.

The properties of political equilibria depend on the degree of electoral uncertainty and on the membership cost. As the electoral risk increases, the attractiveness of moderate platforms is weakened, and more extreme platforms can be supported in equilibrium. Put differently, if the electoral outcome becomes less predictable, the upper bound on the platform distance becomes larger while the lower bound remains unchanged. In the limiting case of full electoral certainty, both bounds coincide and a unique pair of policy platforms can be offered in equilibrium. With respect to the second exogenous parameter, both boundaries on the platform distance increase as the membership cost gets larger. Intuitively, citizens ask for more difference in the policy platforms and higher policy gains in order to be willing to engage politically. Combining these comparative statics, it can be shown that the classical prediction of full policy convergence to the median voter is only sustainable for the twofold limiting case of full electoral certainty and zero costs of political activity.

The chapter proceeds as follows. After sketching the related literature in section 1.2, the model will be presented in section 1.3. In sections 1.4 to 1.6, the game is analyzed and the main results for a given pair of the exogenous parameters are derived. In section 1.7, I present comparative statics with respect to membership costs and the degree of electoral uncertainty. For the special case of electoral certainty, the existence of a unique political equilibrium is derived. Section 1.8 concludes.

1.2 Related literature

The model builds on the citizen candidate framework introduced by Besley and Coate (1997) and Osborne and Slivinski (1996). In both versions of this model, the set of candidates is determined endogenously from the set of citizens who are not only entitled to vote in a democratic election, but can also decide to run as (individual) candidates, facing an exogenous cost of candidacy. There are no parties, and citizens cannot coordinate their political behavior. The models do not deliver a unique theoretical prediction but a multiplicity of political equilibria with either one or two candidates. Their main insight is that the endogeneity of the candidate set eliminates the possibility of completely convergent platforms in two-candidate equilibria. This impossibility result is in sharp contrast to the classical prediction of the median voter model by Downs (1957) and the probabilistic voting model by Lindbeck and Weibull (1987), but is in line with empirical observations. In both versions of the citizen candidate model, there may however be equilibria with arbitrarily polarized candidates. In the model by Besley and Coate (1997), the platform distance in two-candidate equilibria is only bound by the extremes of the policy space.[1]

A number of papers extend the basic citizen candidate framework to accommodate political parties. For example, Rivière (1999) studies the formation of parties as cost-sharing devices and provides a game-theoretical explanation for Duverger's law, i.e., the prevalence of two-party systems under the plurality rule. The same result is derived in a different environment by Osborne and Tourky (2008), who analyze the incentives to form parties within a group of legislators under the assumptions of costly participation and economies of party size. In contrast, Levy (2004) examines whether the formation of political parties can be effective in the sense that it enables offering platforms that would not be feasible

[1] In the version of Osborne and Slivinski (1996), there is large set of equilibria with potentially large, but limited polarization. In contrast to the analysis in this chapter, however, the upper bound on the platform distance results from the assumption of sincere instead of strategic voting and is not related to the candidates' behavior or coordination.

without parties. Morelli (2004) studies the implications of alternative electoral systems for the formation of parties by agents with heterogeneous policy preferences. Snyder and Ting (2002), as well as Poutvaara and Takalo (2007), show that parties may serve as brand names or screening devices, which provide superior information about the candidates' preferences or quality, respectively.

In contrast to this chapter, these papers do not examine the effects of endogenous formation of political parties on political polarization. Directly related to this issue, they do not show that party formation alleviates the (often criticized) indeterminacy of the basic citizen candidate model. Furthermore, these papers either consider only the case of electoral uncertainty or strongly restrict the type space. In this chapter, I will instead study the implications of endogenous policy formation on platform choice in a general setting, allowing for different degrees of electoral uncertainty as well as a continuum of agents without restrictions on the location of bliss points.[2]

To my knowledge, only one previous paper investigates the effect of political parties on platform choice within the citizen candidate framework. Cadigan and Janeba (2002) study party competition in a US-style presidential election with primary elections and identify a strong connection between membership structures and party platforms. Instead of endogenizing membership decisions, however, they assume exogenous party affiliations of the citizens. The drawback of this model is that any combination of platforms represents a political equilibrium for some membership structures. As they cannot distinguish between stable and unstable membership structures, the model only delivers very limited insights into the effects of party formation. Furthermore, Cadigan and Janeba (2002) do not consider the general case of electoral uncertainty.

In addition, there is a small number of papers on the formation of political parties outside the citizen candidate framework. Most closely related, Roemer (2006) studies the effects of endogenous party formation and campaign contributions by policy-motivated citizens. Similar to my model, the unique political equilibrium of Roemer's model features positive but limited platform distance. However, both models differ considerably in many aspects. Most importantly, Roemer applies the cooperative notion of "Kantian equilibrium" in which agents consider joint (proportional) deviations of all party members at the contribution stage. The implications of this equilibrium concept differ strongly from the non-cooperative notion of Nash equilibrium that I will apply in my model.[3] Furthermore, the plat-

[2] Dhillon (2004) provides an overview over the existing theoretical models with pre-election as well as post-election party formation, with a particular focus on papers that extend the citizen candidate model.

[3] For example, every citizen is member of one party in the model of Roemer (2006) while there is a (large) set of independents in any equilibrium of my model.

forms are chosen through a Nash bargaining process in which the agents' influence is proportional to their individual contributions in his model. In my model, in contrast, there are primary elections wherein each party member has exactly one vote.

In other papers, citizens only decide whether to support exogenously given political parties by contributing to their electoral campaigns (Herrera, Levine, and Martinelli, 2008; Campante, 2011; Ortuño-Ortin and Schultz, 2005). Although there is no endogenous party formation in these models, citizens have an indirect influence on policy platforms, which are chosen by the parties, taking into account the induced contribution behavior. Poutvaara (2003) also models endogenous party formation and predicts a positive but limited platform distance. However, the results are mainly driven by the assumption that agents make their membership decisions based on expressive objectives while, in my model, they follow from strategic membership decision and cooperation between like-minded citizens.[4]

Finally, this chapter also relates to the literature on probabilistic voting and electoral uncertainty, beginning with the seminal paper of Lindbeck and Weibull (1987). Eguia (2007) studies the effect of electoral uncertainty in the citizen candidate model. Without party formation, electoral uncertainty has the effect of increasing the set of political equilibria with two candidates by allowing for asymmetric equilibria. However, electoral uncertainty per se does not lead to additional centripetal forces and does not limit political polarization. Both models focus on the behavior of individual agents and do not examine the effects of party formation.

1.3 The model

There is a continuum of citizens N of mass one. The policy space X is one-dimensional and given by the real line $(-\infty, +\infty)$. The citizens have linear Euclidean preferences and heterogeneous bliss points w_i. Thus, if policy $x \in X$ is implemented, citizen i receives a policy payoff of

$$v_i(x) = -|x - w_i| . \tag{1.1}$$

[4]Besides, there exist a few models on endogenous formation of political parties under proportional electoral systems in which the implemented policy is given as a weighted sum of the party platforms (e.g. Gomberg, Marhuenda, and Ortuño-Ortin 2004; Gerber and Ortuño-Ortin 1998). Due to the incentives given by this electoral system, these models typically predict an extremely high level of political polarization.

The distribution of bliss points in the population has full support on $\mathbb{R}$, but is not known ex ante. The population median m is commonly perceived to be the realization of a random variable with twice continuously differentiable cdf Φ and pdf ϕ. In particular, I assume that m is perceived as normally distributed with mean zero and standard deviation σ.[5] As the median voter will be decisive in the general election, this assumption induces electoral uncertainty.

A general election with plurality ("winner-takes-all") rule takes place to choose a president who is entitled to implement policy. There are two parties, the leftist party L and the rightist party R. The election is party-based in the sense that only the two parties have the right to nominate presidential candidates who run against each other in the general election. In order to nominate a candidate, however, each party is required to pay an exogenous cost C of candidacy, which must be financed jointly by the members and supporters of each party. The presidential candidate of each party is determined in a series of pairwise primary elections in which all party members are entitled to stand for office and to vote. Neither a party nor a candidate is able to make a binding policy commitment prior to the general election. As will become clear later on, the bliss point of the leftist (rightist) party's candidate can consequently be interpreted as policy platform l (r).

The membership structures of both parties are not given exogenously. Instead, they follow endogenously from the citizens' optimizing behavior. Specifically, citizens choose their affiliation by making contributions $\alpha_i^P \in [0, \infty)$ to the parties $P \in \{L, R\}$. The utility of citizen $i \in N$ is linearly decreasing in his contributions, and given by

$$v_i(x) - \alpha_i^L - \alpha_i^R \tag{1.2}$$

if policy x is implemented. Agent i becomes a member of party $P \in \{L, R\}$ if and only if he contributes $\alpha_i^P \geq c$. Thus, c represents the cost of political activity, which may correspond to monetary costs, but can also be interpreted as hours worked and effort spent for the electoral campaign and party meetings. To rule out that only degenerate parties are formed in equilibrium, I assume that $c < C/2$ is satisfied. Each citizen can join one party at most.[6] The result of the

[5]The assumption of a normally distributed population mean is motivated by an extension of the central limit theorem. This theorem states that, for a sample with a sufficient number of independent and identically distributed random variables, the distribution of the sample mean approximates a normal distribution. Ma, Genton, and Parzen (2011) discuss the conditions under which the same result applies for the distribution of the sample median and other sample quantiles.

[6]This assumption simplifies the following analysis without affecting the results. It can be shown that no citizen wants to be a member of both parties in any political equilibrium. Note also that it is possible to support a party without becoming its member (for $\alpha_i^P < c$) or to con-

party formation game is a partition of the set N into the member sets of each party (M^L, M^R) and the set of independents (I) such that $N = M^L \cup M^R \cup I$.[7]

The political process consists of four stages. At the first stage, all agents $i \in N$ simultaneously choose their party affiliation by making contributions to both parties $\alpha_i^L, \alpha_i^R \geq 0$. Party P becomes active and is entitled to nominate a presidential candidate if and only if $\sum_{i \in N} \alpha_i^P \geq C$. At the second stage, a series of pairwise primary elections is conducted in each active party to select the presidential candidate. In the pairwise elections of each party, only the respective party members are entitled to vote. In the subsequent general election, the Condorcet winners of each party's primaries run as presidential candidates.[8]

At the third stage, the population median is drawn and the general election between the nominated presidential candidates takes place. All citizens observe the identities, i.e., the bliss points, of both presidential candidates and simultaneously cast their votes. The winner is determined according to the plurality rule and becomes president. If there is only one active party and presidential candidate, he directly enters the presidential office. At the last stage of the political process, the elected president implements some policy $x \in X$. The candidates are not able to make binding policy commitments at earlier stages of the political process. If there is no active party, a default policy $x_0 \in \mathbb{R}$ is implemented.

Figures 1.1 and 1.2 depict the timing of the game and its information structure graphically. At the first stage, the citizens simultaneously choose their contributions (α^L, α^R), which induce a partition of the agent set N into the membership sets of both parties M^L, M^R and the set of independents. Note that figure 1.1 only depicts two possible membership structures for each party (e.g. M_1^L, M_2^L) in order to illustrate the basic structure, although there is an infinite number of possible membership structures in general. At the first stage, each agent $i \in N$ must hold beliefs about the resulting membership structures and the platforms that would arise in case of his membership in any party as well as in case of his independence. These beliefs determine the expected effect of his political activity on his individual payoff and must be consistent in equilibrium.

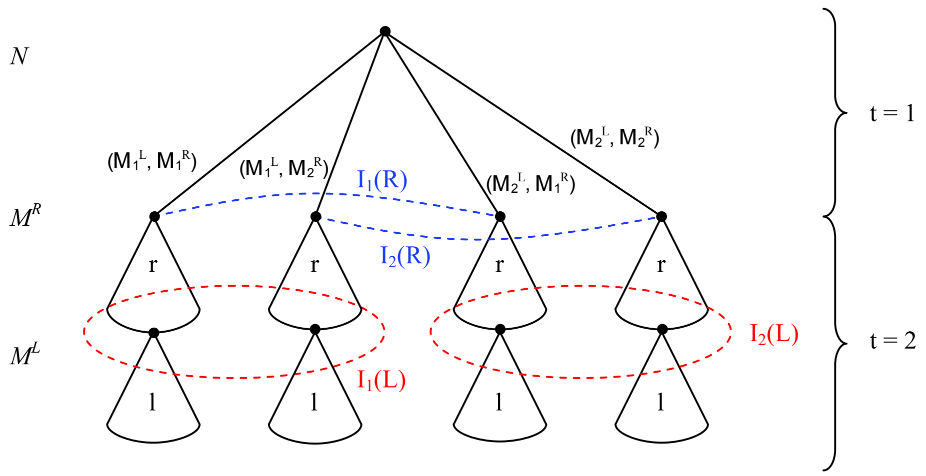
At the second stage, the members of each party jointly choose their presidential candidate and the policy platform, respectively. With respect to the information structure, I assume that at the time of candidate nomination, the members of party $P \in \{L, R\}$ can observe the set of their party fellows (M^P) and their bliss points,

tribute more than the exogenous membership cost. The additional generality of this financing structure has no effect on the result of the model.

[7] As I will show in the following sections, the member sets of both parties are finite in any political equilibrium.

[8] As shown in the following section, the existence of a Condorcet winner is guaranteed for each finite membership set.

Figure 1.1: The party formation subgame



but not the membership set of the competing party (M^{-P}). For the members of party R, all nodes involving membership set M_1^R are thus contained in the same information set $I_1(R) = \{(M^L, M^R) \mid M^R = M_1^R\}$. Similarly, the members of party L can neither observe the membership set M^R nor the chosen platform r of the rightist party when they decide about their own platform l. Rather, the information set $I_1(L) = \{(M^L, M^R) \mid M^L = M_1^L\}$ consists of all nodes involving the same membership set M_1^L, but different sets M^R and platforms r.

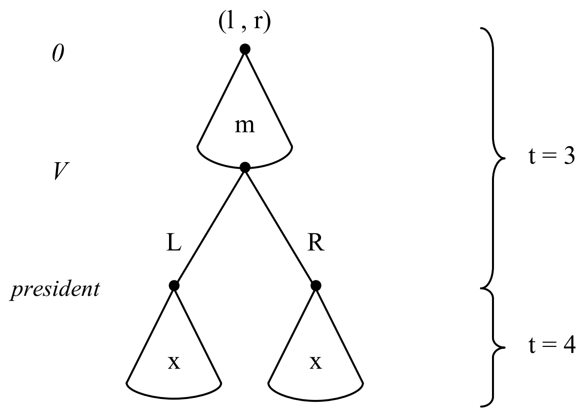
Thus, a specific form of updating takes place at the beginning of the second stage: Members of the leftist party can perfectly update their previous belief about the leftist party's membership structure M^L, while their beliefs about M^L remain constant. Consequently, the members of party L must hold a belief about M^R and the finally chosen platform r in each information set (see figure 1). In the following, I will only consider the belief $\hat{r}$ about the competing party's platform r explicitly, as this is the only payoff-relevant variable (in contrast to M^R). After the primary election stage, the nominated candidates and the associated platforms of both parties become public information, and all citizens update their beliefs $\hat{r}$ as well as $\hat{l}$. The remaining stages of the game are depicted graphically in figure 1.2 below.

This information structure simplifies the analysis while it does not change the qualitative results of the model. In particular, lower and upper bounds on the platform distance in political equilibria could also be identified under the alternative assumption that all agents can observe both member sets M^L and M^R at the primary election stage.[9]

[9]Given this information structure, the analysis of deviations from equilibrium is simplified considerably. If a previously independent agent deviates by joining party L, this will in general

An allocation is given by a tuple of party platforms (r, l) (the presidential candidates' bliss points) and a partition of the population into the sets of party members and independents. A Perfect Bayesian equilibrium of this game is given by a strategy profile and a belief system such that, first, the strategies are sequentially rational given the belief system and, second, the belief system is derived from the optimal strategies everywhere on the equilibrium path. Additionally, I assume that agents do not play weakly dominated strategies at the candidate selection stage and vote sincerely at the general election stage.[10] The goal of this chapter is to identify the set of equilibrium platform combinations and the corresponding set of stable membership structures. I concentrate on political equilibria in pure strategies with two active parties.[11] In the following, I will solve the model backwards starting with the policy implementation stage.

Figure 1.2: The general election subgame



induce a change of platform l. The members of party R cannot react to this deviation by changing their platform r, however, since they are not able to observe the deviation. In the alternative case of fully observable membership sets at the second stage, the same change in the party affiliation of one agent might induce platform changes in both parties. Due to the finite set of feasible platforms, however, the implied reaction function of the competing party would in general be discontinuous and depend strongly on the specific composition of M^R. Accounting for these best responses would thus require a large number of case distinctions.

[10] At the general election stage, the assumption of sincere voting seems innocuous. With any finite set of voters and only two alternatives, sincere voting would be the weakly dominant strategy. With a continuum of voters, the notion of weak dominance is not properly defined since no voter can ever be pivotal. The economic intuition however does not change, leaving sincere voting as the only reasonable equilibrium behavior.

[11] In general, there may also exist political equilibria in mixed strategies and equilibria in which there is only one active party with a sufficiently moderate platform (see appendix).

1.4 Policy implementation and general election

The last two stages of the game can be solved straightforwardly. At the last stage, the elected president decides which policy to implement. Assume agent i with bliss point w_i is the president. Recall that he was unable to commit to any policy before. He can thus maximize his individual payoff $v_i(x) = -|x - w_i|$ by implementing his bliss point $x = w_i$. This policy choice is anticipated by all agents at the previous stages. Thus, the nomination of agent i as presidential candidate by party L implies a (credible) commitment to his individual ideal policy w_i. In the following, I will thus interpret the ideal policies of both presidential candidates as the parties' policy platforms l and r.

At the general election stage, all citizens vote for one of the parties or one of the nominated presidential candidates, respectively. The bliss points of both candidates are known. For clarity, we denote these bliss points by l and r, as they represent the platforms offered by both parties L and R. As a convention, the party with a more leftist platform will be called party L, and its platform will be denoted by l such that $l \leq r$ holds.

All citizens vote sincerely in the general election. Thus, citizen $i \in N$ votes for the party whose platform is closer to his own bliss point w_i, and the median voter's preference prevails. Thus, the leftist party L will win the election if and only if its platform l is located more closely to the median voter's bliss point m (the population median) than platform r, i.e., if $m < \frac{l+r}{2}$ holds.

Ex ante, however, the agents do not know the exact location of the population median $m \in \mathbb{R}$, but only its probability distribution. Thus, the winning probability $p(l,r)$ of party L is equal to the value of the distribution function at the arithmetic mean of both platforms,

$$p(l,r) = \Phi\left(\frac{l+r}{2\sigma}\right). \tag{1.3}$$

Obviously, the winning probability is increasing both in l and r (for $l < r$). Besides, note that the random distribution of m induces electoral uncertainty as all agents assign positive winning probabilities to both parties for any combination of l and r ex ante. As I will show in the following section, this electoral uncertainty implies a smooth trade-off between the subjective desirability and the winning probabilities of alternative party platforms, which is in line with the economic intuition and often referred to in political discussions. To simplify notation, we focus on the case of a standard normal distribution with $\sigma = 1$ in the following.[12]

[12]In section 1.7, I study the effects of variations in electoral risk, as captured by σ.

1.5 Candidate selection

At the candidate selection stage, the members of both parties simultaneously nominate their presidential candidates. As the nomination process in both parties is completely symmetric, I will only consider intra-party decision making in the leftist party L. To avoid case distinctions, I impose the simplifying assumption that each party has an odd number of members.[13]

At this stage, both member sets M^L and M^R have been determined as the outcome of the party formation game at the first stage. By the assumed information structure, the members of party L can only observe the composition M^L of their own party (see figure 1). For each information set $I_k(L)$, however, they hold a belief $\hat{r}$ about the resulting platform of the rightist party.

The presidential candidate is selected by the members of party L in a series of pairwise elections. This procedure will lead to a clear-cut decision if and only if one member represents a Condorcet winner, i.e., if a majority of member prefer one agent $i \in M^L$ to all other potential candidates. Lemma 1.1 states that a Condorcet winner exists for any combination of member set M^L and belief $\hat{r}$.

Lemma 1.1. *Let M^L be the set of members of party L, m^L the party median and $\hat{r} \geq m^L$ the commonly held belief about party R's platform. The selected candidate of party L is given by the member with bliss point $l(M^L, \hat{r}) = \max\left\{m^L, l_M(\hat{r}, M^L)\right\}$, where $l_M(\hat{r}, M^L) \equiv \arg\max_{\{w_i : i \in M^L\}} (\hat{r} - w_i)p(w_i, \hat{r})$*

First, note that candidate selection serves only as a device to commit to the preferred platform, as the agents' utilities do not depend on the identity of the candidates. Conditional on platform l and belief $\hat{r}$, the expected policy payoff to member i of party L is given by

$$\begin{aligned} \tilde{v}_i(l, \hat{r}) &\equiv p(l, \hat{r})(-|l - w_i|) + [1 - p(l, \hat{r})](-|\hat{r} - w_i|) \\ &= p(l, \hat{r})\{|\hat{r} - w_i| - |l - w_i|\} - |\hat{r} - w_i|. \end{aligned} \tag{1.4}$$

Each member would like to choose l in order to maximize his individual policy payoff, given the expected platform of the competing party $\hat{r}$. For illustration, look at the preferences of a leftist citizen such that $w_i < \hat{r}$ holds. Obviously, he would never choose a platform $l > \hat{r}$ as this would imply an even lower policy payoff than a certain implementation of policy $\hat{r}$. Furthermore, no platform to the left of a member's bliss point can be individually optimal, since any platform $l < w_i$ leads to a lower winning probability $p(l, \hat{r})$ as well as a lower policy payoff in case

[13] For an even number of members, only minor changes occur, while the qualitative results remain valid.

of winning (compared to w_i). For platforms in the remaining interval $[w_i, \hat{r}]$, the policy payoff function simplifies to

$$\tilde{v}_i(l, \hat{r}) = p(l, \hat{r})(\hat{r} - l) - (\hat{r} - i).$$

In this interval, the platform preferences involve a trade-off between the probability of winning $p(l, r)$ and the subjective desirability $(l - w_i)$. As platform l approaches $\hat{r}$, member i benefits from an increasing winning probability of party L, but receives a lower payoff in case of electoral success. Each member prefers the platform which induces the largest shift of the expected policy towards his bliss point. In order to measure this shift, I define the policy effect function

$$\Gamma(l, \hat{r}) \equiv (\hat{r} - l)p(l, \hat{r}) = (\hat{r} - l)\Phi\left(\frac{r + l}{2}\right). \tag{1.5}$$

In the appendix, I show that this function is strictly quasi-concave for the case of a normally distributed population median m. I denote its unique maximizer by $l_\Gamma(\hat{r}) = \arg\max_{l \in \mathbb{R}} \Gamma(l, \hat{r})$. Figure 1.3 depicts the policy effect function graphically.

As the party platform must equal the bliss point of some party member $j \in M^L$, however, this platform may not be feasible. Taking this restriction into account, the feasible platform with the highest policy effect is given by $l_M(\hat{r}, M^L) = \arg \max_{\{w_i:\, i \in M^L\}} \Gamma(l, \hat{r})$. By the quasi-concavity of the policy effect function, the policy payoff of agent i is maximized by the platform $l_M(\hat{r}, M^L)$ if this is more moderate than ω_i, and by his own bliss point w_i otherwise.

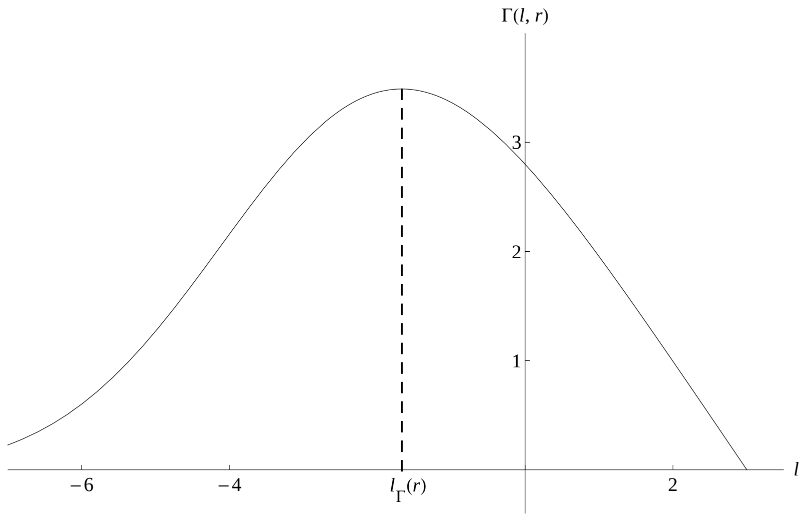
Second, I show in the appendix that the platform preferences satisfy the single-crossing property (see Lemma 1.3). Thus, voting behavior is monotonic in each pairwise election. The preferred candidate of the median party member consequently represents a Condorcet winner and is nominated as presidential candidate. As explained above, the median member prefers the maximum of his own bliss point and platform $l_M(\hat{r}$.

Note that pairwise elections are not the only decision procedure leading to the nomination of the Condorcet winner as presidential candidate. For example, the same platforms arise under the formal rule that the median party member is entitled to nominate his preferred candidate.[14] Furthermore, one could think of a richer model with US-style primary elections in which all party members are entitled to vote and to run as candidates. In such a model, the unchallenged candidature of the Condorcet winner identified above would represent a subgame equilibrium, too.[15]

[14]This decision rule is applied in the model of Poutvaara (2003).

[15]With such a primary election stage, there may be additional equilibria with two winning can-

Figure 1.3: The policy effect function



Horizontal axis: Platform l of the leftist party. Vertical axis: Policy effect function $\Gamma(l, r)$ for $r = 3, \sigma = 1$.

1.6 Political equilibria

Political equilibria can be characterized by membership structures M^L, M^R and the resulting platforms l, r. In the previous section, I identified the presidential candidates that are nominated by the members of party L in each information set, i.e., for any combination of member set M^L and belief $\hat{r}$. In a political equilibrium, the platform beliefs must be consistent. This implies that the equilibrium platform l must be the Condorcet winner in set M^L, given the correct belief $\hat{r} = r$ (accordingly for platform r). If membership structures were given exogenously by some partition (M^L, M^R), then this condition would already pin down the unique equilibrium combination of policy platforms.

At the first stage of the game studied here, however, policy-motivated citizens choose their party affiliation endogenously. In a political equilibrium, membership structures must therefore be stable in the sense that

(I) no member of any party can profitably deviate by becoming politically independent,

(II) no independent citizen can profitably deviate by joining any party,

(III) no member of any party can profitably deviate by changing his party affiliation.

didates between whom the median member m^L is indifferent.

Conditions (I) to (III) are necessary and sufficient conditions for any political equilibrium. However, they do not give many insights by themselves, as the effects of the mentioned deviations depend in a non-trivial way on the complete vector of contributions α^L, α^R and the implied membership sets M^L, M^R. In the following, I will examine the implications of these conditions on the set of policy platforms that can be supported in equilibrium. After deriving necessary conditions for political equilibria, I prove equilibrium existence.

Consider some vector of contribution decisions (α_0^L, α_0^R) and the induced membership structure M_0^L, M_0^R. Let the resulting policy platforms be given by l_0 and r_0. This constellation can only represent an equilibrium if there is no profitable deviation at the party formation stage, i.e., if no agent would benefit from changing his party affiliation. Party L is active if and only if the sum of its contributions is larger than the exogenous cost of running: $\sum_{i \in N} \alpha_i^L \geq C$. It is *efficient* if $\sum_{i \in N} \alpha_i^L \in [C, C + c)$ holds, which implies that there is no wasteful overcontribution and that the withdrawal of any member would induce the inactivity of its former party. Conditions (I) and (II) jointly lead to the following lemma.

Lemma 1.2. *In any equilibrium with two active parties, both parties are efficient, i.e.,* $\sum_{i \in N} \alpha_i^P \in [C, C + c)$ *for* $P \in \{L, R\}$.

Lemma 1.2 can be proven by contradiction. In order to do this, assume that there is a political equilibrium with non-efficient contributions. Let the party platforms be given by l_0 and r_0. In equilibrium, the members of both parties hold correct beliefs $\hat{r} = r_0$ and $\hat{l} = l_0$. Now, consider two specific deviations. First, the exit of the most leftist member j of party L would not induce L's inactivity but shift its party median to a more rightist position $m_1^L > m_0^L$. As party R cannot react to this deviation and belief $\hat{r}$ remains unchanged, the withdrawal of agent j will induce the nomination of a weakly more moderate candidate $l_1 \geq l_0$ by Lemma 1.1. Agent j will prefer to maintain his membership in L if and only if the shift from l_0 to l_1 is so large that the reduction in his policy payoff outbalances the saved membership cost.

Next, consider a more rightist, independent agent k with bliss point $w_k \in (l_1, r)$If he would join party L, this would have the same effects on the party median and, consequently, on the nominated candidate as the previously considered exit of the leftist member j. Thus, the policy platform shifts from l_0 to l_1 once again, inducing an increase of k's policy payoff. Agent k profits from this deviation if this effect outweighs the cost c of joining party L. In the appendix, I show that the payoff increase to the entrant k is strictly larger than the payoff decrease to j from leaving party L (in absolute values). Thus, whenever agent j prefers not to become independent, it is profitable for k to join party L. Since either j or k will always have an incentive to change his party affiliation, there cannot be a political equilibrium with inefficient parties.

Lemma 1.2 implies the number of party members will be smaller than $\frac{C}{c}+1$ in any political equilibrium. Consequently, the sets of members of both parties will always be finite, and there will be independent agents in any equilibrium.

Party structures can thus only be stable if the exit of any member of L causes the inactivity of his party and guarantee the implementation of the opposing platform r. Given this pivotality, agent i prefers to stay a party member if the policy gains induced by his activity outweigh the cost c of his membership. In equilibrium, this can only be true for some party members if the policy effect $\Gamma(l,r)$ of each party is sufficiently large. Furthermore, membership structures can only be stable if no independent agent has an incentive to join one of the parties. By the following proposition, each party's platform has to satisfy a set of necessary conditions, conditional on the platform of the opposing party.

Proposition 1.1. *In every political equilibrium in which party R offers platform r, the leftist platform l satisfies the following two conditions:*

(i) Moderate and extreme boundary: $l \in [\eta_1(r,c), \eta_2(r,c)]$, where both thresholds are given by both roots of function $A(l,r,c) = \Gamma(l,r) - c$ in l and satisfy $\eta_1(r,c) \leq l_\Gamma(r) \leq \eta_2(r,c)$.

(ii) Extreme boundary: $l > \lambda(r,c)$, where the threshold $\lambda(r,c)$ is given by the unique root of function $B(l,r,c) = \Gamma(l_\Gamma(r),r) - \Gamma(l,r) + 2p(l,r)(l_\Gamma(r) - l) - c$ in l and satisfies $\lambda(r,c) < l_\Gamma(r)$.

Proposition 1.1 implies that the leftist party's platform must be located in some well-defined interval, which depends on the opposing platform r. Part i is a consequence of the efficiency of parties derived in Lemma 1.2 and condition (I) on party members. No member (including the presidential candidate himself) would be willing to maintain his political activity if the activity of party L would not increase its policy payoff sufficiently strong. For every party member, the induced policy gain is weakly smaller than the policy effect function $\Gamma(l,r) = (r-l)p(l,r)$, which must exceed the membership cost c, thus. The moderate bound $\eta_2(r,c)$ follows from the necessity to have a sufficiently large platform difference $(r-l)$. No agent would be willing to bear the cost of c if the offered platforms were too similar. In particular, the positive costs of political activity eliminate the possibility of full policy convergence, the classical result due to Downs (1957). Additionally, there is an extreme boundary $\eta_1(r,c)$ since the members of party L would not be willing to support a party with negligible electoral prospects. By the quasi-concavity of the policy effect function $\Gamma(l,r)$, both boundaries are well-defined (see figure 1.3).

The second part of Proposition 1.1 follows from condition (II), according to which no independent agent must have an incentive to join a party. The extreme boundary $\lambda(r,c)$ is derived in two steps. Consider an allocation in which platform l is located to the left of the maximum effect platform $l_\Gamma(r)$. By Lemma 1.1, this platform will be chosen if and only if (a) it provides a higher policy effect than any other available platform and (b) the party median is even more extreme: $l = l_M(M^L, r) \geq m^L$. It available, the median party member would prefer to offer the platform with maximum policy effect $l_\Gamma(r)$.

If an independent agent with bliss point $w_i = l_\Gamma(r)$ were to join party L, he would thus become presidential candidate. Thus, an equilibrium with platform l' only exists if this agent cannot benefit from joining party L. On the one hand, he can clearly achieve a policy gain by joining. On the other hand, he can save the cost c and free-ride on the provision of party L by other leftist citizens by staying independent. The net gain from entering party L is given by

$$\begin{aligned} B(l,r,c) =& \tilde{v}_i\left(l_\Gamma(r),r\right) - c - \tilde{v}_i(l,r) \\ =& \underbrace{\Gamma\left[l_\Gamma(r),r\right] - p(l,r)\left[r + l - 2l_\Gamma(r)\right]}_{\text{policy gain}} - c\,. \end{aligned} \tag{1.6}$$

In any political equilibrium, $B(l,r,c)$ must be negative. Thus, platform l has to be sufficiently moderate. For values of l close enough to $l_\Gamma(r)$, the membership cost dominates the achievable policy gain. If platform l becomes more extreme, the net gain will however strictly increase for two reasons. First, as the distance between l and $l_\Gamma(r)$ increases, platform l becomes less attractive to the potential entrant. Second, the probability of party L's victory in the general election becomes smaller. Consequently, there is a unique cut-off value $\lambda(r,c)$ such that there is an incentive to deviate whenever $l \leq \lambda(r,c)$ holds. Thus, the function $\lambda(r,c)$ represents an extreme boundary for platform l, conditional on the platform of party R.[16]

As the game is completely symmetric between both political parties, corre-

[16]Note that $l > \lambda(r,c)$ is a necessary but not a sufficient condition for the stability of party L's membership structure. More exactly, one can show that agents with slightly more moderate bliss point $w_i > l_\Gamma(r)$ have an even larger incentive to join party L and still prefer to join party L in constellations with a slightly more moderate platform $l = \lambda(r,c) + \varepsilon$. While the construction of a sufficient condition is possible, it does not provide additional economic insights.

sponding necessary conditions have to be fulfilled for the equilibrium platform of the rightist party R. The following corollary recapitulates the analysis so far and identifies a set of potential political equilibria.

Corollary 1.1. *In any political equilibrium, the party platforms l and r satisfy the following necessary conditions:*

1. *Platform l of the leftist party L is located in the interval*

$$B^L(r,c) = [max\{\eta_2(r,c), \lambda(r,c)\}, \eta_1(r,c)]$$

2. *Platform r of the rightist party R is located in the interval*

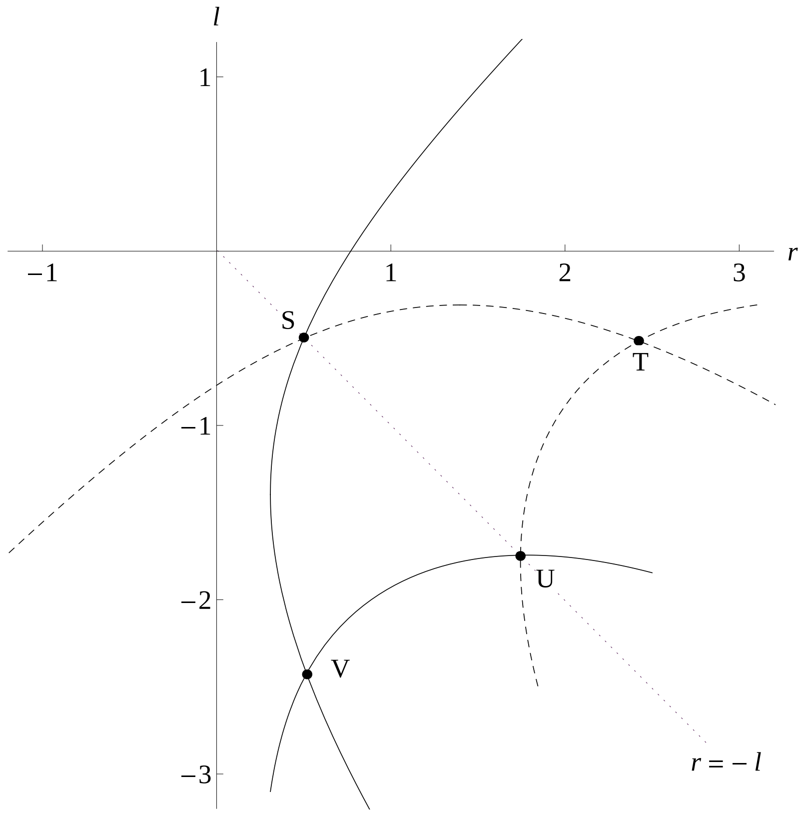
$$B^R(l,c) = [-\eta_1(-l,c), min\{\eta_2(-l,c), -\lambda(-l,c)\}]$$

Note that for any given membership structure, there is a unique reaction function $l(M^L, r)$ with respect to the platform of the competing party R. Since the party structures are not given exogenously, however, the correspondences $B^L(r,c)$ and $B^R(l,c)$ represent the collection of all reaction functions for the complete set of stable membership structures. Figure 1.6 depicts these correspondences for both both parties in a diagram with platform r on the horizontal and platform l on the vertical axis. The upper and lower bounds for platform l are given by the solid lines, the bounds for platform r by the dashed lines. Consider an allocation with any pair of platforms l and r. If the point (r,l) is not located in the area between both solid lines, platform l cannot be supported in any equilibrium, i.e., by any membership structure.

In figure 1.6, region $STUV$ represents the intersection of the correspondences $B^L(r,c)$ and $B^R(l,c)$ for the parameter values $c = 0.5$ and $\sigma = 1$. It contains the set of all tuples (l,r) that satisfy the necessary platform conditions established in Proposition 1.1. The set of political equilibria is a subset of this intersection, as the conditions identified in Proposition 1.1 are necessary, but not sufficient for equilibrium. For any combination of platforms outside this interval, in contrast, there is a profitable deviation for at least one agent.

Figure 1.6 graphically shows that the distance between both party platforms is bounded from above as well as from below for the considered example. Formally, upper and lower boundaries for the platform distance can be derived from the conditional boundary functions $\eta_2(r,c)$ and $\lambda(r,c)$ for all parameter values. In the *minimal distance equilibrium* S, both parties offer the most moderate platforms that can be supported against each other. This implies that the policy effect delivered by both platforms is exactly sufficient to cover the membership cost c. In the

Figure 1.4: Stable parties and supportable platforms



Horizontal axis: Rightist party platform r. Vertical axis: Leftist party patform l. Region $STUV$: Potential equilibrium platforms for $c = 0.5$, $\sigma = 1$.

maximum distance equilibrium U, both parties nominate the most extreme presidential candidates for which the necessary conditions in Proposition 1.1 hold. By the symmetry between both parties, the rightist party's platforms in both constellations is a fixed point of the conditional boundary function: $\underline{r}(c) = -\eta_2(\underline{r}(c), c)$ and $\bar{r}(c) = -\lambda(\bar{r}(c), c)$.

Proposition 1.2. *In every political equilibrium, the platform distance* $r - l$ *is*

(i) weakly larger than $2c > 0$*, and*

(ii) smaller than $2\bar{r}(c)$*, where* $\bar{r}(c) > c$ *is defined as the unique root of* $G(r, c) = \lambda(r, c) + r$ *in* r*.*

For part (i), note that the function $-\eta_2(r, c)$ is strictly decreasing in r. Thus, it has at most one fixed point. It is easy to show that this fixed point is given by the exogenous membership cost c.[17]

The proof for part (ii) of Proposition 1.2 is slightly more complicated. First, I show that the derivative of the maximum effect platform $l_\Gamma(r)$ (the best answer) with respect to r is always larger than -1. Intuitively, whenever the platform of party R becomes more extreme, party L would achieve a higher winning probability ceteris paribus. While the members of party L might prefer to change their platform as well, their best response will never involve a more extreme shift that would eliminate this advantage. Second, the incentives for the potential entrant with bliss point $l_\Gamma(r)$ change: his policy payoffs both in case of joining party L and in case of staying independent increase because the rightist platform r becomes less competitive. Altogether, the derivative of the extreme boundary function $-\lambda(r, c)$ in r is smaller than 1 such that there can be at most one fixed-point. Exploiting this fixed-point property of $\bar{r}$, it can finally be shown that the defining function $G(r, c) = \lambda(r, c) + r$ has a unique root for any $c \geq 0$.

Proposition 1.2 establishes the main result of this chapter and represents a qualification of the insights provided in the basic citizen candidate model (Besley and Coate, 1997). As in the citizen candidate model, there can only be limited policy convergence due to the costs of political activity. In contrast, there can only be limited polarization in my model because of the coordination possibilities provided by political parties. The following proposition establishes the existence of equilibria for all parameter constellations, ensuring the relevance of these insights.

[17] More concretely, it can be shown that either $\eta_2(r, c)$ has a unique fixed point in r or that the associated boundary $\eta_1(r, c)$ has a unique fixed point in r. In both cases, the fixed point is given at $r = c$.

Proposition 1.3. *The set of political equilibria is non-empty for all levels of the membership cost* $c \geq 0$*.*

By this proposition, platform tuples (l, r) and stable membership structures exist such that the sufficient conditions (I)-(III) are satisfied. To give an intuition for this result, consider a political constellation where $l = m^L \in [l_\Gamma(r), \eta_2(r, c)]$ and party L is efficient according to Lemma 1.2. In this situation, the policy platform l is given by the bliss point of the median party member who prefers this constellation to any other platform (see Lemma 1.1). Consequently, the offered platform will not change as long as the party median is constant. Clearly, it is possible to construct membership structures (with multiple party members that share the party median's bliss point m^L) such that m^L does not change due to the entry of any independent agent. This implies that neither an independent agent nor a current member of party R has an incentive to join party L. Moreover, if party L is efficient and the bliss points of all its member are sufficiently leftist, no member of L would benefit from becoming independent (as the moderate boundary condition $l < \eta_2(r, c)$ is satisfied). If platform r and membership set M^R satisfy equivalent conditions, no agent $i \in N$ can profitably change his party affiliation. Thus, the existence of a political equilibrium with policy platforms l and r is guaranteed.

1.7 Comparative statics

In the previous chapters, I have established the existence of political equilibria and their properties given some fixed membership cost c. Moreover, I have focused on the specific case of the standard normal distribution, $\sigma = 1$. This section investigates the effects of changes with respect to both exogenous parameters, c and σ. In particular, I am interested in the effects on the boundary functions $\underline{r}(c, \sigma)$, $\bar{r}(c, \sigma)$ and the implications for equilibrium platform distance.[18] First, I consider variations in the membership cost c, a crucial ingredient of the citizen candidate framework.

Proposition 1.4. *The minimal distance boundary* $\underline{r}(c, \sigma)$ *and the maximal distance boundary* $\bar{r}(c, \sigma)$ *are strictly increasing in* c*:*

$$\frac{d\underline{r}(c, \sigma)}{dc} = 1 > 0 \, , \; \frac{d\bar{r}(c, \sigma)}{dc} > 0.$$

[18]In the previous chapters, σ was set equal to one in order to simplify notation, and all boundaries were written as functions of c only. In the following, I will allow for variations in σ. With some abuse of notation, I redefine the boundaries η_1, η_2, λ, $\underline{r}$ and $\bar{r}$ to be the corresponding functions of c and σ.

For c approaching zero, the limits of both boundaries are given by

$$\lim_{c \to 0} \underline{r}(c,\sigma) = 0, \text{ and}$$
$$\lim_{c \to 0} \bar{r}(c,\sigma) = \frac{0.5\sigma}{\phi(0)}.$$

In equilibrium, party members are only willing to maintain their activity if each party's activity has a sufficiently large effect on expected policy, i.e., if the platform distance is large enough. As the cost of political activity becomes larger, party members demand increasing policy effects and platform distances. Thus, the minimal distance boundary increases. If, however, the membership cost approaches zero, the members will be willing to accept increasing convergence. In the limit, party membership is costless and is even consistent with full policy convergence.

With respect to the maximal distance boundary, increasing membership costs tighten the combined coordination and free-riding problem faced by potential activists. Whenever platform l is located to the left of the maximal effect position $l_\Gamma(r)$, all party members unanimously prefer to have a presidential candidate with bliss point $l_\Gamma(r)$ instead. As long as l does not exceed the extreme boundary, however, agents with this bliss point prefer to free-ride on the current party members, because the feasible policy gain is outweighed by the membership cost. With increasing c, an even larger policy gain is required to make political activity profitable. Thus, more extreme platforms can be supported in equilibrium and the maximal distance between both parties increases. When the membership cost converges to zero, on the other hand, this coordination problem vanishes and an agent with a desirable bliss point $w_i = l_\Gamma(r)$ will be willing to join party L whenever he is sure that he will become presidential candidate, i. e. whenever the initial platform is more extreme. With $c \to 0$, an independent agent with bliss point $l_\Gamma(r)$ benefits from entering the party whenever this has an effect on the party's platform. Thus, the party median members can always recruit their preferred candidates. Proposition 1.4 gives the mutually best platform choices, $l = -\frac{0.5\sigma}{\phi(0)}$ and $r = \frac{0.5\sigma}{\phi(0)}$, that would be chosen by party medians with extreme policy preferences, $m^L \to -\infty$ and $m^R \to \infty$.

Proposition 1.5. *The minimal distance boundary $\underline{r}(c,\sigma)$ is independent of the degree of electoral uncertainty while the maximal distance boundary $\bar{r}(c,\sigma)$ is strictly increasing in σ:*

$$\frac{d\underline{r}(c,\sigma)}{d\sigma} = 0\,, \quad \frac{d\bar{r}(c,\sigma)}{d\sigma} > 0.$$

In the case of full electoral certainty, both boundaries coincide:

$$\lim_{\sigma\to 0} \underline{r}(c,\sigma) = \lim_{\sigma\to 0} \bar{r}(c,\sigma) = c.$$

In figure 1.6, both the *minimal* and the *maximal distance equilibrium* involve symmetric platforms $l = -r < 0$, and equal winning probabilities for both parties (independently of σ). In the minimal distance case, the platform distance must be large enough so that party members do not benefit from leaving their party, and causing its inactivity. Thus, the membership cost c must not outweigh the policy effect $\Gamma(-r, r, \sigma) = [r - (-r)]\frac{1}{2} = r$, which is not affected by increasing uncertainty in this symmetric constellation. The same policy effect is even given for $\sigma = 0$, the case of a perfectly known population median.[19]

In contrast, the maximal distance boundary is derived by considering a shift from an extreme to a more centrist platform, i.e., a deviation from a symmetric to an asymmetric allocation. This platform shift is profitable to the party members and the potential entrant if and only if the winning probability increases sufficiently. Higher electoral risk however reduces the increase in winning probability and the incentive for independent agents to join a political party. Overall, increasing electoral risk diminishes the inherent centripetal forces of platform choice in endogenous parties, and more polarized platforms can be supported in equilibrium.

Corollary 1.2. *With electoral uncertainty, $\sigma = 0$, the platforms of both parties are given by $r = c$ and $l = -c$ in every two-party equilibrium.*

For the case of electoral certainty, all voters know the median voter position $m = 0$ ex ante. The uniqueness of party platforms for this case, $\sigma = 0$, is directly implied by the limits of both boundaries in Proposition 1.5. With σ approaching zero, the lower and upper boundaries $\underline{r}(c,\sigma)$, $\bar{r}(c,\sigma)$ converge and, in the limit, coincide. The economic intuition for this case is however simple, and can be provided directly.

If the bliss point of the population median is ex ante known, the political equilibrium can only involve two active parties if, first, those offer symmetric platforms $l = -r$, giving rise to identical winning probabilities. For any other constellation, one party would inevitably loose the general election and have no effect on the implemented policy. Thus, no agent would be willing to bear the cost of engaging in this party. Second, there cannot be an equilibrium with $r = -l < c$, as the distance between both parties and the implied policy effect would be too small for any agent to be willing to bear the cost of political activity.

[19]Note, however, that the moderate boundary $\eta_2(r, c, \sigma)$ changes for all values $r \neq \underline{r}$. Specifically, the moderate boundary function rotates clockwise with increasing σ.

Finally, platform polarization is limited by the possibility to recruit and nominate moderate independent citizens. Under electoral certainty, if any entrant with bliss point $w_i \in (l, 0)$ were to join party L and to be nominated as presidential candidate, he would certainly win the general election against platform $r = -l$. Since this would induce a shift of the expected policy $E(x) = 0$ to $w_i \in (l, 0)$, all members of party L would strictly prefer his nomination. Thus, an equilibrium with divergent platforms exists if and only if no independent agent can benefit from this deviation. For the potential entrant, entering party L improves the policy payoff by r. For $r > c$, joining party L would clearly be a profitable deviation. Thus, there is a unique political equilibrium with $r = c$ and $l = -c$.

Consequently, the effect of endogenous party formation is most obvious in the case of electoral certainty, which is the case on the basic citizen candidate model concentrates. The first two arguments also apply in the model by Besley and Coate (1997), implying that the platform distance must exceed a lower bound. Without party formation, however, there is no mechanism limiting policy polarization in equilibrium. Consequently, every symmetric allocation with platform distance beyond the lower bound represents a political equilibrium.[20] The resulting multiplicity of equilibria contrasts sharply with the unique determination of equilibrium platforms derived in Corollary 1.2.[21]

1.8 Conclusion

Building on the citizen candidate framework, this chapter has investigated political competition between endogenously formed parties. There seems to be little doubt that modeling political competition between parties instead of individual candidates brings theory closer to real-world politics. The model possesses a number of compelling properties. The analysis has focused on equilibria with two active parties, which are shown to exist for all levels of membership costs and electoral uncertainty. In contrast to the median voter model (Downs, 1957), there can never be full convergence of party platforms in equilibrium. Thus, the party

[20]In the model by Besley and Coate (1997), the lower bound on the platform distance depends on the cost of running in the general election, which has to be paid by a single candidate. Here, the lower bound instead depends on the cost of party membership. Intuitively, the latter cost should be considerably smaller.

[21]Note that, in the Osborne and Slivinski (1996) version of the citizen candidate model, the platform distance is not uniquely determined, but is nevertheless bound from below and from above. The upper bound follows from the assumption of sincere voting. Intuitively, extreme polarization is prevented by the assumption that voters are able to coordinate in Osborne and Slivinski (1996), while it is hindered (more completely) by the coordination of party members in my model.

formation model reproduces one of the main results of the basic citizen candidate model without parties (Besley and Coate, 1997). At the same time, allowing for party formation alleviates the major drawback of the citizen candidate model, the extreme multiplicity of equilibria. This becomes most obvious in the benchmark case of full electoral certainty, i.e., perfect information about the median voter's preferences. In this case, infinitely many equilibria with two running candidates exist in the basic citizen candidate model. In contrast, the party formation model possesses a unique equilibrium with two active parties.

This chapter has concentrated on a particularly simple framework to enhance the clarity of the arguments. A richer model could allow for, e.g., a larger number of potential parties, a multi-dimensional policy space, more general rules with respect to intra-party decision-making, more general policy preferences, or different modeling of electoral uncertainty. Further analyses show that the economic intuition and the main results are robust with respect to all these modifications.

Appendix 1.A Proofs

Proof of Lemma 1.1

Lemma 1.1 identifies the optimal choice of party platform l in the primary election of party L, conditional on the membership structure M^L and belief $\hat{r}$. It is proven through a series of lemmas.

Lemma 1.3. *Given any platform belief $\hat{r}$, the platform preferences of party members over the set of potential platforms fulfill the single crossing property.*

Proof. The single-crossing property implies that the preferences of agent i with respect to pairwise comparisons between two alternatives are monotonic in his bliss point ω_i. Consider the case $l_1 < l_2 < \hat{r}$. An agent with bliss point w_i prefers l_1 to be the platform of party L instead of l_2 if and only if the following condition holds:

$$\begin{aligned} F(l_1, l_2, \hat{r}, w_i) &= \tilde{v}_i(l_1, \hat{r}) - \tilde{v}_i(l_2, \hat{r}) \\ &= p(l_1, \hat{r})(|w_i - \hat{r}| - |w_i - l_1|) - p(l_2, \hat{r})(|w_i - \hat{r}| - |w_i - l_2|) \\ &> 0 \end{aligned}$$

First, note that $F(l_1, l_2, \hat{r}, w_i) \mid_{w_i<l_1} = p(l_1, \hat{r})(\hat{r} - l_1) - p(l_2, \hat{r})(\hat{r} - l_2) = \Gamma(l_1, \hat{r}) - \Gamma(l_2, \hat{r}) = -F(l_1, l_2, \hat{r}, w_i) \mid_{w_i>\hat{r}}$. Thus, agents with bliss points at both extremes of the policy space will always have conflicting preferences.

Second, the derivative of function F with repect to w_i is given as

$$\frac{dF(\,)}{dw_i} = \begin{cases} 0 & \text{for} \quad w_i \leq l_1 \\ -2p(l_1, \hat{r}) < 0 & \text{for} \quad w_i \in (l_1, l_2] \\ 2\left[p(l_2, \hat{r}) - p(l_1, \hat{r})\right] > 0 & \text{for} \quad w_i \in (l_2, \hat{r}] \\ 0 & \text{for} \quad w_i \geq \hat{r} \end{cases}$$

As long as platforms l_1 and l_2 provide different policy effects, there is exactly one cut-off value $\psi(l_1, l_2, \hat{r})$ such that $F(l_1, l_2, \hat{r}, \psi(\,)) = 0$ holds.

For $\Gamma(l_1, \hat{r}) > \Gamma(l_2, \hat{r})$, the cut-off is located in the interval (l_1, l_2). All agents with bliss points to the left of l_1 prefer l_1 and we get the following version of the single-crossing property:

$$\begin{aligned} F(l_1, l_2, \hat{r}, w_i) \leq 0 \Rightarrow & F(l_1, l_2, \hat{r}, w_j) < 0 \; \forall \; w_j > w_i, \text{ and} \\ F(l_1, l_2, \hat{r}, w_i) \geq 0 \Rightarrow & F(l_1, l_2, \hat{r}, w_k) > 0 \; \forall \; w_k < w_i \end{aligned}$$

For $\Gamma(l_1, \hat{r}) > \Gamma(l_2, \hat{r})$, the cut-off is located in the interval $(l_2, \hat{r})$. This time, all agents to the left of l_1 prefer platform l_2 and the preferences exhibit the following monotonicity:

$$F(l_1, l_2, \hat{r}, w_i) \leq 0 \Rightarrow F(l_1, l_2, \hat{r}, w_j) < 0 \; \forall \; w_j < w_i, \text{ and,}$$
$$F(l_1, l_2, \hat{r}, w_i) \geq 0 \Rightarrow F(l_1, l_2, \hat{r}, w_k) > 0 \; \forall \; w_k > w_i$$

For the special case of identical policy effects $\Gamma(l_1, \hat{r}) = \Gamma(l_2, \hat{r})$, all agents with bliss points to the left of l_1 as well as to the right of $\hat{r}$ are indifferent between both platforms while the moderate agents in the interval $(l_1, \hat{r})$ strictly prefer the moderate platform l_2. Trivially, the preferences satisfy the single-crossing property in the following sense:

$$F(l_1, l_2, \hat{r}, w_i) \geq 0 \Rightarrow F(l_1, l_2, \hat{r}, w_k) \geq 0 \; \forall \; w_k \in \mathbb{R}$$

Similar arguments apply for other constellations, e. g. $l_1 < \hat{r} < l_2$. □

Lemma 1.4. *For any member set M^L and platform belief $\hat{r}$, there is a Condorcet winner in the primary election of party L.*

Proof. Let the finite set of feasible platforms, i.e., the set of bliss points of party L's members, be given by A. Denote by l^* the platform in A that maximizes the utility of the median party member with platform $w_i = m^L$:

$$l^* = \arg\max_{l \in A} \tilde{v}_i(l, \hat{r}) = -p(l, \hat{r}) \left| \hat{r} - m^L \right| - [1 - p(l, \hat{r})] \left| l - m^L \right|$$

By the single-crossing property established in Lemma 1.3, platform l^* is preferred by a majority of party members (the median member plus either all members with $w_j \leq m^L$ or all members with $w_j \geq m^L$) to any other available platform $l' \in A$. Consequently, l^* wins any pairwise election and represents a Condorcet winner. □

Lemma 1.5. *On $(-\infty, r)$, the policy effect function $\Gamma(l, r) = p(l, r)(r - l)$ is strictly quasi-concave in l and has a unique maximizer $l_\Gamma(r)$.*

Proof. For $l < r$, the policy effect function and its first and second derivatives with respect to w_i are given as

$$\Gamma(l, r) = (r - l)\Phi\left(\frac{r + l}{2}\right),$$
$$\Gamma_1(l, r) = \frac{d\Gamma(l, r)}{dl} = -\Phi\left(\frac{r + l}{2}\right) + \frac{r - l}{2}\phi\left(\frac{r + l}{2}\right), \text{ and}$$
$$\Gamma_{11}(l, r) = \frac{d^2\Gamma(l, r)}{dl^2} = -\frac{1}{2}\phi\left(\frac{r + l}{2}\right) - \frac{1}{2}\phi\left(\frac{r + l}{2}\right) + \frac{r - l}{4}\phi'\left(\frac{r + l}{2}\right)$$

$$= -\left(1 + \frac{r^2 - l^2}{8}\right) \underbrace{\phi\left(\frac{r+l}{2}\right)}_{>0}.$$

As l approaches $-\infty$ and r, respectively, the first derivative goes to:

$$\begin{aligned}
\lim_{l \to -\infty} \Gamma'(l, r) &= \lim_{l \to -\infty} \frac{r-l}{2} \phi\left(\frac{r+l}{2}\right) = \lim_{l \to -\infty} \frac{(r-l)/2}{1/\phi\left(\frac{r+l}{2}\right)} \\
&= \lim_{l \to -\infty} \frac{\phi\left(\frac{r+l}{2}\right)^2}{\phi\left(\frac{r+l}{2}\right)} = \lim_{l \to -\infty} \frac{\phi\left(\frac{r+l}{2}\right)^2}{-\frac{r+l}{2}\phi\left(\frac{r+l}{2}\right)} \\
&= \lim_{l \to -\infty} \frac{\phi\left(\frac{r+l}{2}\right)}{-\frac{r+l}{2}} = 0, \text{ and} \\
\lim_{l \to r} \Gamma'(l, r) &= -\Phi\left(\frac{2r}{2}\right) < 0.
\end{aligned}$$

The second derivative $\Gamma''(l, r)$ is negative if and only if

$$\begin{aligned}
1 + \frac{r^2 - l^2}{8} > 0 \\
\Rightarrow l \in (-\sqrt{r^2 + 8}, +\sqrt{r^2 + 8})
\end{aligned}$$

For $l < -\sqrt{r^2 + 8}$, the policy effect function is thus strictly convex. Moreover, it is strictly increasing in this region, since $\lim_{l \to r} \Gamma_1(l, r) = 0$. In the interval $(-\sqrt{r^2 + 8}, r)$, the function is strictly concave. Combining these results we know that $\Gamma(l, r)$ is strictly quasi-concave on $(-\infty, r)$.

As $\Gamma_1(l, r)$ is positive for $l < -\sqrt{r^2 + 8}$ and negative for $l \to r$, the policy effect function must have a unique maximum on $l \in (-\infty, r)$. This maximum must be located in the interval $(-\sqrt{r^2 + 8}; r)$. □

Lemma 1.6. *For any membership set M^L and belief $\hat{r}$, the policy payoff of the median party member with bliss point $m^L < \hat{r}$ is maximized by platform $l(M^L, \hat{r}) = \max\left\{m^L, l_M(\hat{r}, M^L)\right\}$, where $l_M(\hat{r}, M^L) = \arg\max_{l \in A} \Gamma(l, r)$.*

Proof. The party median's policy payoff is given by $\tilde{v}_{m^L}(l, r) = p(l, r)\left\{\left|r - m^L\right| - \left|l - m^L\right|\right\} - \left|r - m^L\right|$. For $l \leq 2m^L - \hat{r}$ and $l \geq r$, the payoff is smaller than in the case of certain implementation of policy r, while it is strictly larger for any platform in the interval $(2m^L - \hat{r}, \hat{r})$. We can thus focus on this interval, where the derivative of $\tilde{v}_{m^L}(l, r)$ with respect to l is given by:

$$\frac{d\tilde{v}_{m^L}(l,r)}{dl} = \begin{cases} \frac{dp(l,r)}{dl}(r + l - 2m^L) + p(l,r) > 0 & \text{for} \quad l < m^L \\ \Gamma_1(l,r) & \text{for} \quad l \in \left(m^L, \hat{r}\right) \end{cases}$$

Consequently, the median member prefers its own bliss point m^L to any more leftist platform, independently of the implied policy effects. Furthermore, he will prefer a more moderate platform l' to his own bliss point if and only if l' provides a larger policy effect $\Gamma(l, r)$. Thus, $\tilde{v}_{m^L}$ will be maximized by the maximum of m^L and the maximum effect platform $l_M(\hat{r}, M^L)$. □

Proof of Lemma 1.2

Assume there is a two-party equilibrium with membership structures M_0^L, M_0^R, party medians m_0^L, m_0^R and platforms $l_0 = l(M_0^L, r_0)$, $r_0 = r(M_0^R, l_0)$ such that party L is not efficient, i.e., $\sum_{i \in N} \alpha_i^L \geq C + c$. Then, neither any member $j \in M_0^L$ nor any independent citizen with arbitrary bliss point w_k must have an incentive to deviate. In particular, this must be true for the most extreme member j with bliss point $w_j = \min\left\{w_i : i \in M^L\right\} \leq m^L$. Assume he deviates by reducing his contribution by c, which has two effects. First, the agent saves the membership cost of c. Second, if this deviation implies he leaves party L, the party median changes and becomes more moderate $m_1^L \geq m_0^L$. Since the party is not efficient by assumption, it will still be active in the general election. However, the adopted platform changes to $l_1 \geq l_0$. The deviation is non-profitable for agent j if and only if the induced policy loss would be larger than the membership cost. Otherwise he could profitably deviate by leaving party L, implying that the initial allocation cannot represent an equilibrium. Thus, the following condition must hold:

$$\tilde{v}_j(l_0, r_0) - \tilde{v}_j(l_1, r_0) = \Gamma(l_0, r_0) - \Gamma(l_1, r_0) > c \tag{1.7}$$

Next, consider the incentives for an independent with bliss point $w_k \in (l_1, r_0)$ to join party L. If he would enter party L, this deviation would have exactly the same effect on the party median as the exit of the extreme member j. Again, we have the new party median $m_1^L \geq m_0^L$. Furthermore, the new platform will either be given by the bliss point of the entrant w_k or by the platform adopted after j's exit, $l_1 \geq l_0$, once again.

Consider the latter is true, and l_1 is the newly adopted platform. The change in agent k's policy payoff is given by

$$\tilde{v}_k(l_1, r_0) - c - \tilde{v}_k(l_0, r_0) = p(l_1, r_0)\left[r_0 + l_1 - 2w_k\right] - p(l_0, r_0)\left[r_0 + l_0 - 2w_k\right] - c$$

$$= \Gamma(l_0, r_0) - \Gamma(l_1, r_0) + 2(r - w_k)[p(l_1, r_0) - p(l_0, r_0)] - c$$
$$> \Gamma(l_0, r_0) - \Gamma(l_1, r_0) - c.$$

The last expression is positive whenever condition (1.7) above holds. If the new platform is instead given by the bliss point of the entrant w_k, the induced increase of the policy payoff to k will be even larger.

Thus, if the extreme member j cannot deviate profitably by leaving party L, joining the party will be a profitable deviation for agent k. In other words, if party L is not efficient, there is always a profitable deviation for at least one of these two agents, which is a contradiction to the equilibrium assumption.

Proof of Proposition 1.1

Part (i) For the first part of Proposition 1.1, assume the policy effect associated with platform l does not exceed the membership cost: $\Gamma(l, r) < c$. By Lemma 1.2, parties are efficient in every political equilibrium. Whenever one member of L deviates by choosing $\alpha_i^L = 0$ and leaving party L, the sum of contributions to party L falls below the amount required cost of running C. Thus, there will no presidential candidate nominated by L in the general election, and the rightist candidate wins certainly. For the presidential candidate or any more extreme member with $\omega_i < l$, this deviation induces a utility change of

$$v_i(r) - (\tilde{v}_i(l, r) - \alpha_i^L) \geq v_i(r) - \tilde{v}_i(l, r) + c = -\Gamma(l, r) + c > 0$$

Thus, leaving party L would be profitable to agent i and, in equilibrium, platform l must fulfill the condition $\Gamma(l, r) \geq c$.

By Lemma 1.5 in the appendix, the policy effect function is strictly quasi-concave and approaches 0 for $l \to -\infty$ and $l \to r$. If $\Gamma(l_\Gamma(r), r) < c$, function A has no root. If $\Gamma(l_\Gamma(r), r) = c$, the maximum effect platform $l_\Gamma(r) = \eta_1(r, c), \eta_2(r, c)$ represents the unique root; the only equilibrium with platform r also involves $l_\Gamma(r)$. If finally $\Gamma(l_\Gamma(r), r) > c$, function $A(x, r, c)$ has two roots in x, denoted by $\eta_1(r, c) < l_\Gamma(r)$ and $\eta_2(r, c) > l_\Gamma(r)$. Then, $A(x, r, c) \geq 0$ holds if and only if $x \in [\eta_1(r, c), \eta_2(r, c)]$. Consequently, the condition stated in the first part of Lemma 1.3 is only fulfilled for platforms l in this interval.

Part (ii) For the second part of Proposition 1.1, consider an allocation with platforms r and $l < l_\Gamma(r)$. This position l can represent the outcome of a primary subgame equilibrium if and only if it provides higher policy effect than the bliss point of any other member in party L and if the party median is even more extreme:

$m^L(M^L) \leq l$. The allocation can only represent an equilibrium, if an independent agent with bliss point $w_i = l_\Gamma(r)$ cannot profitably deviate by joining party L. Given this deviation, agent i would certainly win the primary of party L and run as its presidential candidate in the general election since his bliss point is preferred to any other available platform by the party median. On the one hand, this change in L's platform increases the policy payoff to the entrant i. On the other hand, he has to pay the membership cost c. Overall, the induced change of utility for agent i is given by

$$B(l,r,c) = \tilde{v}_i\left(l_\Gamma(r), r\right) - c - \tilde{v}_i(l,r) = \Gamma\left(l_\Gamma(r), r\right) - p(l,r)\left[r + l - 2l_\Gamma(r)\right] - c$$

Agent i benefits from joining party L if and only if $B(l,r,c) > 0$ holds. The function has a unique root in l which is located in the interval $[2l_\Gamma(r) - r, l_\Gamma(r)]$. First, the deviation is profitable for any $l < 2l_\Gamma(r) - r$ since

$$\begin{aligned} B(l,r,c) =& \Gamma(l_\Gamma(r), r) - p(l,r)\underbrace{\left[r + l - 2l_\Gamma(r)\right]}_{<0} - c \\ >& \Gamma(l_\Gamma(r), r) - c > \Gamma(l,r) - c \geq 0 \end{aligned}$$

The second inequality holds by the definition of $l_\Gamma(r)$, the last one must hold in any potential equilibrium by part (i). Next, joining party L is obviously not profitable for i if l approaches $l_\Gamma(r)$: $\lim_{l \to l_\Gamma(r)} B(l,r,c) = -c < 0$. Finally, the incentive for political activity is strictly decreasing in l in this interval:

$$\frac{dB(l,r,c)}{dl} = -\frac{dp(l,r)}{dl}\underbrace{\left(r + l - 2l_\Gamma(r)\right)}_{>0} - p(l,r) < 0$$

Thus, function $B(l,r,c)$ has a unique root in l in $(-\infty, l_\Gamma(r))$, which we define as the extreme boundary function $\lambda(r,c)$. For all platforms $l < \lambda(r,c)$, joining party L and becoming its presidential candidate at the cost of c is a profitable deviation for agent i. Consequently, there cannot be an equilibrium with platforms r and $l < \lambda(r,c)$.

By the symmetry of both parties, the platform of party R must satisfy $r \leq -\lambda(-l,c)$ in every political equilibrium.

Proof of Proposition 1.2

Proposition 1.2 builds on the following lemma with respect to derivative of function $l_\Gamma(r)$.

Lemma 1.7. *The derivative of the maximum effect platform $l_\Gamma(r)$ with respect to platform r is given by $\frac{dl_\Gamma(r)}{dr} > -1$.*

Proof. The maximum policy effect platform $l_\Gamma(r)$ is implicitly (and uniquely) defined by the equation

$$\frac{d\Gamma(l_\Gamma(r), r)}{dl} = \frac{r - l_\Gamma(r)}{2}\phi\left(\frac{r + l_\Gamma(r)}{2}\right) - \Phi\left(\frac{r + l_\Gamma(r)}{2}\right) = 0$$

Substituting in $x = \frac{r + l_\Gamma(r)}{2}$, we can the following dependence:

$$\begin{aligned} l_\Gamma(r) = f(x) =& x - \frac{\Phi(x)}{\phi(x)}, \text{ and} \\ r = g(x) =& x + \frac{\Phi(x)}{\phi(x)}. \end{aligned}$$

Making use of these function, we can rewrite the function $l_\Gamma(r) = f(x(r) = f(g^{-1}(r))$.

For being able to use the inverse function g^{-1}, function g must be monotonic in x. For the standard normal distribution, $\frac{\phi(x)}{\phi(x)} = -x$ holds, so that the derivative of g equals $g'(x) = 2 + x\frac{\Phi(x)}{\phi(x)}$. We show that $x\frac{\Phi(x)}{\phi(x)} > -1$, which is a sufficient condition for $g'(x) > 0$ everywhere on $\mathbb{R}$. We do this by considering the auxiliary function $a(x) = x\Phi(x) + \phi(x)$, and proving that this function is positive for all $x \in \mathbb{R}$. We can derive the limit of $a(x)$ for $x \to -\infty$ by making use of l'Hopital's rule several times:

$$\begin{aligned} \lim_{x \to -\infty} a(x) =& \lim \left[x\Phi(x) + \phi(x)\right] = \lim \frac{x}{1/\Phi(x)} + 0 \\ =& \lim \frac{1}{-\phi(x)/\left[\Phi(x)\right]^2} = \lim -\frac{\Phi(x)^2}{\phi(x)} = \lim -\frac{2\Phi(x)\phi(x)}{\phi(x)} \\ =& \lim -\frac{2\Phi(x)\phi(x)}{-x\phi(x)} = \lim_{b \to -\infty} \frac{2\Phi(x)}{x} = 0 \end{aligned}$$

Moreover, $a(x)$ is strictly increasing in x, since $a'(x) = \Phi(x) + x\phi(x) + \phi(x) = \Phi(x) > 0$ for any $x \in (-\infty, +\infty)$. Thus, we have shown that $a(x) > 0$ for every $x \in (-\infty, +\infty)$, which is equivalent to

$$\begin{aligned} & x\Phi(x) > -\phi(x) \\ \Leftrightarrow\ & x\frac{\Phi(x)}{\phi(x)} > -1 \end{aligned}$$

Consequently, $g'(x) = 2 + x\frac{\Phi(x)}{\phi(x)} > 0$ holds for all x, and the inverse function $x(r) = g^{-1}(r)$ is well-defined. Its derivative is given by $x'(r) = [g'(g^{-1}(r))]^{-1} > 0$.

Second, consider the maximum effect platform function $l_\Gamma(r) = f(g^{-1}(r))$. Its derivative is given by

$$\frac{dl_\Gamma(r)}{dr} = f'(x)x'(r) = \left[-x\frac{\Phi(x)}{\phi(x)}\right] x'(r)$$
$$= -\frac{x\Phi(x)/\phi(x)}{2 + x\Phi(x)/\phi(x)}$$

As shown above, the denominator of this fraction is strictly positive. The sign of the numerator, and thus the complete expression, equals the sign of x. For all $x < 0$, it follows that $\frac{dl_\Gamma(r)}{dr} > 0$ (note that $x < 0$ results for $r < \frac{0.5}{\phi(0)}$). For $x > 0$, the numerator is positive, but strictly smaller than the denominator, implying that $\frac{dl_\Gamma(r)}{dr} \in (-1, 0)$ holds. Thus, the derivative of $l_\Gamma(r)$ is larger than -1 on its whole domain. □

Making use of Lemma 1.7, Proposition 1.2 can be proven in the following.

Part (i)**: Lower bound** By Lemma 1.3, the moderate platform boundary is defined as the larger root of $A(l, r, c) = \Gamma(l, r) - c$. Making use of this implicit definition, the derivative of $\eta_2(r, c)$ with respect to r is given by

$$\frac{d\eta_2(r,c)}{dr} = -\frac{\partial A/\partial r}{\partial A/\partial \eta_2} = -\frac{\Phi(\frac{\eta_2+r}{2}) + \frac{r-\eta_2}{2}\phi(\frac{\eta_2+r}{2})}{-\Phi(\frac{\eta_2+r}{2}) + \frac{r-\eta_2}{2}\phi(\frac{\eta_2+r}{2})}$$

The numerator of this expression is positive. As η_2 is the larger root of $A(l, r, c)$, the denominator is negative by the strict quasi-concavity of the policy effect function. Thus, $\frac{d\eta_2(r,c)}{dr} > 0$ holds for all $r \in \mathbb{R}$. Thus, the equation $r = -\eta_2(r, c)$ has at most one solution in r. Denote this solution by $\underline{r}(c)$. In this intersection, both platform are equal to their conditional minimal boundaries: $l = \eta_2(r, c)$ and $r = -\eta_2(-l, c)$. The intersection exists if and only if $\eta_2(r, c) < -r$ holds for some $r \in \mathbb{R}$.

On the other hand, the extreme boundary value $\eta_1(r, c)$ (also defined in Lemma 1.7) is strictly decreasing in r, and its derivative is given by $\frac{d\eta_1(r,c)}{dr} < -1$. Thus, the equation $r = -\eta_1(r, c)$ has at most one solution, too. However, this solution exists if and only if $\eta_1(r, c) > -r$ for some $r \in \mathbb{R}$ which is equivalent to $\eta_2(r, c) > -r$ for all $r \in \mathbb{R}$. Thus, the $45°$ line has either a unique intersection with the function $\eta_2(r, c)$ or a unique intersection with the function $\eta_1(r, c)$.

Finally, plugging in $r = \underline{r}$, $l = -\underline{r}$ gives

$$A(-\underline{r}, \underline{r}, c) = \Gamma(-\underline{r}, \underline{r}) - c = \underline{r} - c$$

Obviously, this function has value zero if and only if $\underline{r} = c$. Thus, the minimal distance boundary $\underline{r}$ exists. By the values of the derivatives, the constellation $l = -\underline{r}$, $r = \underline{r}$ in fact constitutes the equilibrium with smallest platform distance $r - l$.

Part (ii)**: Upper bound** The extreme boundary $\lambda(r, c)$ is defined as the unique root of the function $B(l, r, c) = \Gamma(l_\Gamma(r), r) - p(l, r)\left[r + l - 2l_\Gamma(r)\right] - c$ in l. The partial derivatives of $B(\lambda, r, c)$ with respect to λ and r are given by

$$\begin{aligned}
\frac{\partial B(\lambda, r, c)}{\partial \lambda} &= -\frac{1}{2}\phi\left(\frac{\lambda + r}{2}\right)(r + \lambda - 2l_\Gamma(r)) - \Phi\left(\frac{\lambda + r}{2}\right), \text{ and} \\
\frac{\partial B(\lambda, r, c)}{\partial r} &= 2\Phi\left(\frac{l_\Gamma(r) + r}{2}\right) + 2\Phi\left(\frac{\lambda + r}{2}\right)\frac{dl_\Gamma(r)}{dr} \\
&\quad - \frac{1}{2}\phi\left(\frac{\lambda + r}{2}\right)\underbrace{(r + \lambda - 2l_\Gamma(r))}_{>0} - \Phi\left(\frac{\lambda + r}{2}\right) < 0
\end{aligned}$$

Thus, the derivative of $\lambda(r, c)$ with respect to r follows as

$$\begin{aligned}
\frac{d\lambda}{dr} &= -\frac{\partial B(\lambda, r, c)/\partial r}{\partial B(\lambda, r, c)/\partial \lambda} \\
&= \frac{2\Phi\left(\frac{l_\Gamma(r)+r}{2}\right) + 2\Phi\left(\frac{\lambda+r}{2}\right)\frac{dl_\Gamma(r)}{dr} - \frac{1}{2}\phi\left(\frac{\lambda+r}{2}\right)\overbrace{(r + \lambda - 2l_\Gamma(r))}^{>0} - \Phi\left(\frac{\lambda+r}{2}\right)}{\frac{1}{2}\phi\left(\frac{\lambda+r}{2}\right)(r + \lambda - 2l_\Gamma(r)) + \Phi\left(\frac{\lambda+r}{2}\right)}
\end{aligned}$$

This derivative is larger than -1 if

$$\Phi\left(\frac{l_\Gamma(r) + r}{2}\right) + \Phi\left(\frac{\lambda + r}{2}\right)\frac{dl_\Gamma(r)}{dr} > 0.$$

Every term in this expression except $\frac{dl_\Gamma(r)}{dr}$ is strictly positive. For $r \leq \frac{0.5}{\phi(0)}$, the condition above holds because $\frac{dl_\Gamma(r)}{dr} \geq 0$, as shown in the proof of Lemma 1.7. For $r > \frac{0.5}{\phi(0)}$, we have $\frac{dl_\Gamma(r)}{dr} > -1$. Making use of the fact that $\lambda(r, c) < l_\Gamma(r)$, we then get

$$\Phi\left(\frac{l_\Gamma(r)+r}{2}\right)+\Phi\left(\frac{\lambda+r}{2}\right)\frac{dl_\Gamma(r)}{dr}>\Phi\left(\frac{\lambda+r}{2}\right)\left(1+\frac{dl_\Gamma(r)}{dr}\right)>0$$

Thus, $\frac{d\lambda(r,c)}{dr}>-1$ holds for all $r\in\mathbb{R}$.

Consequently, there can be at most one intersection between the boundary function $\lambda(r,c)$ and the 45° line ($l=-r$). Looking at the defining function of $\bar{r}(c)$, this statement is equivalent to the existence of a unique root in the function $\tilde{B}(r,c)=B(-r,r,c)$. This function can be expressed as

$$\begin{aligned}\tilde{B}(r,c)=&\Gamma(l_\Gamma(r),r)-p(-r,r)(r+(-r)-2l_\Gamma(r))-c\\=&\Gamma(l_\Gamma(r),r)+l_\Gamma(r)-c\end{aligned}$$

For $r=\frac{0.5}{\phi(0)}$, we have $l_\Gamma(r)=-\frac{0.5}{\phi(0)}$ and $\tilde{B}(r,c)=-c<0$. Moreover, $\tilde{B}(r,c)$ is strictly increasing in r:

$$\begin{aligned}\frac{d\tilde{B}(r,c)}{dr}=&\frac{d\Gamma(l_\Gamma(r),r)}{dl}\frac{dl_\Gamma(r)}{dr}+\frac{d\Gamma(l_\Gamma(r),r)}{r}+\frac{dl_\Gamma(r)}{dr}\\=&2\Phi\left(\frac{l_\Gamma(r)+r}{2}\right)+\frac{dl_\Gamma(r)}{dr}>0\end{aligned}$$

Once again, this statement holds for all $r\in\mathbb{R}$. Furthermore, it can be shown that there are $\delta>0$, $\varepsilon\in(0,1)$ such that $\frac{d\tilde{B}(r,c)}{dr}>\varepsilon>0$ holds for all $r>\frac{0.5}{\phi(0)}+\delta$. Thus, it is guaranteed that $\lim_{r\to\infty}\tilde{B}(r,c)>0$. We can conclude that there exists a unique threshold $\bar{r}(c)>\frac{0.5}{\phi(0)}$ such that $\tilde{B}(\bar{r},c)=0$ and $\tilde{B}(r,c)>0$ if and only if $r>\bar{r}$. In the maximum distance equilibrium, the party platforms are given by $(l,r)=(-\bar{r},\bar{r})$.

Finally, the threshold $\bar{r}$ strictly exceeds c, the rightist party's platform in the minimum distance equilibrium. At $r=c$, we have

$$\tilde{B}(c,c)=\Gamma(l_\Gamma(c),c)+l_\Gamma(c)-c=\left[p\left(l_\Gamma(c),c\right)-1\right]\left(c-l_\Gamma(c)\right).$$

For all $\sigma>0$, we have $p\left(l_\Gamma(c),c\right)<1$. Moreover, $l_\Gamma(c)<c$ holds generally. Thus, $\tilde{B}(c,c)<0$, which implies that $\bar{r}>c$ by the arguments above.

Proof of Proposition 1.3

We prove the existence of symmetric equilibria with $l=-d$ and $r=d$, where $d\in[c,c+\varepsilon]$ for some $\varepsilon>0$. Consider an allocation in which both parties are efficient, party medians are given by $m^L = -d$, $m^R = d$, and all party members are weakly more extreme, i.e., $\omega_i \le -d$ for all $i \in M^L$ and $\omega_j \ge d$ for all $j \in M^R$. Let all members contribute only the membership fee c to their party, so that each party consists of exactly $C/c{+}1 \ge 3$ members. Thus, the entry of a more moderate agent would not cause a shift in the party medians.

Recall that the platform bounds specified in Proposition 1.1 are only necessary, but not sufficient conditions. The allocation represents an equilibrium if and only if conditions (I), (II) and (III) are satisfied. Conditions (I) and (III) are satisfied for all $d \ge c$, because leaving his party would induce a policy loss of d to each party member, but only save the cost of activity c.

Condition (II) ensures that no independent agent can profitably join one of the parties. First, consider the case $c < \frac{0.5}{\phi(0)}$. For all $d \le \frac{0.5}{\phi(0)}$, we have $l_\Gamma(d) \le -d$. Thus, the party medians prefer there own bliss point to any more moderate platform. If $c < \frac{0.5}{\phi(0)}$, the allocation characterized above represents an equilibrium for all $d \in \left[c, \frac{0.5}{\phi(0)}\right]$, consequently. Now, consider case $c \ge \frac{0.5}{\phi(0)} > 0$, and let $d \ge c$. Then, some independent agents with $\omega_i \in (-d, d)$ could indeed enter party L and be nominated as presidential candidate. Taking into account the membership cost c, their utility would however change by

$$\begin{aligned}\Gamma(\omega_i, r) - \Gamma(-d, d) + 2p(-d, d)(\omega_i - l) - c &= \Gamma(\omega_i, r) + \omega_i - c \\ &= [p(\omega_i, d) - 1](d - \omega_i) + d - c < d - c.\end{aligned}$$

Because $p(\omega_i, d) < 1$, the first term in the last line is strictly negative. Thus, there is a $\varepsilon > 0$ such that the utility change of the joining agent is negative for all $d \in [c, c + \varepsilon]$. If the symmetric allocation satisfies this condition, it consequently represents an equilibrium.

Proof of Proposition 1.4

The minimal distance boundary is given by $\underline{r}(c) = c$, which directly gives the derivative provided in Proposition 1.3.

The maximal distance boundary $\bar{r}(c)$ is defined implicitly as the root of function $\tilde{B}(r, c) = \Gamma(l_\Gamma(r), r) + l_\Gamma(r) - c$ in r. Thus, its derivative is given as

$$\frac{d\bar{r}(c)}{dc} = -\frac{\partial \tilde{B}(\bar{r}, c)/\partial c}{\partial \tilde{B}(\bar{r}, c)/\partial r} = \frac{1}{2\Phi\left(\frac{l_\Gamma(r)+r}{2}\right) + \frac{dl_\Gamma(r)}{dr}} > 0$$

Note that the positive sign of the denominator has already been proven for Proposition 1.2.

For the limit, $\bar{r}(0)$ is given by the root of $\Gamma(l_\Gamma(r), r) + l_\Gamma(r)$. This equation is satisfied for $\frac{0.5}{\phi(0)}$. As there is at most one root as shown above, this must be the limit $\bar{r}(0)$.

Proof of Proposition 1.5

First, note that $\underline{r}(c, \sigma) = c$ for all $\sigma \geq 0$. Next, look at the derivative of $\bar{r}(c, \sigma)$ with respect to σ. The maximum distance boundary is defined by $\bar{r}(c, \sigma) = r \in \mathbb{R} : f(r, \sigma) = \Gamma(l_\Gamma(r, \sigma), r, \sigma) + l_\Gamma(r, \sigma) - c = 0$. By Proposition 1.2, function f has a unique root in r. The derivative follows as $\frac{d\bar{r}(c,\sigma)}{d\sigma} = -\frac{\partial f(\bar{r},\sigma)/\partial\sigma}{\partial f(\bar{r},\sigma)/\partial r}$.

The partial derivative of f in r is given by:

$$\begin{aligned}\frac{\partial f}{\partial r} =& \underbrace{\Gamma_1(l_\Gamma(r, \sigma), r, \sigma)}_{=0} \frac{dl_\Gamma(\bar{r}, \sigma)}{d\bar{r}} + \Gamma_2(l_\Gamma(r, \sigma), r, \sigma) + \frac{dl_\Gamma(r, \sigma)}{dr} \\ =& 2\,\Phi\left(\frac{r + l_\Gamma(r, \sigma)}{2\sigma}\right) + \frac{dl_\Gamma(r, \sigma)}{dr}\end{aligned}$$

As $\bar{r} + l_\Gamma(\bar{r}, \sigma) > 0$ holds in general, the induced winning probability of party L is strictly larger than one half. Thus, the partial derivative in $\bar{r}$ is strictly positive.

$$\frac{\partial f}{\partial \bar{r}} > 1 + \underbrace{\frac{dl_\Gamma(\bar{r}, \sigma)}{dr}}_{>-1} > 0$$

With respect to the partial derivative of f in σ, we get the following expression where I omit the arguments of $l_\Gamma(r, \sigma)$ in order to simplify notation.

$$\begin{aligned}\frac{\partial f}{\partial \sigma} =& \underbrace{\frac{\partial \Gamma(l_\Gamma, r, \sigma)}{\partial l_\Gamma}}_{=0} \frac{dl_\Gamma}{d\sigma} + \frac{\partial \Gamma(l_\Gamma, r, \sigma)}{\partial \sigma} + \frac{dl_\Gamma}{d\sigma} \\ =& -\frac{r^2 - l_\Gamma^2}{2\sigma^2}\phi\left(\frac{r + l_\Gamma}{2\sigma}\right) + \frac{1}{\sigma}\,\frac{8\sigma^2 l_\Gamma + (r - l_\Gamma)(r + l_\Gamma)^2}{8\sigma^2 + r^2 - l_\Gamma^2}\end{aligned}$$

It can be shown that this expression is negative if and only if the following condition holds:

$$\left[1 - \Phi\left(\frac{r + l_\Gamma}{2\sigma}\right)\right](r - l_\Gamma)(r + l_\Gamma)^2 < 8\sigma^2(r + l_\Gamma)\Phi\left(\frac{r + l_\Gamma}{2\sigma}\right) - 8\sigma^2 l_\Gamma$$

Making use of the fact that $p(l_\Gamma(r, c, \sigma), r, \sigma) > \frac{1}{2}$ once again, the following sufficient condition for $\frac{\partial f}{\partial \sigma} < 0$ can be derived:

$$\begin{aligned} \left[1-\Phi\left(\frac{r+l_\Gamma}{2\sigma}\right)\right](r-l_\Gamma)(r+l_\Gamma)^2 &< 4\sigma^2(r-l_\Gamma) \\ \Leftrightarrow \left[1-\Phi\left(\frac{r+l_\Gamma}{2\sigma}\right)\right]\left(\frac{r+l_\Gamma}{2\sigma}\right)^2 &< 1 \end{aligned}$$

Substituting $b = \frac{\bar{r}+l_\Gamma}{2\sigma} > 0$, this condition is given as

$$\hat{f}(b) = b^2\left[1-\Phi(b)\right] < 0$$

It can be shown that this condition holds on the relevant interval (for all $b > 0$). Thus, we have established $\frac{\partial f}{\partial \sigma} < 0$.

Consequently, the maximum distance boundary $\bar{r}(c,\sigma)$ is strictly increasing in σ:

$$\frac{d\bar{r}(c,\sigma)}{d\sigma} = -\frac{\overbrace{\partial f(\bar{r},\sigma)/\partial\sigma}^{<0}}{\underbrace{\partial f(\bar{r},\sigma)/\partial r}_{>0}} > 0$$

Finally, consider the limit of $\bar{r}(c,\sigma)$ for σ converging to zero. Look at platform $\tilde{l} = -r+\varepsilon$ with arbitrarily small $\varepsilon > 0$. The policy effect of this platform is given by

$$\Gamma(\tilde{l},r,\sigma) = (r-\tilde{l})\Phi\left(\frac{r+\tilde{l}}{2\sigma}\right) = (2r-\varepsilon)\Phi\left(\frac{\varepsilon}{2\sigma}\right)$$

For $\sigma \to 0$, this policy effect converges to $\lim_{\sigma\to 0}\Gamma(\tilde{l},r,\sigma) = 2r-\varepsilon$. Thus, $\tilde{l} = l_\Gamma(r,0)$ holds for $\varepsilon \to 0$. Then, looking at the defining function of $\bar{r}(c,\sigma)$ for general r, we get:

$$\lim_{\sigma\to 0} f(r,\sigma) = \Gamma(\tilde{l},r,\sigma) + \tilde{l} - c = 2r-\varepsilon-r+\varepsilon-c$$

Since $\bar{r}(c,\sigma)$ is defined to be the root of function $f(r,\sigma)$ in r, we obviously have the limiting result: $\lim\limits_{\sigma\to 0}\bar{r}(r,\sigma) = c$.

2

Political Selection and the Concentration of Political Power

2.1 Introduction

In representative democracies, political power is exercised by elected politicians. The role of institutions is to enforce the voters' interests within the political process. From the founding of modern democracies in the 18th century to recent constitutional drafts in Egypt and Lybia, political thinkers have been engaged in finding the best institutions. A central question in the debate has been whether political power should be concentrated on one group of political agents, typically the party winning the general election, or dispersed between different groups. Strikingly, there are pronounced cross-country differences along this dimension, with classical extreme cases being the United Kingdom (concentrated power), and Switzerland (dispersed power).[1]

The Federalist Papers highlight two channels through which constitutions affect social welfare: the selection of competent politicians into office and the disciplining of politicians in office.[2] The economic literature on the second chan-

[1]For a discussion of this crucial issue and its relation to various specific institutions, see Lijphart (2012), Lijphart (1999), and Tsebelis (2002).

[2]"The aim of every political Constitution is, or ought to be, first to obtain for rulers men who possess most wisdom to discern, and most virtue to pursue, the common good of society; and in the next place, to take the most effectual precautions for keeping them virtuous whilst they continue to hold their public trust" (Madison, 1788b).

nel consistently finds that power-dispersing institutions increase welfare as they help to discipline egoistic incumbents. In contrast, economists have little to say about the first channel, political selection (see Besley 2005). It cannot be taken for granted that voters are able to identify and empower the most competent politicians. Since voters base their ballot on their perceptions of candidates' competencies (Stokes, Campbell, and Miller, 1958; King, 2002; Pancer, Brown, and Barr, 1999), politicians exert considerable effort to appear competent and virtuous during electoral campaigns. This impedes the voters' capacity to empower able candidates. A comprehensive appraisal of political institutions thus has to account for whether or not institutions enforce the selection of competent candidates for office.

The aim of this chapter is to study the effects of power-concentrating institutions on the politicians' campaign behavior and on political selection. We consider a pre-election setup in which candidates are privately informed about their quality and partly motivated by office rents. Voters infer candidates' qualities from their policy proposals. We identify a trade-off that arises for changes in the level of power concentration. On the one hand, higher concentration of power implies a better allocation of power to competent candidates. We refer to this positive effect on welfare as the *empowerment effect* of power concentration. On the other hand, more concentration of power increases the desire of office-motivated candidates to win the election. Mimicking of competent candidates becomes more profitable, resulting in increasingly distorted policy choice. Thus, campaigns convey less information about the candidates' competence, and the voters' capacity to select high-ability politicians is reduced. We label this negative effect on welfare the *behavioral effect* of power concentration.

We formalize our argument by a simple model in which two candidates compete in a public election by making binding policy proposals. In particular, they can either commit to risky reforms or to the (riskless) status quo. Candidates differ in their abilities, which are unobservable to the electorate. Only highly able candidates can increase expected welfare by implementing a reform, while less able candidates should stick to the status quo. Voters observe policy proposals, draw inferences about the candidates' abilities, and vote accordingly. In equilibrium, a reform proposal is associated with high ability, and reforming candidates win the election more often than those proposing the status quo. Politicians do not only care about welfare, but are also office-motivated. This creates incentives for low-ability candidates to mimic the policy choice of their more able counterparts at the cost of adopting inefficient policies.

Variations in the level of power concentration induce the empowerment effect and the counteracting behavioral effect. The relative sizes of these effects depend on the importance of office motivation in politicians' preferences. The optimal institution balances both effects. We find that the optimal level of power dispersion is higher, the more politicians are driven by office rents. If and only if politicians care predominantly about implementing efficient policies, it is optimal to concentrate power completely in the hands of the election winner. Conversely, if office rents are a strong component of the candidates' motivation, some dispersion of political power enhances voter welfare.

The basic intuition behind this result is the following. Candidates' office motivation induces mimicking and distorts policy choices. Higher concentration of power strengthens the electoral incentives and, consequently, aggravates these distortions. For sufficiently high levels of office motivation, it is optimal to reduce the resulting inefficiencies by decreasing the concentration of power, even though this involves the delegation of some power to low-ability candidates. Hence, welfare can be enhanced through power-dispersing institutions.

To provide an economic intuition, we first analyze a simple model which abstracts from many important aspects of the political process. The qualitative results are however robust to a number of extensions and modifications. In section 2.8, we discuss two modifications of the theoretical model. First, we allow for heterogeneous policy preferences in the electorate. In this setting, we additionally show that power-dispersing institutions help to reduce inequality in the society, giving rise to a second rationale for power dispersion. In other words, if the social objective involves inequality aversion and a desire for balancing the interests of different groups in the society, less power should be concentrated in the hands of the election winner. Second, we relax the assumption of binding policy commitments. The qualitative results derived for the main model continue to hold if candidates are able to sometimes withdraw a proposed reform after the election, as long as policy proposals are not pure cheap talk.

Data from international surveys like the *International Social Survey Panel* indicate considerable differences across countries in how voters assess the motives of their politicians. Assuming that these differences mirror actual heterogeneity in politicians' motivation, our theoretical analysis gives rise to a testable hypothesis: Countries with strongly policy-motivated politicians benefit from power concentration. In contrast, countries with predominantly office-motivated politicians suffer from reduced welfare if they concentrate political power.

In a cross-country design, we investigate whether the welfare effect of power concentration indeed depends on politicians' motivation. For this purpose, we combine data on the perceived motivation of politicians with measures of political institutions.[3] As a measure for the performance of the political system, we

[3]We use Lijphart's index of the executive-parties dimension, which orders political systems ac-

use growth in real GDP per capita. Due to data availability restrictions, our analysis is restricted to eighteen established democracies, which rules out a rigorous statistical test. Nevertheless, the data for this limited set of countries are in line with our hypothesis.

The chapter proceeds as follows. The next section reviews the related literature. Section 2.3 presents the model. Section 2.4 delivers the benchmark of perfect information. Thereafter, we analyze the equilibrium behavior of privately informed politicians in Section 2.5. We proceed by examining the effects of institutions in Section 2.6. In Section 2.7, we present design and results of the empirical analysis. Section 2.8 provides two modifications of the theoretical model, and Section 2.9 concludes. All formal proofs are provided in the Appendix.

2.2 Related literature

In this chapter, we identify the economic effects of power-dispersing institutions, which limit the office-holders discretion. Many economists have addressed this question for a homogeneous set of politicians, thereby abstracting from political selection. With homogenous politicians, power-dispersing institutions increase voter welfare. For example, Lizzeri and Persico (2001) demonstrate in a pre-election setting that office-motivated politicians provide more of an efficient public good and less pork barrel under proportional representation than under plurality voting. In a post-election setting, Persson and Tabellini (2003) show that voters are more able to discipline an incumbent if power is separated between multiple political agents.

These papers abstract from any heterogeneity in candidate quality and thus from the role of political selection.[4] The importance of incorporating selection into the analysis of political institutions is demonstrated by Besley (2005). Selecting competent politicians has two aspects. First, to choose among competing candidates the one who holds most promise to design and implement efficient policies. This aspect is based on the candidates' campaign behavior. Second, to keep in office only politicians who perform adequately during the term. In other words, politicians are screened both before and after an election.

The role of institutions for political selection has so far only been studied in post-election settings. A first model addressing this question is Maskin and Ti-

cording to the implied dispersion or concentration of power, considering five categories of political institutions (Lijphart, 1999).

[4] The assumption that candidates differ in a quality dimension, sometimes referred to as "valence issue", is applied in a large number of papers, including Adams (1999), Ansolabehere and Snyder (2000), Aragones and Palfrey (2002), Messner and Polborn (2004), Sahuguet and Persico (2006), and Krasa and Polborn (2011).

role (2004). It investigates conditions under which the voter prefers political decisions to be taken by accountable "politicians" instead of non-accountable "judges". Maskin and Tirole (2004) argue that holding public officials accountable in reelections provides incentives to pander to public opinion and is thus not optimal for some kinds of political decisions. While they do not compare alternative democratic institutions, this approach is taken by Smart and Sturm (2013). They study variations in the level of accountability through the introduction of term limits. Depending on the share of public-spirited politicians, a limit of two terms as applied in many modern democracies is shown to be optimal.

Closest to this chapter is the analysis by Besley and Smart (2007), who study the effects of several fiscal restraints on political selection in a post-election setting. Similar to Maskin and Tirole (2004) and Smart and Sturm (2013), they identify a trade-off between disciplining incumbents and improving political selection. Whenever an institution allows the disciplining of bad incumbents, i.e., makes them adopt welfare-enhancing policies, this prevents effective political selection because voters are unable to distinguish a disciplined but bad politician from a good one. Our pre-election model produces a different trade-off. If voters have to infer the ability of candidates from their campaigns, dispersing power leads to both better policy choice and better selection, but comes at the cost of giving some political power to low-ability candidates. Besley and Smart (2007) consider four fiscal restraints that limit the office-holders' discretion, including limits on government size and transparency. Our focus, in contrast, is on power-dispersing institutions, such as proportional representation, federalism, or public referenda. Interestingly, Besley and Smart (2007) find that three of the four restraints only increase voter welfare if there are sufficiently many benevolent politicians. This contrasts our result according to which power dispersion is optimal if and only if the candidates are strongly driven by egoistic motives.

In this chapter as in the models discussed so far, the pool of political candidates is taken to be exogenous. In general, however, political institutions could also affect the quality of political selection by changing the set of agents that decide to enter politics. This aspect has been studied by another branch of the literature (Iaryczower and Mattozzi, 2008; Mattozzi and Merlo, 2008; Caselli and Morelli, 2004; Messner and Polborn, 2004). Closest to our analysis is the paper by Iaryczower and Mattozzi (2008) who investigate the effects with alternative voting systems on the quality of political candidates. They find that neither the majoritarian electoral system nor proportional representation clearly brings forth a higher-quality pool of candidates. Other papers study whether the quality of the candidate pool can be improved by financial incentives, i.e., changing the levels of wages paid to elected officials (Mattozzi and Merlo, 2008; Caselli and Morelli, 2004;

Messner and Polborn, 2004). In contrast, our model analyzes a broader range of political institutions and their effect on the competence in government, but takes the pool of candidates as given.

Finally, we relate to a growing empirical literature on democratic systems and their effects on fiscal policy. The analyses often focus on specific political institutions (see, e.g., Feld and Voigt, 2003; Persson and Tabellini, 2004; Enikolopov and Zhuravskaya, 2007; Blume et al., 2009; Voigt, 2011). In contrast, we apply a classification of political systems based on the implied dispersion of political power, thus encompassing a broad range of institutions. Using the same classification, Lijphart (1999; 2012) as well as Armingeon (2002) examine the influence of power dispersion on various political and economic outcomes. While Lijphart (1999) finds no effect of power dispersion on measures of economic performance, Armingeon (2002) finds a negative effect of power dispersion on unemployment and inflation. Complementing these findings, we show that the effect of power dispersion on growth in real GDP per capita positively depends on the strength of politicians' office motivation.

2.3 The model

Our model studies the effects of institutions on candidates' campaigns and political selection. Candidates differ in quality, more precisely in the ability to implement welfare-enhancing policies. They are privately informed about their abilities and commit to a policy prior to the election. The policy space is given by the unit interval, where the end points correspond to the status quo and a large-scale reform. Voters observe candidates' campaigns and vote based on the welfare they expect each candidate to provide. We depict political institutions in reduced form, by means of how much political power is concentrated in the political system. With higher concentration of power, the candidate receiving a majority of votes is more capable to enforce his agenda.

The game consists of three stages. At the first stage, nature independently draws both candidates' abilities a_1 and a_2, which are privately revealed to the candidates. At the second stage, both candidates simultaneously make binding policy proposals, x_1 and x_2. At the third stage, the voters observe the proposals, update their beliefs about the candidates' abilities and cast their votes. Based on the election result, the set of political institutions determines how political power is divided between both candidates. In general, the election winner as well as the loser win have some influence on policy choice.

While the basic model serves to clarify the main arguments, we allow for heterogeneity in the voters' policy preferences and for a form of limited commitment in Section 2.8. While providing some additional insights, these modifications do not alter the main results derived in the following.

2.3.1 Voters

There is a continuum of fully rational and risk-neutral voters of mass one who have preferences both over policy and candidates. In the considered policy field, an amount $x \in [0,1]$ of a risky reform can be implemented, where $x = 0$ and $x = 1$ represent the status quo and a complete reform, respectively. If a reform is successful, all voters receive a return of x, while a failed reform yields a return of zero. Whenever policy x is adopted, all voters bear a cost of cx. Maintaining the status quo is thus costless and yields a certain payoff of zero.

Voters are not exclusively interested in their expected payoff in the considered policy field. Instead, they also care about other policy fields and about the candidates' ideologies or personal characteristics other than ability. We account for these preferences by assuming that voters have heterogeneous candidate preferences, following the probabilistic voting approach (see Lindbeck and Weibull 1987). If policy is set by candidate 1, voter k receives an additional utility of μ_k, while we normalize the additional utility if candidate 2 determines policy to zero. Let μ_k be distributed according to some continuous pdf that is symmetric around zero and has full support on the interval $[-1,1]$. This guarantees heterogeneity in the resulting voting preferences.[5]

If candidate i implements policy x_i, voter k receives a utility of

$$V_k(x_i, i) = \begin{cases} 1_{i=1}\,\mu_k + x_i\,(1-c) & & \text{reform succeeds} \\ 1_{i=1}\,\mu_k - c & \text{if} & \text{reform fails} \\ 1_{i=1}\,\mu_k & & \text{status quo is maintained,} \end{cases} \tag{2.1}$$

where $1_{i=1}$ denotes the indicator function which is one if $i = 1$ and zero otherwise. Voter k prefers candidate 1 if and only if he expects $V_k(x_1, 1)$ to be larger than $V_k(x_2, 2)$. We assume sincere voting, i.e., each voter casts his vote for his preferred candidate. Hence, candidate i's vote share depends positively on the voters' belief about the payoff he provides, and negatively on the belief about the payoff provided by his opponent.

[5]Note that our results are independent of whether these candidate preferences are subject to an additional aggregate shock as in Lindbeck and Weibull (1987).

2.3.2 Candidates

Two candidates run for office. Each candidate i can commit to a policy $x_i \in [0, 1]$. More able candidates design better reforms, i.e., reforms that are more likely to succeed. We measure candidate i's ability by the implied probability of a successful reform, $a_i \in [0, 1]$. Hence, policy x_i set by candidate i provides an expected payoff of $x_i(a_i - c)$ to each voter. Since candidate preferences are symmetrically distributed around zero, this is candidate i's expected welfare contribution.

Both agents' abilities are realizations of two independent random variables with identical cumulative distribution Φ. Let the corresponding density function ϕ have full support on $[0, 1]$ and be continuously differentiable. To make the problem interesting, we assume that the expected ability $\int_0^1 a\phi(a)da$ is below c. Thus, the voters benefit from a reform in expectation if and only if it is designed by a high-ability candidate. After observing his ability, each candidate i commits to policy $x_i \in [0, 1]$. Thus, the strategy X_i of politician i is a mapping from abilities to policy proposals.

Each candidate cares about his amount of political power (office motivation) as well as about his expected welfare contribution (policy motivation). The utility function of politician i is given by

$$U_i(a_i, x_i) = f(v_i, \rho)\left[\theta + x_i(a_i - c)\right], \tag{2.2}$$

where $\theta > 0$ denotes the relative weight of office motivation. Put differently, θ describes the increase in the candidates' office utility due to a marginal increase in his amount of power.[6] The term $f(v_i, \rho) \in (0, 1)$ captures candidate i's political power, which we define as his influence on policy implementation after the election. As explained in more detail below, the implemented policy is in general a compromise between both candidates' proposals. More precisely, we assume that the implemented policy is a linear combination of the proposals with weights being equal to the power of the candidates. Each candidate's power depends on his vote share v_i and the parameter of power concentration ρ, representing the set of political institutions. To simplify notation, this utility function is formulated at an ex interim stage, i.e., taking the expected payoff after the election but before the reform outcome has been realized. For readability, we have also omitted the dependence of v_i on both candidates' strategies and actions.

Note that candidate i only cares about how expected welfare is affected by his policy choice, not about voter welfare in general. This way to formulate policy

[6]We assume that θ mirrors the candidates' intrinsic utility of power exertion and can thus not be manipulated by institutions. If the preference parameter θ could instead be decreased by changing political institutions or the politicians' wages, this would obviously be welfare-enhancing.

preferences of politicians has been introduced by Maskin and Tirole (2004), who label it *legacy motive*. It captures the politicians' desire to leave a positive legacy to the public.[7]

2.3.3 Political institutions

We model political institutions by a power allocation function f that translates election results into an allocation of political power, i.e., each politician's probability to implement his policy proposal. Formally, candidate i's power $f(v_i, \rho)$ depends on his vote share v_i and on the level of power concentration ρ implied by the set of political institutions.

Definition 2.1. The continuously differentiable function $f : [0, 1] \times \mathbb{R}_+ \to [0, 1]$ is a power allocation function if it satisfies

(i) symmetry in v_i: $f(v_i, \rho) = 1 - f(1 - v_i, \rho)$,

(ii) monotonicity in v_i: $\frac{\partial f(v_i,\rho)}{\partial v_i} \geq 0 \; \forall \; \rho$,

(iii) piece-wise monotonicity in ρ: $\frac{\partial f_i(v_i,\rho)}{\partial \rho} > 0 \; \forall \; v_i \in (1/2, 1)$, and

(iv) $\lim_{\rho\to\infty} f(v_i, \rho)$ is a step function with a discontinuity at $v_i = \frac{1}{2}$.

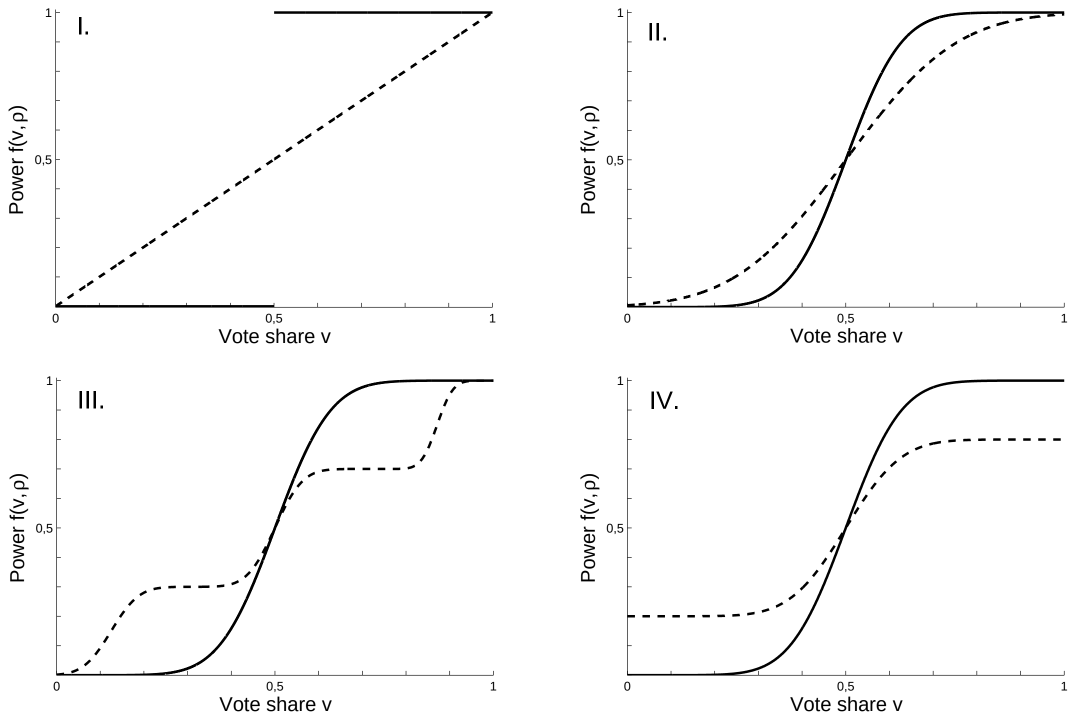
Property (i) establishes anonymity, i.e., the constitution does not treat candidates differently. Property (ii) rules out that candidates receive a larger amount of political power if they gain less votes in the election. By property (iii), the parameter ρ can be interpreted as a measure of power concentration. The higher ρ, the larger is the amount of power assigned to the election winner, i.e., the candidate who gains more than half of the votes. Finally, (iv) implies that in the case of full power concentration, the electoral margin does not matter. In other words, the election winner receives all power that is allocated through the corresponding election.

This modeling approach allows to study a large variety of institutional differences. Figure 2.3.3 illustrates how political institutions can be represented by power allocation functions. Each panel depicts two examples. Throughout, the solid line represents institutions that concentrate power more strongly than those corresponding to the dashed line.

Panel I depicts two stylized allocation rules frequently used to compare electoral systems in the theoretical literature (see, e.g., Lizzeri and Persico 2005). The

[7] Alternatively, we could assume that candidates directly care for welfare. While complicating the analysis, this assumption would not change the qualitative results.

Figure 2.1: Political institutions and corresponding power allocation functions.



solid line represents institutions that fully concentrate power in the hands of the election winner. In political economy, this step function is the standard way to model plurality voting. The dashed line represents proportional representation, which implies a lower concentration of political power and is often modeled by the identity function $f(v_i, \rho) = v_i$.

A less simplistic representation of these two systems is shown in Panel II. Here, the winner's amount of power depends on his margin of victory, e.g., because delegates might vote against the party lines. Plurality voting tends to generate clear-cut majorities, as the winning party typically receives a share of parliamentary seats beyond its vote share. In contrast, the allocation of seats corresponds closely to vote shares under proportional representation. Thus, the dashed curve for the proportional system is flatter than the one for plurality voting.

In Panel III, the dashed line represents a political system with a supermajority requirement for certain policy decisions (as employed in Germany and the US). This requirement generates additional steps in the power allocation function, since some policies can only be enforced after a landslide victory. In contrast, the solid line corresponds to a system as applied in the UK, where any decision can be taken by a simple majority.

Finally, the dashed line in Panel IV depicts the use of direct democratic institutions as employed for example in Switzerland. Even after a landslide victory in the election, the winning party cannot always implement its agenda. The opposition party can block policies via a referendum or even enforce its own proposals. Thus, only a limited part of political power is at stake in the parliamentary election (similar arguments can be made with respect to federalism, bicameralism or a constitutional court).

2.3.4 Equilibrium concept and normative criterion

To solve the game, we apply the notion of Perfect Bayesian equilibria and the D1 refinement proposed by Cho and Kreps (1987). A Perfect Bayesian equilibrium of the game defined above consists of a strategy profile (X_1^*, X_2^*) and a belief system σ^* such that (1) both candidates play mutually best responses when announcing their policy proposals, anticipating the winning probabilities for each vector (x_1, x_2) that are implied by the voters' beliefs σ^*, and (2) the voters' belief system σ^* is derived from the candidates' strategies X_1^*, X_2^* according to Bayes' rule everywhere on the equilibrium path. A D1 equilibrium is a Perfect Bayesian equilibrium that is robust to the D1 criterion, which restricts the set of beliefs off the equilibrium path. Intuitively, the D1 criterion rules out "unreasonable" belief systems by requiring that each deviation from equilibrium actions must be attributed to types that profit from it under the most general conditions.[8]

We investigate the effects of changes in power concentration, i.e., in the parameter ρ. As normative criterion, we use a utilitarian welfare function in ex ante perspective, i.e., expected welfare before candidates' abilities are drawn:

$$W(\rho, \theta) = \int_0^1 \int_0^1 \phi(a_1)\phi(a_2) \sum_{i=1}^{2} f(v_i, \rho) X_i(a_i)(a_i - c)\, da_2 da_1. \tag{2.3}$$

Welfare is hence given by the weighted sum of the politicians' welfare contributions, integrated over all possible combinations of the candidates' ability. The weights correspond to the candidates' power, $f(v_i, \rho)$. Note that welfare is calculated using equilibrium strategies, which are functions of the parameters ρ (power concentration) and θ (candidates' office motivation).

[8] More precisely, D1 specifies the beliefs associated to each off-equilibrium action x as follows: First, identify for each type a the set of beliefs $\Sigma(a, x)$ for which the action would be profitable. Second, type a belongs to set $A_d(x)$ if there is some other type a' such that $\Sigma(a, x)$ is a strict subset of $\Sigma(a', x)$. D1 requires that action x must not be associated to any type in the dominated set A_d.

2.4 Benchmark case: perfect information

If individual abilities are observable to the electorate, voters condition their ballot on candidates' abilities and reform proposals. In particular, the fraction of citizens voting for a candidate is increasing in his welfare contribution.

With perfect information, the agents' office and policy motives are fully aligned: Each candidate maximizes his power by proposing the policy with the highest welfare contribution. Hence, a complete reform ($x_i = 1$) is proposed by high-ability candidates with $a_i \geq c$. In contrast, a candidate with ability $a_i < c$ gains more power by proposing the status quo instead of a reform with a negative welfare contribution. Thus, equilibrium policy choices are undistorted: A politician proposes to implement a complete reform if and only if the reform enhances welfare. As a consequence, candidates with higher ability, i.e., those who propose to reform, receive higher vote shares in the election.

This result has a direct welfare implication. While variations in power concentration ρ do not distort candidates' behavior, a higher concentration of power allocates more power to candidates with higher welfare contribution. Hence, welfare strictly increases with the level of power concentration. The following Proposition summarizes these results.

Proposition 2.1. *Under perfect information, candidates propose a complete reform if and only if $a_i \geq c$. Welfare is maximized if political power is completely concentrated.*

2.5 Imperfect information

For the remainder of this chapter, we assume that both candidates are privately informed about their abilities. Voters observe the policy proposals x_1, x_2 and form beliefs about the candidates' abilities a_1, a_2, on which they base their voting decisions. As a consequence, the vote shares of both candidates depend on the belief system σ. More concretely, the belief system determines the expected welfare contributions $\hat{\pi}_i(x_i) = x_i\,[\hat{a}_i(x_i) - c]$ for $i \in \{1, 2\}$, where $\hat{a}_i(x_i)$ denotes the expected ability of candidate i proposing policy x_i implied by belief σ. In a Perfect Bayesian equilibrium, these beliefs are consistent with the candidates' strategies everywhere on the equilibrium path.

Proposition 2.2. *In every Perfect Bayesian equilibrium, the strategy of candidate $i \in \{1, 2\}$ can be characterized by a cutoff $\alpha_i \in (0, c)$ and a reform magnitude*

$b_i \in [0,1]$ *such that*

$$X_i^*(a_i) = \begin{cases} 0 & \textit{for } a_i < \alpha_i, \\ b_i & \textit{for } a_i \geq \alpha_i. \end{cases} \tag{2.4}$$

By Proposition 2.2, both agents play simple cut-off strategies that involve at most two actions. In the appendix, we first show that the policy preferences of candidate i satisfy a single-crossing property, given any belief system σ and strategy X_{-i} played by the opponent. Second, we find that candidate i will propose the same policy $X_i^*(a_i) = x$ whenever his ability a_i is above the reform cost c. Consequently, any other equilibrium proposal x' can only be proposed for some ability below c. As the voters anticipate this, any proposal $x' \neq x$ will be associated with an expected ability $\hat{a}_i(x')$ below c, and a negative welfare contribution $\hat{\pi}_i(x')$. But this implies that, for all agents with ability below c, proposing any policy $x' > 0$ gives a lower welfare contribution as well as a lower vote share than the status quo proposal $x_i = 0$. Thus, candidate i will never play any other action than the status quo and a unique reform announcement $b_i > 0$ in equilibrium.

The concept of Perfect Bayesian equilibrium does not pin down the exact level of the reform magnitude b_i, however. For all parameter values, there exists in contrast a large set of equilibria with differing levels of b_1 and b_2. For some combinations of the parameters ρ and θ, there even exist equilibria with $b_i = 0$, in which candidate i plays a pooling strategy, proposing the status quo independently of his ability level. These equilibria are however supported by very pessimistic out-of-equilibrium beliefs, associating a very negative ability to candidate i whenever he proposes some other policy.

Because the incentive to deviate by proposing a reform is strictly increasing in the ability of candidate i, such pessimistic beliefs seem "unreasonable" though. Indeed, the D1 criterion proposed by Cho and Kreps (1987) rules out these beliefs and consequently eliminates all equilibria with pooling strategies. More generally, we find that only a small subset of the Pareto efficient equilibria are robust to the D1 criterion.

Proposition 2.3. *For all combinations of θ and ρ, the set of D1 equilibria is non-empty. In every political D1 equilibrium, both candidates play identical strategies*

$$X_i^*(a_i) = \begin{cases} 0 & \textit{for } a_i < \alpha, \\ 1 & \textit{for } a_i \geq \alpha, \end{cases} \tag{2.5}$$

with a symmetric cutoff $\alpha \in (\underline{a}, c)$, where $\underline{a} > 0$ is implicitly defined by $\int_{\underline{a}}^1 a d\Phi(a) = 0$.

The D1 criterion restricts the beliefs for off-equilibrium actions. Intuitively, it requires that the voters must associate the deviation to any off-equilibrium action x' to the types that benefit from this deviation under the largest set of beliefs (Cho and Kreps, 1987). In this model, this equilibrium refinement eliminates all equilibria with $b_i \neq 1$ for one of the candidates. If b_i would in contrast differ from unity, the D1 criterion would require that the voters must associate an expectation of $\hat{a}_i(1) = 1$ to the off-equilibrium action $x_i = 1$. Given this belief, the deviation would however be profitable for high-ability agents. Thus, only Perfect Bayesian equilibria with full reforms by both candidates are robust to the D1 criterion.[9]

By Proposition 2.3, the equilibrium behavior of both candidates is completely characterized by the symmetric cutoff $\alpha \in (\underline{a}, c)$. If candidate i has an ability above the cutoff α, he can achieve a higher payoff by proposing a full reform $x_i = 1$ than by proposing the status quo. Formally, the utility difference

$$E_{a_{-i}}[U_i(a_i, 1)|X^*_{-i}(a_{-i}), \sigma] - E_{a_{-i}}[U_i(a_i, 0)|X^*_{-i}(a_{-i}), \sigma] \tag{2.6}$$

is positive if $a_i > \alpha$, and negative if $a_i < \alpha$. At the cutoff level $a_i = \alpha$, candidate i is indifferent between both actions. For any level of rho, α is implicitly defined by the equation

$$\begin{aligned} R(\alpha, \rho) =& E_{a_{-i}}[U_i(\alpha, 1)|X^*_{-i}(a_{-i}), \sigma] - E_{a_{-i}}[U_i(\alpha, 0)|X^*_{-i}(a_{-i}), \sigma] \\ =& \underbrace{\theta\left(f(v^r(\alpha), \rho) - \frac{1}{2}\right)}_{\text{Change in office utility}} + \underbrace{\left[\frac{1}{2} + \Phi(\alpha)\left(f(v^r(\alpha), \rho) - \frac{1}{2}\right)\right](\alpha - c)}_{\text{Change in welfare contribution}} \\ =& 0 \end{aligned} \quad . \tag{2.7}$$

Here, $v^r(\alpha)$ denotes the vote share of a reforming candidate when facing an opponent who proposes the status quo, given that the voters' beliefs are consistent with the equilibrium cutoff α. In the following, we refer to function $R(\alpha, \rho)$ as the reform incentive function.

Equation (2.7) distinguishes between both aspects of the politicians' preferences. By proposing a reform instead of the status quo, a candidate can gain more political power, but will also provide a different welfare to the voters. For the cutoff type, both effects exactly outweigh each other. As a consequence, the cutoff α satisfies two conditions in every equilibrium. First, the reform proposal must be associated with a positive welfare contribution, i.e., $\hat{a}_i(1) = \int_\alpha^1 a d\Phi(a) > c$. This

[9] In contrast, equilibria with $b_1 = b_2 = 1$ are robust to the D1 criterion. The formal arguments are provided in more detail in the appendix.

requires α to exceed the lower bound $\underline{a} > 0$, as the average ability is assumed to be below c. In equilibrium, policy proposals thus provide at least some information to the voters. Assume in contrast a negative welfare contribution, $\hat{a}_i(1) < c$. Then, reforming candidates would achieve a lower vote share than those proposing the status quo. Thus, no type with ability below c would prefer proposing a reform, which is inconsistent with the assumption $\hat{a}_i(1) < c$.

Thus, a reform proposal is always associated with a positive contribution, and leads to a higher vote share than the status quo. It follows that candidates with ability above c always choose to reform. They gain not only from their positive welfare contribution, but also from an increase in expected office rewards. Thus, the equilibrium cutoffs must be strictly below c. For all abilities below c, in contrast, a reform proposals leads to two counteracting effects in office utility and his welfare contribution. In particular, each candidate is willing to commit to an inefficient reform for all abilities in the interval $[\alpha, c)$ in order to achieve a higher vote share.

In general, the number of D1 equilibria depends on the properties of the ability distribution Φ. The following regularity condition ensures equilibrium uniqueness.

Assumption 2.1. *The pdf of the ability distribution $\phi(a)$ is bounded from above with $\phi(a) < \frac{1+\Phi(a)}{c-a}$ for all $a < c$.*

Proposition 2.4. *Under Assumption 2.1, there is a unique D1 equilibrium.*

As argued above, any root of the reform incentive function $R(\alpha, \rho)$ in its first argument represents a D1 equilibrium. Assumption 2.1 ensures that R is monotonically increasing in its first argument for all levels of ρ and θ. This directly rules out the possibility of multiple equilibria. Assumption 2.1 is fulfilled, e.g., for the uniform distribution. For the remainder of the chapter, we take it as given and derive the effects of variations in the institutional setting in the unique D1 equilibrium.[10]

2.6 Effects of power-concentrating institutions

Empirically, democratic countries differ strongly in their political institutions and the implied power concentration. As we have argued in Subsection 2.3.3, our

[10]If Assumption 2.1 is not satisfied, multiple equilibria may arise. The following analysis of the effects of power concentration is still valid if we restrict our attention to the equilibrium with the highest cutoff level, which also involves the highest level of welfare.

framework allows to represent these differences by an appropriate power allocation function $f(v_i, \rho)$. In this section, we study the effects of variations in power concentration ρ.

2.6.1 Effects on candidates' behavior

The power allocation function f determines the electoral incentives of political candidates. Under perfect information, variations in power concentration leave the behavior of candidates unaffected and policy choice is efficient (see Proposition 2.1).

With asymmetric information and office-motivated candidates, in contrast, policy choice is distorted as some low-ability candidates propose welfare-reducing reforms. Political institutions affect the magnitude of these policy distortions.

Proposition 2.5. *Increasing power concentration ρ leads to the proposal of strictly more inefficient reforms:* $\frac{d\alpha}{d\rho} < 0$.

Consider some level of power concentration ρ_0. The cutoff type with ability $a_i = \alpha_0 < c$ is defined by the condition (2.7), i.e., is indifferent between a reform proposal and the status quo. We find that after an increase in power concentration, the cutoff type strictly prefers to propose a reform. In particular, his utility of proposing the status quo decreases while his utility of proposing a reform increases.

If the cutoff type proposes the status quo, his welfare contribution is equal to zero. Thus, he only draws utility from office rents, which are in expectation positive in any equilibrium. With increasing power concentration, these office rents are reduced because he receives less power when running against a reforming opponent.

If the cutoff type proposes a reform, he again receives office rents, but also incurs a utility loss due to his negative welfare contribution. His overall utility is given by the sum of these two components, which is positive because it must be equal to the utility from a status quo proposal. With increasing power concentration, both the office rents and the negative welfare contribution increase by the same factor. Hence, his utility from a reform proposal also increases by this factor.

Consequently, with higher levels of power concentration, status quo proposals yield lower utility while reform proposals become more attractive. The equilibrium cutoff thus decreases with the level of power concentration.

2.6.2 Effects on welfare

In the following, we study the effects of power-concentrating institutions on ex ante welfare. With privately informed candidates, the relation between power concentration and welfare is not as clear-cut as under perfect information.

On the one hand, there is still a positive *empowerment effect* of power concentration. Whenever both policies are proposed, the majority of votes goes to the reforming candidate, who provides higher expected welfare than the candidate proposing the status quo (see Section 2.5). Consequently, any increase in power concentration ρ assigns more power to the appropriate candidate.

On the other hand, the previous section demonstrated a negative *behavioral effect* of power concentration. By reinforcing the electoral stakes, stronger concentration of power induces the proposal of more inefficient reforms. This reduces the information revealed during the campaigns and limits the voters' capacity to allocate power to high-ability candidates.

We impose the following regularity condition on the ability distribution.

Assumption 2.2. *The ability distribution is log-concave, i.e.,* $\Phi(a)/\phi(a)$ *is nondecreasing in* a*.*

Assumption 2.2 is satisfied for many common distributions, including the normal, the uniform and the exponential distributions.

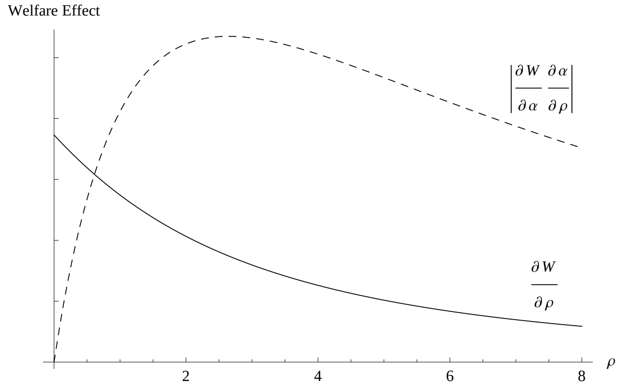
Lemma 2.1. *The welfare function* W *is strictly quasi-concave in* ρ*.*

Lemma 2.1 implies that the welfare function has a unique maximum in ρ. Its proof involves analyzing how power concentration influences the empowerment effect and the behavioral effect.

First, consider the positive empowerment effect. With increasing ρ, a reforming candidate receives more power if he runs against an opponent proposing the status quo. Welfare is increased by this reallocation of power because reforms are associated with a positive expected welfare contribution as argued above. The size of the empowerment effect is determined by the average reform payoff. At higher levels of ρ, more inefficient reforms are proposed, so that the average reform payoff is diminished. Consequently, the empowerment effect is strictly decreasing in ρ, as illustrated by the solid line in Figure 2.6.2.

Second, consider the negative behavioral effect. It results because increasing power concentration leads to a reduction in the cutoff α. The size of this effect depends on, first, the marginal welfare loss from a decline in α, and second, the sensitivity of α with respect to changes in power concentration. Both factors are affected differently by increasing power allocation.

Figure 2.2: Empowerment effect and behavioral effect



The welfare effects of a change in ρ for a logistic power allocation function with mean $\mu = 0.5$ and scale parameter $\beta = 1/\rho$, μ_k distributed according to $\mathcal{N}(0, 0.5)$, a uniform ability distribution, $\theta = 1$, and $c = 0.6$. The solid line represents the (positive) empowerment effect, the dashed line represents the (negative) behavioral effect. The optimal level of ρ is attained at the intersection of both lines.

Regarding the first factor, higher power concentration induces the cutoff to depart further from its efficient level c. As increasingly inefficient reforms are proposed, the marginal welfare loss from reductions in α increases with ρ. Regarding the second factor, the sensitivity of α depends on the additional vote share a candidate gains by proposing a reform, which is directly related to the average reform payoff. For higher levels of power concentration, the average reform payoff becomes smaller and so does the additional vote share. Thus, higher levels of ρ come along with a reduced sensitivity of α, which attenuates the behavioral effect. As a consequence, the behavioral effect is non-monotonic in ρ (see the dashed line in Figure 2.6.2).

The sign of the overall effect of power concentration on welfare depends on the relative sizes of both effects. With a log-concave ability distribution, the ratio of empowerment effect and behavioral effect is strictly decreasing in ρ at every local extremum, as we show in Appendix A. Thus, the welfare function cannot have an interior minimum and at most one interior maximum in ρ, which corresponds to the definition of quasi-concavity.

Proposition 2.6. *If and only if office motivation is below some threshold level $\bar{\theta}$, welfare is maximized by full concentration of power. If instead $\theta > \bar{\theta}$, it is optimal to disperse power, $\rho^*(\theta) \in (0, \infty)$, and the optimal concentration of power is strictly decreasing in the candidates' office motivation, $\frac{d\rho^*}{d\theta} < 0$.*

Proposition 2.6 establishes a relation between parameter θ, capturing the candidates' motivation, and the optimal level of power concentration. Intuitively, higher office motivation makes mimicking more attractive and induces more inefficient reforms. Allocating power to reforming candidates is consequently less beneficial, so that the positive empowerment effect decreases in θ. Furthermore, higher office motivation reinforces the negative behavioral effect, since candidates respond more strongly to the electoral incentives.

Regarding the optimal constitution, we have to distinguish two cases. First, consider the case of mainly policy-oriented candidates, $\theta < \bar{\theta}$, in which mimicking is not prevalent and the average reform payoff is large. In this case, the negative behavioral effect is sufficiently small to be dominated by the positive empowerment effect for all levels of ρ. Consequently, welfare is maximized by full concentration of power. Second, consider the case of mainly office-motivated candidates, $\theta > \bar{\theta}$, in which mimicking is widespread. Hence, the behavioral effect is reinforced relative to the empowerment effect. It is then optimal to attenuate electoral incentives by decreasing power concentration. Both effects outbalance each other at some interior level $\rho^* \in (0, \infty)$ that represents the optimal institution. By the same logic, the optimal level of power concentration is reduced with any further increase in the level of office motivation.

So far, this section has studied the optimal institutional setup, given that political power is delegated through democratic elections. However, our model also allows to investigate whether the democratic selection of political leaders is desirable at all. A similar question has been addressed by Maskin and Tirole, 2004, who compare decision-making by accountable "politicians" and non-accountable "judges". A non-democratic regime allocates all political power to a randomly chosen dictator. While such a regime obviously rules out selection, it also eliminates incentives for inefficient policy choice. In our model, this non-democratic system yields the same welfare as the limiting case of a democratic system with fully dispersed power. Consequently, Proposition 2.6 implies that democratic systems with appropriately chosen power concentration always dominate the non-democratic alternative. In contrast, Maskin and Tirole (2004) find that, under certain circumstances, political decisions should rather be delegated to "judges" than to "politicians".

2.7 Empirical analysis

In this section, we analyze whether data for established democracies support our model predictions. Proposition 2.6 states that power concentration is conductive to the implementation of efficient policies if politicians exhibit low levels of office motivation, $\theta < \bar{\theta}$. At higher levels of office motivation, in contrast, it is optimal to disperse power. Moreover, the optimal degree of power concentration declines for further increases in office motivation. The implications of our model can be summarized in the following Hypothesis.

Hypothesis 2.1. *The effect of power concentration on welfare depends on the level of politicians' office motivation. If politicians are mainly policy-motivated, power concentration has significantly positive effects on welfare. If politicians are in contrast mainly office-motivated, the welfare effect of power concentration is significantly smaller or even negative.*

While this theoretical prediction can in principle be confronted with empirical data, several restrictions to data availability make a rigorous stochastical test infeasible. In particular, objective measures for the politicians' office motivation or the ability of empowered politicians do not exist for obvious reasons. We are however able to resort to some indirect measures that exist for a (only) limited set of established democratic countries. In the following, we suggest an empirical strategy based on a cross-country analysis that illustrates the consistency of our model predictions with the data.

2.7.1 Operationalization

The empirical analysis requires three basic measures. As the dependent variable, we need a measure of efficient policies. Key independent variables are the degree of power concentration within the political system and the extent of politicians' office motivation.

As a measure for *efficient policies*, we use growth in real GDP per capita (World Bank). It provides a concise and objective measure of developments that bear the potential of welfare improvements for the general public. Growth has been used as outcome variable by a number of other empirical studies on political institutions as Feld and Voigt (2003) and Enikolopov and Zhuravskaya (2007). Other frequently used outcome measures relate to fiscal policy (see Voigt, 2011), which is not addressed in our model.

Several measures of democratic institutions have been discussed in the literature. Lijphart's index of the executive-parties dimension captures the *concentration of power* that is implied by the set of political institutions (Lijphart, 1999). This well-established measure quantifies how easily a single party can take complete control of the government. We revert the original index, so that high values of our explanatory variable correspond to high concentration of power within the political system. Index values are provided for 36 economically developed countries with a long democratic tradition. The measure is based on the period

1945-1996. New Zealand underwent major constitutional changes after 1996 and is thus excluded from the analysis. Its inclusion, however, does not change the qualitative results.

While *office motivation* cannot be measured objectively, indication for it may come from voter surveys. The International Social Survey Programme (ISSP) includes questions on voters' opinions about politicians.[11] The item relevant to our study was included in its 2004 survey (ISSP Research Group, 2012), which was performed in most democratic states: "Most politicians are in politics only for what they can get out of it personally." Agreement with this statement is coded on a five point scale. We use the mean points of all survey participants in a country as our measure for the importance of office motivation. That means we assume that differences in this item reflect differences in politicians' motives.[12]

For an easy interpretation of regression results, we normalize the indices for both office motivation and power concentration to range between zero and one. High values indicate pronounced office motivation of political leaders or a strong concentration of political power, respectively.

2.7.2 Design

Our analysis focuses on countries with a similar degree of democratization. We require that all countries be established democracies as identified by the 2002 Polity IV Constitutional Democracy index (Marshall and Jaggers, 2010). All countries have to feature an index of 95 or higher, which excludes Venezuela from the sample. The remaining 18 countries in the sample are similar with respect to their economic characteristics. In particular, all countries are economically highly developed, as classified by the World Bank. They furthermore feature a Human Development Index (HDI) of at least 0.9 in 2004, which places them in the top quintile of all countries.[13]

The time-invariant regressors require a cross-country analysis. All explanatory variables correspond to 2004 or earlier years. As dependent variable, we use average economic growth per year after 2004. By this choice of the time horizon, we

[11] Other surveys such as the World Values Survey, the Global Barometer Survey, the Eurobarometer, or the European Value Survey query trust or confidence in institutions, such as the political parties and the national parliament. Such questions only indirectly relate to politicians' motivation.

[12] Alternatively, one could use measures that are based on experts' assessments like the Corruption Perception Index from Transparency International and the Worldwide Governance Indicators from Kaufmann, Kraay, and Mastruzzi (2009). However, these indices focus on rent extraction and not on private motivations of politicians in general.

[13] The similarity in socioeconomic development was formulated as a major prerequisite for cross-country analyses in Armingeon (2002).

limit address potential problems of reverse causality as the explanatory variables cannot be affected by our explained variable.[14]

To test for an interaction effect between our main explanatory variables, power concentration and office motivation, we include the product of both variables in the regression.[15] We control for variables that may be correlated with both our explanatory variables and our explained variable. Most notably, past economic performance affects growth (see, e.g., Barro, 1991; Sala-i-Martin, 1994) and may also alter voters' perception of politicians. We hence control for GDP per capita in 2004. Besides, the empirical growth literature has identified other variables that robustly affect growth, such as capital accumulation, school enrollment rates, life expectancy, or openness of the economy (see, e.g., Sala-i-Martin, 1997). Given the size of our country sample, it is however impossible to control for all these variables in one regression. To capture these influences and to keep the number of explanatory variables low, we add past growth in real GDP per capita (from 1991 to 2004) to the regression.[16]

2.7.3 Results

For a first description of the data, we split the country set at the median value of politicians' office motivation. Figure 2.3 plots growth against concentration of power separately for the two sets of countries. The left panel depicts the relationship for countries in which politicians' office motivation is below its median value. For this group of countries, the figure does not reveal a clear pattern. The right panel depicts the relationship for countries in which politicians' office motivation is above its median value. For this set, countries with more concentrated power seem to experience less economic growth. For both groups of countries, the bivariate correlations between power concentration and growth support this observation.[17]

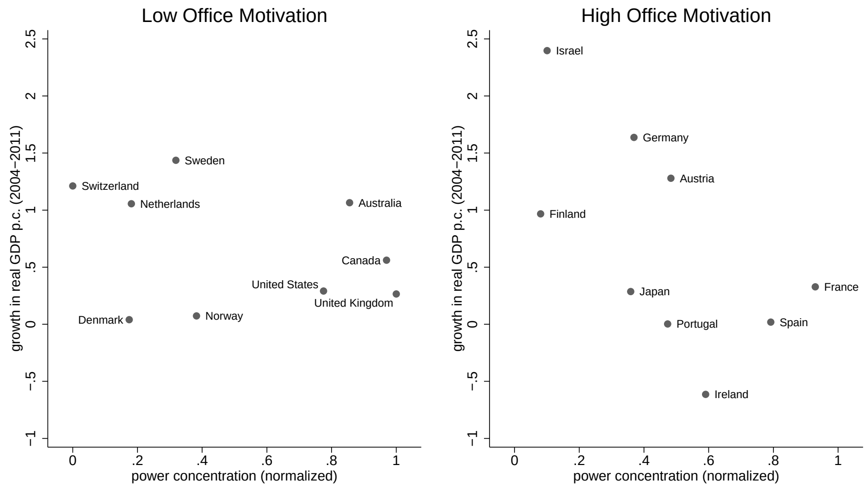
[14]It is instead possible that our explained variables are affected by past growth prior to 2004. If growth before and after 2004 were correlated, this might lead to a spurious correlation between our explanatory and explained variables. We eliminate this possibility by controlling for growth before 2004 explicitly in our regression model.

[15]The analysis of an interaction effect can be problematic if the interacting variables are highly correlated. However, we find no significant correlation between power dispersion and office motivation (Pearson's correlation coefficient $\rho = -0.199$, $p = 0.427$). This also suggests that historically developed political institutions within a country do not exhibit the optimal level of power concentration.

[16]Descriptive statistics for all variables are provided in Appendix C.

[17]For countries with high levels of politicians' office motivation, there is a negative and weakly significant relationship between growth and power concentration (Pearson's correlation coefficient, $\rho = -0.618$, $p = 0.076$), while there is no significant relationship between the two variables for countries with low levels of politicians' office motivation (Pearson's correlation

Figure 2.3: Power concentration, office motivation and growth: Empirical patterns



For a statistical test of the effects of power concentration on economic growth, we conduct an regression analysis that also controls for relevant covariates. For this analysis, we use office motivation as a continuous explanatory variable instead of the binary measure used above. Table 2.1 presents the regression results. Test statistics are based on White heteroscedasticity-consistent standard errors.

Column (a) displays the results of a regression model without interaction term. In this regression, the coefficient of power concentration estimates the effect on economic growth under the assumption that this effect does not depend on the level of office motivation. We find that this coefficient is insignificant.

This picture changes if the interplay between power concentration and politicians' motivation is taken into account. Column (b) reports the corresponding regression results. Most importantly, the coefficient of the interaction term between power concentration and office motivation is negative and significant. Thus, power concentration is more negatively related to growth, the more office-motivated politicians are. The inclusion of the interaction term in the regression also strongly increases the explanatory power of the econometric model. The adjusted R^2 increases from 0.19 to 0.49, even though no additional information is used.

As it turns out, power concentration comes along with either increased or decreased growth prospects depending on the level of politicians' office motivation. The conditional effect of power concentration at the lowest and the highest level of office motivation in our country set are reported in Table 2.2. At the lowest level of office motivation, power concentration is positively related to growth. In contrast, this relationship is negative at the highest level of office motivation. Our

coefficient, $\rho = -0.291$, $p = 0.447$).

Table 2.1: Power concentration and growth: OLS regression results

	Growth in real GDP per capita (2004-2011) (a)	(b)
Power concentration	−0.852	3.565*
	(0.490)	(1.637)
Office motivation	−0.125	2.566*
	(1.090)	(1.182)
Power concentration · office motivation		**-8.948****
		(3.520)
Real GDP per capita in 2004 (in $ 1000)	−0.025	−0.0436*
	(0.025)	(0.021)
Growth in GDP per capita (1991-2004)	−0.267***	−0.352***
	(0.079)	(0.059)
Constant	2.382**	1.845**
	(1.090)	(0.806)
adjusted R^2	0.19	0.49
F	4.38	14.17
N	18	18

Standard errors are provided in brackets. ***, **, * indicate significance at the 1-, 5-, and 10-percent level, respectively.

analysis thus leads to the following result.

Result 2.1. *The higher is office motivation, the more negative is the relation between power concentration and growth. Furthermore, power concentration is negatively related to growth if politicians' office motivation is high. If politicians are mainly policy-motivated, power concentration comes along with increased growth.*

We conclude that the data is in line with the model presented in this chapter. We do not only observe a negative and significant interaction effect, but also that the effect of power concentration changes its sign as suggested by the theory. Moreover, taking this interaction effect into account increases explanatory power considerably. In the analysis of political institutions, neglecting the interplay between power concentration and politicians' office motivation thus conceals actual patterns and yields misleading conclusions.

Table 2.2: Conditional effects of power concentration

	lowest office motivation	highest office motivation
Coefficient	3.565*	−5.382**
Standard error	1.637	1.955

The table depicts the coefficient of power concentration for the lowest level of office motivation ($\theta = 0$) and for the highest level of high office motivation ($\theta = 1$). ***, **, * indicate significance at the 1-, 5-, and 10-percent level, respectively.

2.7.4 Discussion of empirical results

We conduct several robustness checks for our empirical analysis. In the following, we discuss the use of different indicators for our main variables, a possible impact of the financial crisis on our results, and an alternative explanation for our result.

First, we check whether the negative and significant interaction term between power concentration and politicians' office motivation is robust to the use of different measures for our key variables. Instead of politicians' motivation from the ISSP, we also use confidence in political parties as contained in the third wave of the World Values Survey (WVS) concluded in 1998 (WVS, 2009). Using this measure and adjusting the GDP and growth variables to the survey date, the interaction effect remains negative and significant (p=0.009, F=1141.31, N=10). Unfortunately, the set of countries covered both by the third wave of the WVS as well as by Lijphart is even smaller than for our preferred model. Other surveys on politicians' office motivation have been conducted only very recently and are thus not applicable within our research design.

The measure for power concentration by Lijphart (1999) is available in a more current version from Armingeon et al. (2011). The use of this indicator yields a highly significant interaction term (p=0.009, F=9.95, N=17). Armingeon et al. (2011) also provide a modified index that focuses on institutional factors only. It is based on the variables "electoral disproportionality" and "number of parties" and is invariant to behavioral factors such as "absence of minimal winning coalitions" included in the original index. Using this measure instead, the results remain significant (p=0.061, F=15.99, N=17). We also use three different measures that capture important aspects of power concentration and find similar patterns. For the index for checks and balances (Keefer and Stasavage, 2003) and a plurality electoral system dummy (Beck et al., 2001), the interaction effect is significant and shows the expected sign ($p = 0.046, F = 17.26, N = 18$ and $p = 0.087, F = 8.91, N = 18$, respectively). For the nine-categorial type of electoral system (IDEA, 2004), however, the coefficient is insignificant ($p = 0.159, F = 8.53, N = 18$).

Second, one might fear that our result is influenced by the financial crisis which affected output beginning in 2008. To ensure that the financial crisis does not drive patterns in the data, we may restrict explained GDP growth to the years 2004-2007. For this shorter period, the interaction term between power concentration and office motivation remains weakly significant (p=0.092, F=4.20, N=18). Using the World Values Survey for our measure of politicians' office motivation we can expand explained GDP growth to the years 1998-2007. These data provide a similar picture (p=0.062, F=127.54, N=10). An alternative approach to deal with the financial crisis is to exclude countries that were particularly affected. The result is robust to the one-by-one exclusion of each country in our sample from the analysis (all p-values below 0.068, F above 4.82) as well as to the exclusion of any subset of the countries Ireland, Spain, and Portugal (all p-values below 0.031, F above 4.82), which were hit most severely by the financial crisis.

Finally, the empirical result could be explained by a different channel. In particular, it could be argued that the main role of political institutions is to discipline rent-seeking politicians. In particular, power-dispersing institutions may restrict rent extraction in office, which would be more important the more politicians value rents. However, the altered behavior of politicians would also affect the possibility to screen politicians in office and to reelect only good ones. Empirically, we cannot distinguish between our explanation and this alternative, since measures for politicians' office motivation may capture not only preferences for power per se, but also for rent extraction. Besley and Smart (2007) however study this alternative channel in a post-election model, investigating the effects of four fiscal restraints that limit the office holders' discretion. According to their theoretical results, three of these constraints enhance welfare only if the share of benevolent politicians is sufficiently large. This suggests that power dispersion enhances welfare only if office motivation is low, which is in contrast to our theoretical results and the empirical findings.

2.8 Extensions

In the following section, we discuss two modifications of the theoretical model studied above. First, we allow for heterogeneous policy preferences of voters in the sense that reforms may benefit a majority of voters, but harm some minority in the society. We show that all result derived above continue to hold. With heterogeneous policy preferences, however, the question arises whether political institutions can help to secure the rights of the minority. Second, we show that limited commitment, i.e., the possibility to withdraw a proposal after the election with a certain probability, does not change the results. For both extensions, we slightly simplify the model by abandoning the assumption of a continuous policy space. Instead, we assume that candidates can only choose between two policy proposals, the status quo and the full reform.[18]

2.8.1 Heterogeneous preferences

In political philosophy as well as public debate, a major virtue of power dispersion is seen in the political representation of minorities and the prevention of a tyranny of the majority. For example, James Madison argues in the Federalist #51 that "the rights of the minority will be insecure" without proper checks and balances (Madison, 1788a). So far, our analysis has abstracted from this aspect of political institutions in order to emphasize effects of power dispersion that are independent of minority rights.

To incorporate heterogeneity in voters' policy preferences into our model, we may assume that voters differ in their benefit from a reform rather than in their candidate preferences. In particular, voter k receives a payoff of μ_k if a reform is successfully implemented. Let the preference parameter μ_k be symmetrically distributed according to the pdf $\xi(\mu)$ and the cdf $\Xi(\mu)$ with full support on some interval $\left[\underline{\mu}, \bar{\mu}\right]$. We assume that the mean preference is larger than the reform cost c, while $\underline{\mu} \in (0, c)$. This implies that a majority of voters is in favor of the reform, as long as it is adopted by a sufficiently able candidate, while a minority unambiguously prefers the status quo.

Proposition 2.7. *If the voters have heterogeneous policy preferences according to distribution $\Xi(\mu)$, Propositions 2.1 and 2.3 to 2.6 continue to hold.*

Essentially, the proofs for all previous results hold whenever the expected vote share of a reforming candidate i is increasing in the average ability of candidates that propose a reform, i.e., in the equilibrium cutoff α_i. The basic model can be seen as the special case with a degenerate distribution function with $\mu_k = 1$ for all voters.[19]

[18] Recall that, for the main model, these are the only proposals that are made along the equilibrium path in any D1 equilibrium.

[19] Our model also allows for additional (ideological) heterogeneity with respect to the candidates. Let the reforms advocated by both candidates be targeted towards different groups of voters and let μ_{ki} denote the payoff to voter k from a successful reform by candidate i. If both parameters share the unconditional distribution $\Xi(\mu)$ defined above, Proposition 2.7 continues to hold for any correlation between μ_{k1} and μ_{k2}. With negative correlation, the candidates' reform proposals differ strongly or are even diametrically opposed (as in a stylized left-right policy space).

Given these heterogeneous policy preferences, our model allows to reconsider Madison's conjecture. Increasing power dispersion leads to higher amounts of power for candidates proposing the status quo, which is the minority's preferred option. As a consequence, the status quo is proposed more often yielding an additional increase in the minority's welfare.

Lemma 2.2. *In any informative equilibrium, the utility of each minority voter k with $\mu_k \leq c$ is strictly decreasing in the concentration of political power.*

The quote above suggests that the Founding Fathers of the United States were interested in the protection of minority rights per se. Formally, this objective can be captured by introducing inequality aversion into the welfare function, using a strictly increasing, strictly concave, and twice continuously differentiable weighting function w:

$$W_{IA} = \int_{\underline{\mu}}^{\bar{\mu}} w(V(\mu_k, \rho))\xi(\mu_k)d\mu_k.$$

In this function, $V(\mu_k, \rho)$ represents the expected utility of a voter with preference μ_k. Following Atkinson (1973) and Hellwig (2005), the relative curvature of w can be interpreted as a measure of inequality aversion. Compared to the inequality-neutral welfare function, W_{IA} puts higher weights on voters with low expected utility.

Proposition 2.8. *Any welfare function W_{IA} with inequality aversion is maximized at a lower level of power concentration than the inequality-neutral function W.*

Intuitively, power-dispersing institutions reduce the discretion of the election winner, who is chosen by the majority. The expected utility of the majority of voters is hence reduced while the minority is better off. The utility of the minority is valued strongly by an inequality averse constitutional designer. Thus, he will choose to disperse power more strongly than if he were inequality-neutral.[20]

2.8.2 Limited commitment

The assumption of full commitment is widely used to ensure tractability of models (see, e.g., Persson and Tabellini 2003). However, it may seem too restrictive that politicians can never change or adapt their agenda. In our setting, candidates

[20] Note that W_{IA} is maximized at a strictly lower level than W for any $\theta > \bar{\theta}$. For the opposite case, even constitutional designers with small degrees of inequality aversion will prefer to concentrate power completely.

with ability lower than c have an incentive to withdraw a reform proposal when they gain power. A straightforward way to introduce limited commitment into the model is to assume that, with probability $\lambda > 0$, the environment changes after the election and politicians may deviate from their proposal. For example, this could be due to an unexpected shock in the policy field or a major event in another policy field. With probability $1 - \lambda$, on the contrary, they have to carry out their proposal.

Proposition 2.9. *Suppose policy proposals are binding with probability λ. Then Propositions 2.1 and 2.3 to 2.6 continue to hold.*

This form of limited commitment increases incentives to propose a reform for low ability candidates, since they may be able to withdraw their proposal after the election. However, this only affects the level of equilibrium cutoffs and not the qualitative results.

Note that the welfare effect of reduced commitment is ambiguous. On the one hand, all candidates with ability $a_i < c$ withdraw their reforms with probability λ, thereby increasing welfare. On the other hand, as limited commitment diminishes the negative welfare contribution of a reform proposal for low ability candidates, more inefficient reforms are proposed. Thus, reform proposals become less informative to the voters, and high-ability candidates receive less political power. The worse selection of politicians as well as the more inefficient reform proposals per se represent negative effects on welfare.

2.9 Conclusion

We have investigated how the level of power concentration affects campaign behavior of politicians and social welfare if candidates are office-motivated and privately informed about their ability. Increasing the concentration of power has two effects. On the one hand, it has a positive *empowerment effect* because more power is given to election winners, who provide higher welfare in expectation. On the other hand, it also has a negative *behavioral effect*. Stronger concentration of political power reinforces the incentive for low-ability candidates to mimic more able ones. This limits the voters' capacity to identify and empower high-ability politicians.

The optimal institutional design balances both effects. We have shown that the optimal level of power concentration is negatively related to the extent of office motivation. If politicians care mainly about welfare, power concentration yields strictly positive effects. In politicians are mainly office-motivated, on the contrary, welfare is maximized by institutions that divide power between election winner and loser. Intuitively, the concentration of power induces distortions in policy choice of office-motivated candidates. The more office-motivated the candidates are, the more beneficial it is to reduce these distortions by means of power-dispersing political institutions.

In the empirical part, we have confronted these predictions with data for eighteen established democracies. Our findings are in line with the theoretically derived hypothesis. In a regression with economic growth as dependent variable, we find a negative and significant interaction effect between office motivation and power concentration. For the highest levels of office motivation, power-concentrating institutions come along with significantly lower economic growth, while we find a positive correlation for countries with the lowest levels of office motivation.

Appendix 2.A Proofs for main model

Proof of Proposition 2.1

Efficient policy choice

Since voters can directly observe candidates' abilities as well as their policies, they are able to anticipate their expected policy payoffs, $x_i\,(a_i - c)$ for $i \in \{1, 2\}$. The vote share of candidate 1 is consequently given by

$$v_1(x_1, x_2, \sigma) = 1 - \Omega(x_2(a_2 - c) - x_1(a_1 - c)),$$

which is strictly increasing in his welfare contribution $x_1\,(a_1 - c)$.

Candidate 1 chooses x_1, taking into account his opponent's strategy X_2^*, to maximize

$$\mathbb{E}_{a_2}\left[U_1(x_1, a_1)\right] = \underbrace{\int_0^1 \phi(a_2) f\left(v_1(x_1, X_2^*(a_2), \sigma), \rho\right) da_2}_{f^{FI}(x_1)} \left(\theta + x_1(a_1 - c)\right).$$

By Definition 2.1, f is monotonically increasing in v_1. As long as $f^{FI}(x_1) > 0$, this implies that candidate 1 is only interested in maximizing his welfare contribution. Moreover, candidate i can always achieve $f^{FI}(x_1) > 0$ by proposing the status quo policy, $x_1 = 0$.

Thus, the dominant strategy is given by

$$X_1^{FI}(a_1) = \begin{cases} x_i = 0 & \text{for} \quad a_i < c \\ x_i = 1 & \phantom{\text{for}} \quad a_i > c. \end{cases}$$

By symmetry, the same reasoning applies for candidate 2.

Positive welfare effect of increasing power concentration

Under full information, welfare is given by

$$W^{FI}(\rho) = \int_0^1 \int_0^1 \phi(a_1)\phi(a_2)\left[f(v_1, \rho) X_1^{FI}(a_1) + (1 - f(v_1, \rho))\, X_2^{FI}(a_2)\right] da_2 da_1$$

As argued above, equilibrium behavior is independent of ρ. The derivative of the welfare function with respect to power concentration is thus given by

$$\frac{\partial W^{FI}(\rho)}{\partial \rho} = \int_0^1 \int_0^1 \phi(a_1)\phi(a_2) \frac{\partial f(v_1, \rho)}{\partial \rho} \left[X_1^{FI}(a_1) - X_2^{FI}(a_2)\right] da_2 da_1 > 0$$

If $X_1^* > X_2^*$, we have $v_1 > \frac{1}{2}$, and $\frac{\partial f}{\partial \rho}$ is positive by Definition 2.1. If instead $X_1^* < X_2^*$, $v_1 < \frac{1}{2}$ implies that $\frac{\partial f}{\partial \rho}$ is negative. Thus, all terms below the integral with $X_1^* \neq X_2^*$ are strictly positive, while all others are zero, ensuring a strictly positive derivative.

Proof of Proposition 2.2

The following notation is used in the proofs below. We denote by $\hat{a}_i(x) \equiv \mathbb{E}\left[a_i \in [0,1] : X_i^*(a_i) = x\right]$ the expected ability that voters associate to candidate i if he proposes policy x. The vote share of candidate 1 results as $v_1(x_1, x_2, \sigma^*) = 1 - \Omega\left[x_2\left(\hat{a}_2(x_2) - c\right) - x_1\left(\hat{a}_1(x_1) - c\right)\right]$. By $\hat{f}(x) = \mathbb{E}_{a_{-i}}\left[f\left[v_i\left(x, X_{-i}^*(a_{-i})\right), \rho\right]\right]$, we denote the expected power share that agent i can gain by proposing x, given his opponent's equilibrium strategy X_{-i}^*. The proof involves three steps.

First, $X_i^*(a_i) = 0$ is true for some $a_i \in [0,1]$ in each equilibrium. Assume otherwise that $X_i^*(a_i) > 0$ for all ability levels. Because the expected ability is below c, there must be some equilibrium action x' such that $\hat{a}_i(x') < c$. Because this implies a negative expected payoff $x'\left(\hat{a}(x') - c\right)$, proposing x' leads to a smaller vote share for candidate i than the status quo proposal $x_i = 0$, given any action x_{-i} by the opponent. By the definition 2.1, we also have $\hat{f}(0) \geq \hat{f}(x')$. Moreover, $\hat{f}(0) > 0$ is true because, by the same arguments, there must be some equilibrium action x'' such that $\hat{a}_{-i} < c$. Whenever $a_i \leq c$, candidate i is strictly better off with the status quo proposal than with x':

$$\hat{f}(0)\theta > \hat{f}(x')\left[\theta + x'\left(a_i - c\right)\right]$$

Thus, x' can at most be proposed by some agent with ability $a_i \geq c$, which contradicts $\hat{a}_i(x') < c$. More generally, this implies that candidate i with $a_i < c$ strictly prefers the status quo proposal to any action $x > 0$ associated with $\hat{a}_i(x) < c$.

Second, each candidate i can propose at most one positive reform proposal $x > 0$. Assume that the strategy of candidate i involvestwo different actions $X_i^*(a') = x'$ and $X_i^*(a'') = x''$ for two ability levels $a'' > a'$. By the optimality of X_i^* for each ability level, this requires that

$$\begin{aligned}\left[\hat{f}(x'')x'' - \hat{f}(x')x'\right]\ \left(a'' - c\right) &\geq \left[\hat{f}(x') - \hat{f}(x'')\right]\theta \\ &\geq \left[\hat{f}(x'')x'' - \hat{f}(x')x'\right]\ \left(a' - c\right)\end{aligned}$$

Both conditions can only be satisfied if $\hat{f}(x'')x'' \geq \hat{f}(x')x'$. If the latter inequality were satisfied with equality, we would have $\hat{f}(x'') \neq \hat{f}(x')$. In this case, candidate i would either strictly prefer x' to x'' for all ability levels, or vice versa. This contradicts the initial assumption that x' and x'' are both played in equilibrium.

If instead $\hat{f}(x'')x'' > \hat{f}(x')x'$, which is only possible with $x'' > 0$, there is a unique cutoff α' such that candidate i prefers x'' to x' if and only if $a_i > \alpha'$. This implies that $\hat{a}_i(x'') > \hat{a}_i(x')$. As x'' can only be an equilibrium action if $\hat{a}_i(x'') \geq c$ as argued above, this implies $\hat{f}(x'') \geq \hat{f}(x')$. Consequently, the cutoff α' is below c, so that candidate i strictly prefers x'' to x' for all abilities $a_i > c$. But then, x' can at most be the optimal action for some ability levels below c, implying $\hat{a}_i < c$. Whenever $x' > 0$, however, candidate i strictly prefers the status quo proposal to x' whenever $a_i \leq c$ as argued above. Thus, $X_i^*(a_i) = x'$ can only be satisfied for some $a_i \in [0, 1]$ if $x' = 0$.

Third, assume that $X_i^*(a) = 0$ and $X_i^*(a') = b > 0$ for some abilities $a' \neq a$. The second condition can only be satisfied if $\hat{f}(b) > 0$. Then, candidate i prefers b to 0 if and only if

$$a_i \geq \frac{\hat{f}(0) - \hat{f}(b)}{\hat{f}(b)}\theta + c = \alpha_i.$$

Thus, each candidate plays a cutoff strategy as claimed in Lemma 2.2.

For $\hat{f}(0) \geq \hat{f}(b)$, we would have $\alpha_i > c$ and $\hat{a}_i(b) > c$. But this implies that $\hat{f}(b) > \hat{f}(0)$, a contradiction. For $\hat{f}(0) < \hat{f}(b)$, the cutoff is instead below c (see Lemma 2.2). To be consistent with $\hat{f}(0) < \hat{f}(b)$, α_i must however satisfy $\int_{\alpha_i}^{1} a\phi(a)da > 0$, which is equivalent with $\alpha > \underline{a}$.

Proof of Proposition 2.3

Non-robustness of $(0, b)$ equilibria with $b < 1$

The D1 criterion introduced by Cho and Kreps (1987) refines the equilibrium concept by restricting off-equilibrium beliefs. Intuitively, it requires that each deviation from equilibrium strategies must be associated to the set of types that would benefit from this deviation for the largest set of beliefs. Put differently, a deviation to some action cannot be associated to a type t if there is some other type t' such that the deviation would be profitable for an agent with type t', first, for all beliefs such that the deviation would be profitable to type t, and second, for some beliefs such that the deviation would not be profitable to type t.

Generally, the set of D1 equilibria is a subset of the set of Perfect Bayesian equilibria. In our model, this criterion eliminates all equilibria in which $X_i^*(1)$ is unequal to 1. Consider some equilibrium with $b < 1$. For an agent with $a_i < \alpha_i$, a deviation to $x = 1$ would be profitable for any belief such that

$$\hat{f}(1) > \hat{f}(0)\frac{\theta}{\theta + a_i - c} \geq \hat{f}(b)\frac{\theta + b(a_i - c)}{\theta + a_i - c}.$$

For an agent with $a_i > \alpha_i$, instead, a deviation to $x = 1$ would be profitable for any belief such that

$$\hat{f}(1) > \hat{f}(b)\frac{\theta + b(a_i - c)}{\theta + a_i - c}.$$

The right-hand side is below $\hat{f}(b)$ for all $a_i > \alpha_i$, and strictly decreasing in a_i. Thus, the set of beliefs giving rise to a profitable deviation to $x = 1$ is strictly larger for $a_i = 1$ than for all $a_i < 1$. The D1 criterion thus stipulates $\hat{a}_i(1) = 1$, implying $1\,(\hat{a}_i(1) - c) > b\,(\hat{a}_i(b) - c) > 0$ and $\hat{f}(1) > \hat{f}(b)$. Given this belief, the deviation from b to 1 is however profitable for candidate i whenever $a_i \geq c$. Thus, no equilibrium with $b \in (0, 1)$ is robust to the D1 criterion. By similar arguments, equilibria with pooling by one candidate are not robust with respect to D1.

Robustness of $(0, 1)$ equilibria

Second, the equilibria identified in Lemma 2.3 satisfy D1. Consider a deviation to any $b' \in (0, 1)$. For agents below α_i, this deviation is profitable if and only if

$$\hat{f}(b') > \hat{f}(0)\frac{\theta}{\theta + b'(a_i - c)} > \hat{f}(0).$$

For agents above α_i, the deviation is profitable if

$$\hat{f}(b') > \hat{f}(1)\frac{\theta + a_i - c}{\theta + b'(a_i - c)}.$$

As the right-hand side is strictly increasing in a_i, the deviation must be attributed to type $\alpha_i < c$ according to D1. Given this belief, we have $\hat{f}(b') < \hat{f}(0)$, so that the deviation is not profitable to candidate i for any $a_i \in [0, 1]$.

Symmetry of cutoffs

By the arguments above, candidate i proposes $x_i = 1$ if and only if his ability is above $\alpha_i \in (0, c)$, and the status quo policy otherwise. The vote share of candidate 1 depends positively on the difference $x_1\,[\hat{a}_1(x_1) - c] - x_2\,[\hat{a}_2(x_2) - c]$ and parameter ρ. Let g be the expected power share above one half, defined by

$$g(x_1 \left[\hat{a}_1(x_1) - c\right] - x_2 \left[\hat{a}_2(x_2) - c\right], \rho) \equiv f\left(v_1(x_1, x_2, \sigma), \rho\right) - \frac{1}{2}.$$

Denote by $\pi_i \equiv \hat{a}_i(1) - c$ the expected welfare contribution that the voters expect from candidate i given proposal $x_i = 1$. Using these functions, the cut-off abilities α_1 and α_2 are implicitly defined as follows.

For candidate 1, proposing the status quo gives an expected utility of

$$\left\{\Phi\left(\alpha_2\right)\frac{1}{2} + \left[1 - \Phi\left(\alpha_2\right)\right]\left[\frac{1}{2} - g\left(\pi_2, \rho\right)\right]\right\}\theta,$$

while the reform proposal $x_1 = 1$ gives an expected utility of

$$\left\{\Phi\left(\alpha_2\right)\left[\frac{1}{2} + g\left(\pi_1, \rho\right)\right] + \left[1 - \Phi\left(\alpha_2\right)\right]\left[\frac{1}{2} + g\left(\pi_1 - \pi_2, \rho\right)\right]\right\}\left[\theta + a_1 - c\right].$$

The reform incentive function R_1 measures the utility gain of candidate 1 from proposing a reform instead of the status quo, depending on a_1, and the cutoffs α_1 and α_2:

$$\begin{aligned} R(a_1, \alpha_1, \alpha_2) =& (1 - \Phi(\alpha_2))\left[\left(g(\pi_1 - \pi_2, \rho) + \frac{1}{2}\right)(\theta + a_1 - c)\right] \\ &+ \Phi(\alpha_2)\left[\left(g(\pi_1, \rho) + \frac{1}{2}\right)(\theta + a_1 - c)\right] \\ &- (1 - \Phi(\alpha_2))\left(\frac{1}{2} - g(\pi_2, \rho)\right)\theta - \Phi(\alpha_2)\theta\frac{1}{2}. \end{aligned}$$

For the cutoff ability α_1, the reform incentive is zero in equilibrium.

$$\begin{aligned} &R_1(\alpha_1, \alpha_1, \alpha_2) = 0 \\ &\Leftrightarrow \frac{\theta\left[\Phi(\alpha_2)g(\pi_1, \rho) + (1 - \Phi(\alpha_2))\left[g(\pi_1 - \pi_2, \rho) + g(\pi_2, \rho)\right]\right]}{c - \alpha_1} = \\ &\frac{1}{2} + \Phi(\alpha_2)g(\pi_1, \rho) + (1 - \Phi(\alpha_2))g(\pi_1 - \pi_2). \end{aligned}$$

Subtracting the corresponding equation for R_2, we get

$$\begin{aligned} &\frac{\theta\left[\Phi(\alpha_2)g(\pi_1, \rho) + (1 - \Phi(\alpha_2))\left[g(\pi_1 - \pi_2, \rho) + g(\pi_2, \rho)\right]\right]}{c - \alpha_1} \\ &- \frac{\theta\left[\Phi(\alpha_1)g(\pi_2, \rho) + (1 - \Phi(\alpha_1))\left[-g(\pi_1 - \pi_2, \rho) + g(\pi_1, \rho)\right]\right]}{c - \alpha_2} = \end{aligned}$$

$$
\begin{aligned}
&\Phi(\alpha_2)g(\pi_1,\rho) + (1-\Phi(\alpha_2))g(\pi_1-\pi_2,\rho) - \Phi(\alpha_1)g(\pi_2,\rho) \\
&+ (1-\Phi(\alpha_1))g(\pi_1-\pi_2,\rho) \\
&\Leftrightarrow \left[\frac{\theta\Phi(\alpha_2)}{c-\alpha_1} - \frac{\theta(1-\Phi(\alpha_1))}{c-\alpha_2} - \Phi(\alpha_2)\right] g(\pi_1,\rho) \\
&- \left[\frac{\theta\Phi(\alpha_1)}{c-\alpha_2} - \frac{\theta(1-\Phi(\alpha_2))}{c-\alpha_1} - \Phi(\alpha_1)\right] g(\pi_2,\rho) \\
&\underbrace{+ \left[(1-\Phi(\alpha_2))\left(\frac{\theta}{c-\alpha_1}-1\right) + (1-\Phi(\alpha_1))\left(\frac{\theta}{c-\alpha_2}-1\right)\right] g(\pi_1-\pi_2,\rho)}_{>0} = 0.
\end{aligned}
$$

If $\alpha_1 = \alpha_2$, this condition is trivially fulfilled. Assuming wlog $\alpha_1 > \alpha_2$, the equality above can only be satisfied if

$$
\begin{aligned}
&\left[\frac{\theta\Phi(\alpha_2)}{c-\alpha_1} - \frac{\theta(1-\Phi(\alpha_1))}{c-\alpha_2} - \Phi(\alpha_2)\right] g(\pi_1,\rho) < \\
&\quad \left[\frac{\theta\Phi(\alpha_1)}{c-\alpha_2} - \frac{\theta(1-\Phi(\alpha_2))}{c-\alpha_1} - \Phi(\alpha_1)\right] g(\pi_2,\rho).
\end{aligned}
$$

However, we have $\pi_1 > \pi_2$ by assumption, which implies $g(\pi_1,\rho) > g(\pi_2,\rho)$. Furthermore, we can show that the factor before $g(\pi_1,\rho)$ is larger than the one before $g(\pi_2,\rho)$:

$$
\begin{aligned}
&\frac{\theta}{c-\alpha_1}\Phi(\alpha_2) - \frac{\theta}{c-\alpha_2}(1-\Phi(\alpha_1)) - \Phi(\alpha_2) > \\
&\frac{\theta}{c-\alpha_2}\Phi(\alpha_1) - \frac{\theta}{c-\alpha_1}(1-\Phi(\alpha_2)) - \Phi(\alpha_1) \\
\Leftrightarrow &\frac{\theta}{c-\alpha_1} + \Phi(\alpha_1) > \frac{\theta}{c-\alpha_2} + \Phi(\alpha_2).
\end{aligned}
$$

The last inequality is clearly fulfilled, generating a contradiction. Thus, the reform incentive functions R_1 and R_2 cannot simultaneously attain zero for different cutoffs, and there are only symmetric equilibria.

Existence

Let π denote the difference in welfare contributions between a reform and a status quo proposal. Making use of the symmetric cutoffs, the incentive function simplifies to

$$
R(\alpha,\rho) = \left[\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right](\alpha - c) + \theta g(\pi,\rho) = 0.
$$

Note that $R(1,\rho)$ is always positive. If $R(0,\rho) < 0$, the reform incentive is equal to zero at least once due to continuity, and there exists an interior equilibrium. If $R(0,\rho) \geq 0$, it is an equilibrium that candidates of all abilities choose to reform. Hence, there is at least one equilibrium.

Proof of Proposition 2.4

Next, we establish uniqueness. The derivative of the incentive function with respect to α is

$$\frac{\partial R}{\partial \alpha} = \underbrace{(\theta + (\alpha - c)\Phi(\alpha))g_\pi(\pi,\rho)\frac{\partial \pi}{\partial \alpha}}_{A} + \underbrace{\left(\frac{1}{2} + (\Phi(\alpha) + (\alpha - c)\phi(\alpha))g(\pi,\rho)\right)}_{B}.$$

The reform incentive function yields that A is always larger than zero in equilibrium for the cutoff type. B is also larger than zero, due to Assumption 2.1. The reform incentive is thus throughout increasing in the cutoff. Consequently, the reform incentive attains zero for at most one cutoff value.

We use implicit differentiation to prove that there is a unique $\tilde{\theta}(\rho)$ such that the unique equilibrium is informative if and only if $\theta < \tilde{\theta}(\rho)$. If $\theta = \tilde{\theta}(\rho) < \infty$, the reform incentive is exactly zero for $\alpha = 0$. In an informative equilibrium, the derivative of the cutoff in θ is given by

$$\frac{d\alpha}{d\theta} = -\frac{g(\pi,\rho)}{(\theta + (\alpha - c)\Phi(\alpha))g_\pi(\pi,\rho)\frac{\partial \pi}{\partial \alpha} + \left(\frac{1}{2} + (\Phi(\alpha) + (\alpha - c)\phi(\alpha))g(\pi,\rho)\right)} < 0.$$

The denominator is positive (see above), as is the numerator. Thus, this derivative is strictly negative in any informative equilibrium, and $\alpha > 0$ for any $\theta < \tilde{\theta}(\rho)$. Moreover, the reform incentive function implies that $\alpha \to c$ if $\theta \to 0$. By continuity, there is a unique $\tilde{\theta}(\rho) > 0$ such that the unique equilibrium is informative if $\theta < \tilde{\theta}(\rho)$.

Proof of Proposition 2.5

Again, we use implicit differentiation to evaluate the derivative.

$$\frac{d\alpha}{d\rho} = -\frac{(\theta + (\alpha - c)\Phi(\alpha))g_\rho(\pi,\rho)}{(\theta + (\alpha - c)\Phi(\alpha))g_\pi(\pi,\rho)\frac{\partial \pi}{\partial \alpha} + \left(\frac{1}{2} + (\Phi(\alpha) + (\alpha - c)\phi(\alpha))g(\pi,\rho)\right)} < 0.$$

While the numerator is unambiguously positive, the positive sign of the denominator follows from Assumption 2.2. Hence, the overall effect is negative.

Proof of Lemma 2.1

Using the symmetry in equilibrium, welfare can be simplified considerably.

$$\frac{W(\rho)}{2} = \underbrace{\int_{\alpha}^{1} \phi(a)(a-c)da}_{z(\alpha)} \left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right).$$

Note that there is a direct effect on welfare, since the function $g(\pi,\rho)$ depends on ρ, and an indirect effect, since ρ changes the strategies of the politicians. Hence, we evaluate the total derivative of $W(\rho)$:

$$\frac{dW}{d\rho} = \frac{\partial W}{\partial \rho} + \frac{\partial W}{\partial \alpha}\frac{d\alpha}{d\rho}.$$

In the following, we denote by $D > 0$ the denominator of the derivative of α with respect to ρ.

$$\begin{aligned}
\frac{dW}{d\rho} =& \Phi(\alpha)z(\alpha)g_\rho(\pi,\rho) + \\
&+ \left\{(c-\alpha)\phi(\alpha)\left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right) + z(\alpha)\left(\phi(\alpha)g(\pi,\rho) + \Phi(\alpha)g_\pi(\pi,\rho)\frac{\partial\pi}{\partial\alpha}\right)\right\}\frac{d\alpha}{d\rho} \\
=& \left\{\Phi(\alpha)z(\alpha)\left[\theta + (\alpha-c)\Phi(\alpha)\right]g_\pi\frac{d\pi}{d\alpha} + \Phi(\alpha)z(\alpha)\left[\frac{1}{2} + (\Phi(\alpha) + (\alpha-c)\phi(\alpha))g(\pi,\rho)\right]\right. \\
&- \left[(c-\alpha)\phi(\alpha)\left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right) + z(\alpha)\left(\phi(\alpha)g(\pi,\rho) + \Phi(\alpha)g_\pi(\pi,\rho)\frac{\partial\pi}{\partial\alpha}\right)\right] \\
&\left.[\theta + \Phi(\alpha)(\alpha-c)]\right\}\frac{g_\rho(\pi,\rho)}{D} \\
=& \frac{g_\rho(\pi,\rho)}{D}\left\{\Phi(\alpha)z(\alpha)\left[\frac{1}{2} + \Phi(\alpha)g(\pi,\rho) + (\alpha-c)\phi(\alpha)g(\pi,\rho)\right]\right. \\
&\left. - \left[\phi(\alpha)\theta g(\pi,\rho) + z(\alpha)\phi(\alpha)g(\pi,\rho)\right]\frac{c-\alpha}{2g(\pi,\rho)}\right\} \\
=& \frac{g_\rho(\pi,\rho)}{D}\left\{\Phi(\alpha)z(\alpha)\left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right) - \phi(\alpha)(c-\alpha)\left[\frac{\theta}{2} + z(\alpha)\left(\frac{1}{2} + \Phi(\alpha)\right)\right]\right\} \\
=& \frac{g_\rho(\pi,\rho)}{D}\left\{\Phi(\alpha)\frac{W(\rho)}{2} - \phi(\alpha)(c-\alpha)\left[\frac{\theta}{2} + \frac{W(\rho)}{2}\right]\right\} \\
=& \frac{g_\rho(\pi,\rho)}{2D}\left\{\Phi(\alpha)W(\rho) - \phi(\alpha)(c-\alpha)\left(\theta + W(\rho)\right)\right\}
\end{aligned}$$

In any interior ($\rho < \infty$) extreme value of the welfare function, the term in brackets has to equal zero, since its factor is positive. Rearranging, we get the following necessary and sufficient condition for interior extreme values of the welfare function:

$$h(\rho) \equiv \frac{\Phi(\alpha)}{\phi(\alpha)(c-\alpha)} - \left(1 + \frac{\theta}{W(\rho)}\right) = 0.$$

Next, we prove that function h has at most one root in ρ, i.e., the welfare function attains at most one maximum. Assumption 2.2 is a sufficient condition for the first term to be decreasing in ρ and, thus, increasing in α. In any interior extreme value of the welfare function, the second term is constant in ρ. Thus, h is decreasing in ρ at each interior root and so is the term in brackets. As $h(\rho)$ is continuous in ρ, this implies that the welfare function has at most one interior maximum and no interior minimum, i.e., it is strictly quasi-concave.

Proof of Proposition 2.6

In the next step, we show how the optimal level of ρ shifts with changes in θ. The optimal level of ρ is interior whenever $\lim_{\rho\to\infty} h(\rho) < 0$, since $\lim_{\rho\to 0} h(\rho) > 0$ for all θ. For $\theta \to 0$, we get $\alpha = c$ from the equilibrium condition. The limit of $h(\rho)$ at $\theta = 0$ is given by $\lim_{\rho\to\infty} \Phi(c)W(\rho)$. This is strictly positive for all $\rho < \infty$. Hence, the optimal institution embodies full concentration of power for $\theta \to 0$. Due to continuity, this is also true for an interval around 0. In contrast, for large θ, α is close to zero if power concentration is high and $\lim_{\rho\to\infty} h(\rho) < 0$. Hence, the optimal level of power concentration is interior. Finally, we show that the optimal ρ decreases monotonically in θ at any interior maximum. Implicit differentiation gives

$$\frac{d\rho^*}{d\theta} = -\frac{\frac{dh(\rho)}{d\theta}}{\left.\frac{dh(\rho)}{d\rho}\right|_{\rho=\rho^*}}.$$

As argued before, the term in the denominator is negative. With respect to the numerator, note that the equilibrium cutoff α is decreasing in θ, $\frac{d\alpha}{d\theta} = -\frac{g(\pi,\rho)}{D} < 0$. Consequently, the same is true for welfare, $\frac{dW}{d\theta} = \frac{\partial W}{\partial \alpha}\frac{d\alpha}{d\theta} < 0$. Hence, h is monotonically decreasing in θ. In total, we conclude that $\frac{d\rho^*}{d\theta} < 0$. Overall, this implies that there is a cutoff $\bar{\theta}$ such that if and only if $\theta > \bar{\theta}$, the optimal level of ρ is smaller than infinity and strictly falling in θ.

Appendix 2.B Proofs for extensions

Proof of Proposition 2.7

The proofs in Appendix A only use one important feature of the vote share function $v_i(x_1, x_2, \sigma)$. Namely, we use that the vote share of candidate i is weakly increasing in his expected ability given a reform proposal, and thus in the difference in welfare contributions between a reform and a status quo proposal. In the following, we show that this still holds for the case of heterogeneous policy preferences. All other proofs do not change. In the new setting, voter k votes for candidate 1 if

$$x_1(\mu_k \hat{a}_1(x_1) - c) \geq x_2(\mu_k \hat{a}_2(x_2) - c),$$

where $\hat{a}_j(x_j)$ denotes the expected ability of candidate $j \in \{1, 2\}$.

If both candidates propose a reform, candidate 1 receives all votes if $\hat{a}_1(1) > \hat{a}_2(1)$, and zero if the opposite is true. Thus, the vote share is monotonically increasing in $\hat{a}_1(1)$.

If candidate 1 faces a status quo proposing opponent, his expected vote share is

$$v_1(x_1 = 1, x_2 = 0, \sigma) = \int_{\frac{c}{\hat{a}_1(1)}}^{l} \xi(\mu_k) d\mu_k.$$

The derivative with respect to $\hat{a}_1$ is strictly positive.

If candidate 1 instead proposes the status quo, he provides a certain payoff of zero to all voters, independent of his ability. Thus, the vote share does not depend on the expected ability $\hat{a}_i(0)$. Hence, the expected overall vote share of candidate i is weakly increasing in his expected competence and, thus, his welfare contribution.

Proof of Lemma 2.2

In an informative equilibrium, the expected utility of voter k with reform preference μ_k is given by

$$V(\mu_k, \rho) = 2 \int_{\alpha}^{1} \phi(a)(\mu_k a - c) da \left(\frac{1}{2} + \Phi(\alpha) g(\pi, \rho) \right).$$

It is strictly increasing in μ_k, and negative for any $\mu_k \leq c$. Its derivative with respect to power concentration follows as

$$\begin{aligned}
\frac{dV(\mu_k,\rho)}{d\rho} &= 2\Phi(\alpha)\frac{dg}{d\rho}\int_\alpha^1 \phi(a)(\mu_k a - c)da \\
&+ 2\left[\left(\phi(\alpha)g + \Phi(\alpha)\frac{dg}{d\rho}\right)\int_\alpha^1 \phi(a)(\mu_k a - c)da \right.\\
&\left. -\phi(\alpha)(c-\mu\alpha)\left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right)\right]\frac{d\alpha}{d\rho} \\
&= 2\left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right)(\Phi(\alpha) + \phi(\alpha - c))\frac{dg}{d\rho}\int_\alpha^1 \phi(a)(\mu_k a - c)da \\
&+ 2\left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right)\phi(\alpha)(c-\mu\alpha)\frac{d\alpha}{d\rho} \\
&= \frac{g_\rho}{D}\left\{[\Phi(\alpha) - \phi(\alpha)(c-\alpha)]\left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right)\int_\alpha^1 \phi(a)(\mu_k a - c)da \right.\\
&\left. -\phi(\alpha)(c-\mu\alpha)\frac{\theta}{2}\right\}.
\end{aligned}$$

For any $\rho \leq \rho^*(\theta)$, the term $\Phi(\alpha) - \phi(c-\alpha)$ is positive by Proposition 2.6 and Assumption 2.2. Thus, the expected utility of every voter with $\mu_k < c$ is strictly decreasing in ρ on this interval. By a similar argument as used in Lemma 2.1, it can be shown that $V(\mu_k,\rho)$ has at most one minimizer. For the limit case $\rho \to \infty$, however, we find that $\frac{dV}{d\rho} \leq 0$. In this limit, we have $g(\pi,\rho) = 1$, which implies $\theta \geq [1 + \Phi(\alpha)](c-\alpha)$ and a negative sign of the bracket in the last line above. Thus, $V(\mu_k,\rho)$ is monotonically decreasing in ρ.

Proof of Proposition 2.8

The proof consists of two steps. First, we show that there exists at least one maximum for some $\rho < \rho^*$. Second, we ensure that there can never be a maximum for any $\rho \geq \rho^*$. Note that the expected utility $V(\mu_k,\rho)$ is increasing in the reform preference μ_k. Due to the strict concavity of w, this directly implies that $w'(V(\mu_k,\rho)) > w'(V(\mu_k',\rho))$ for any $\mu_k < \mu_k'$. Moreover, the cross derivative of expected utility with respect to ρ and μ_k is

$$\begin{aligned}
\frac{d^2V(\mu_k,\rho)}{d\rho\, d\mu_k} =& \frac{2g_\rho(\pi,\rho)}{D}\left[\left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right)(\Phi(\alpha) + \phi(\alpha)(\alpha - c))\int_\alpha^1 a\phi(a)da \right.\\
&\left. + \frac{\theta}{2}\alpha\phi(\alpha)\right].
\end{aligned}$$

Since this term does not depend on μ_k, the marginal effect of ρ on expected utility is monotonic in μ_k. Take any welfare function of an inequality averse society:

$$W_{IA}(\rho) = \int_{\underline{\mu}}^{\bar{\mu}} w(V(\mu_k, \rho))\xi(\mu_k)d\mu_k.$$

Its derivative with respect to ρ is

$$\frac{dW_{IA}(\rho)}{d\rho} = \int_{\underline{\mu}}^{\bar{\mu}} w'(V(\mu_k, \rho))\frac{dV(\mu_k, \rho)}{d\rho}\xi(\mu_k)d\mu_k.$$

For the case of $\rho' < \rho^*$, the cross derivative $\frac{d^2V(\mu_k,\rho)}{d\rho\ d\mu_k}$ is larger than zero. All terms of it are always positive except for $(\Phi(\alpha) + \phi(\alpha - c))$. This, however, is positive for all $\rho' \leq \rho^*$ (see Proposition 2.6 and Assumption 2.2). The positive cross derivative yields

$$\frac{dW_{IA}(\rho)}{d\rho} = \int_{\underline{\mu}}^{\bar{\mu}} w'(V(\mu_k, \rho))\frac{dV(\mu_k, \rho)}{d\rho}\xi(\mu_k)d\mu_k < \frac{dW(\rho)}{d\rho} = \int_{\underline{\mu}}^{\bar{\mu}} \frac{dV(\mu_k, \rho)}{d\rho}\xi(\mu_k)d\mu_k.$$

The derivatives of the expected utility are smaller for voters with smaller μ_k. Exactly these utilities are weighted more strongly in the case of inequality aversion, since $w'(V(\mu_k, \rho)) > w'(V(\mu_k', \rho))$. Hence, the derivative of the welfare function at ρ^* is negative, and there exists at least one local maximum for some level $\rho' < \rho^*$.

Now consider the case of $\rho' > \rho^*$, where $\frac{dW}{d\rho} < 0$. From above, we know that the cross derivative is throughout either positive or negative. Suppose the cross derivative is positive. Then, we have that $\left.\frac{dW_{IA}}{d\rho}\right|_{\rho'} < \left.\frac{dW}{d\rho}\right|_{\rho'} < 0$, and there cannot be a maximum at ρ'. Suppose that the cross derivative is negative at ρ'. From Lemma 2.2, we know that the marginal effect of ρ is throughout negative for voters with $\mu < c$. As a consequence of the negative cross derivative, the marginal effect is negative for all agents. Thus, the derivative of W_{IA} is certainly negative. Overall, there cannot be any maximum in the range $[\rho^*, \infty)$.

Proof of Proposition 2.9

For the case of limited commitment, the proofs of Propositions 2.1 and 2.3 to 2.6 need to be considered one by one. We shorten the proof whenever it is analogous or very similar to the case with full commitment. The proof of Proposition 2.1 does not rely on full commitment and thus carries over to the new setting.

Proof of Proposition 2.3 with limited commitment

We just need to prove symmetry of cutoffs. The proof with regard to the classification of equilibria is identical to the case with full commitment. In equilibrium, the reform incentive with limited commitment simplifies to

$$\begin{aligned}
&R_1(\alpha_1, \alpha_1, \alpha_2) = \\
&\theta\left[\Phi(\alpha_2)g(\pi_1,\rho) + (1-\Phi_2)\left[g(\pi_1-\pi_2,\rho)+g(\pi_2,\rho)\right]\right] + \\
&\lambda(\alpha_1 - c)\left[\frac{1}{2} + \Phi(\alpha_2)g(\pi_1,\rho) + (1-\Phi(\alpha_2))g(\pi_1-\pi_2)\right] = 0 \\
&\Leftrightarrow \frac{\theta\left[\Phi(\alpha_2)g(\pi_1,\rho) + (1-\Phi(\alpha_2))\left[g(\pi_1-\pi_2,\rho)+g(\pi_2,\rho)\right]\right]}{\lambda(c-\alpha_1)} = \\
&\frac{1}{2} + \Phi(\alpha_2)g(\pi_1,\rho) + (1-\Phi(\alpha_2))g(\pi_1-\pi_2).
\end{aligned}$$

Subtracting the corresponding equation for the second player and proceeding as in the proof with full commitment, we obtain

$$\begin{aligned}
&\left[\frac{\theta\Phi(\alpha_1)}{\lambda(c-\alpha_2)} - \frac{\theta(1-\Phi(\alpha_2))}{\lambda(c-\alpha_1)} - \Phi(\alpha_1)\right] g(\pi_2,\rho) = \\
&\left[\frac{\theta\Phi(\alpha_2)}{\lambda(c-\alpha_1)} - \frac{\theta(1-\Phi(\alpha_1))}{\lambda(c-\alpha_2)} - \Phi(\alpha_2)\right] g(\pi_1,\rho) + \\
&\underbrace{\left[(1-\Phi(\alpha_2))\left(\frac{\theta}{\lambda(c-\alpha_1)} - 1\right) + (1-\Phi(\alpha_1))\left(\frac{\theta}{\lambda(c-\alpha_2)} - 1\right)\right] g(\pi_1-\pi_2,\rho)}_{>0}
\end{aligned}$$

If $\alpha_1 = \alpha_2$, this condition is trivially fulfilled. Assuming wlog $\alpha_1 > \alpha_2$, the equality above implies that

$$\begin{aligned}
&\left[\frac{\theta\Phi(\alpha_2)}{\lambda(c-\alpha_1)} - \frac{\theta(1-\Phi(\alpha_1))}{\lambda(c-\alpha_2)} - \Phi(\alpha_2)\right] g(\pi_1,\rho) < \\
&\quad \left[\frac{\theta\Phi(\alpha_1)}{\lambda(c-\alpha_2)} - \frac{\theta(1-\Phi(\alpha_2))}{\lambda(c-\alpha_1)} - \Phi(\alpha_1)\right] g(\pi_2,\rho).
\end{aligned}$$

However, we have $\pi_1 > \pi_2$. Moreover, we can show that

$$\frac{\theta}{\lambda(c-\alpha_1)}\Phi(\alpha_2) - \frac{\theta}{\lambda(c-\alpha_2)}(1-\Phi(\alpha_1)) - \Phi(\alpha_2) >$$
$$\frac{\theta}{\lambda(c-\alpha_2)}\Phi(\alpha_1) - \frac{\theta}{\lambda(c-\alpha_1)}(1-\Phi(\alpha_2)) - \Phi(\alpha_1)$$
$$\Leftrightarrow \frac{\theta}{\lambda(c-\alpha_1)} + \Phi(\alpha_1) > \frac{\theta}{\lambda(c-\alpha_2)} + \Phi(\alpha_2).$$

Thus, the reform incentive functions R_1 and R_2 can never simultaneously attain zero for $\alpha_1 > \alpha_2$. Thus, there can only be symmetric equilibria.

With respect to equilibrium existence, the reform incentive function simplifies to

$$R(\alpha,\rho) = \left[\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right]\lambda(\alpha - c) + \theta g(\pi,\rho) = 0.$$

Note that it is always positive if $\alpha = 1$. If $R(0,\rho) < 0$, the reform incentive is equal to zero at least once, due to the continuity and there exists an interior equilibrium. If $R(0,\rho) \geq 0$, it is an equilibrium that all candidates choose to reform. Hence, there is at least one equilibrium.

Proof of Proposition 2.4 with limited commitment

Next, we establish uniqueness. The derivative with respect to α is

$$\frac{\partial R}{\partial \alpha} = \underbrace{(\theta + (\alpha - c)\lambda\Phi(\alpha))g_\pi(\pi,\rho)\frac{\partial \pi}{\partial \alpha}}_{A} + \underbrace{\lambda\left(\frac{1}{2} + (\Phi(\alpha) + (\alpha - c)\phi(\alpha))g(\pi,\rho)\right)}_{B}.$$

The remainder of the proof is analogous to the case with perfect commitment.

Proof of Proposition 2.5 with limited commitment

We use implicit differentiation to prove the proposition:

$$\frac{d\alpha}{d\rho} = -\frac{(\theta + \lambda(\alpha - c)\Phi(\alpha))g_\rho(\pi,\rho)}{(\theta + (\alpha - c)\lambda\Phi(\alpha))g_\pi(\pi,\rho)\frac{\partial\pi}{\partial\alpha} + \lambda\left(\frac{1}{2} + (\Phi(\alpha) + (\alpha - c)\phi(\alpha))g(\pi,\rho)\right)}$$
$$< 0.$$

While the numerator is unambiguously positive, the positive sign of the denominator follows from Assumption 2.1.

Proof of Proposition 2.6 with limited commitment

Inserting equilibrium strategies, the welfare function can be simplified to

$$\frac{W(\rho)}{2} = \underbrace{\left(\lambda \int_{\alpha}^{1} r(a)da + (1-\lambda) \int_{c}^{1} r(a)da\right)}_{z(\lambda,\alpha)} \left(\frac{1}{2} + \Phi(\alpha)g(\pi,\rho)\right),$$

where $r(a) = \phi(a)(a-c)$. The total derivative of the welfare function can be simplified along the same lines as with full commitment and yields the same necessary and sufficient condition for extreme values:

$$h(\rho) = \frac{\Phi(\alpha)}{\phi(\alpha)(c-\alpha)} - \left(1 + \frac{\theta}{W(\rho)}\right) = 0.$$

Thus, the rest of the proof is equivalent.

For the second step, we have to show how the unique maximum changes with θ. For $\theta \to 0$, we get from the reform incentive $\alpha = c$ and

$$\left.\frac{dW(\rho)}{d\rho}\right|_{\theta=0} = \frac{g_\rho(\pi,\rho)}{D}\Phi(\alpha)\lambda W(\rho).$$

This is positive. For $\theta \to 0$, the optimal institution hence fully concentrates power. Due to continuity, we get that this is also true for an interval around 0. Since $h(\rho,\alpha)$ does not change with limited commitment, we again refer the reader to the proofs for full commitment to see that the optimal ρ is monotonically decreasing in θ.

Appendix 2.C Data

Description and sources of variables

Main variables

Growth in real GDP per capita	Average growth rate. Computed based on per capita GDP in constant 2000 US$. World Bank (2012).
GDP per capita	Denominated in constant 2000 TUS$. World Bank (2012).
Office motivation	International Social Survey Programme 2004: Citizenship I. ISSP Research Group (2012).
Power concentration	Lijphart's index for executive-parties dimension, reverted. Lijphart (1999).

Variables for robustness checks

Trust in political parties	World Values Survey, third wave. WVS (2009).
Power dispersion	Time-variant proxy for Lijphart's executive-parties dimension, year 2004. Armingeon et al. (2011).
Power dispersion, institutional	Time-variant proxy for Lijphart's executive-parties dimension, institutional factors, year 2004. Armingeon et al. (2011).
Checks and balances	Number of veto players. Keefer and Stasavage (2003).
Plurality electoral system	Dummy variable. Beck et al. (2001).
Electoral system	Type of electoral system, 9 minor categories. IDEA (2004).

Country list

Australia	Austria	Canada	Denmark
Finland	France	Germany	Ireland
Israel	Japan	Netherlands	Norway
Portugal	Spain	Sweden	Switzerland
United Kingdom	United States		

Summary of variables

	Mean	Std. dev.	Min	Max	Poss. values
Power concentration	−0.31	0.98	−1.77	1.21	[-2,2]
Office motivation	3.37	0.37	2.61	4.20	[1,5]
GDP p.c.	26.98	7.69	11.55	39.83	
GDP p.c. growth (2004-2011)	0.68	0.74	−0.61	2.40	
GDP p.c. growth (1991-2004)	2.08	1.07	0.56	5.59	

For the regression analysis, the variables power concentration and office motivation are rescaled to range between 0 and 1.

Correlation table

	Power concentration	Office motivation	GDP p.c.	GDP p.c. growth (2004-2011)
Office motivation	0.20 (0.43)	1		
GDP p.c.	−0.20 (0.43)	−0.58 (0.01)	1	
GDP p.c. growth (2004-2011)	−0.44 (0.07)	0.072 (0.78)	−0.15 (0.56)	1
GDP p.c. growth (1991-2004)	0.27 (0.28)	−0.10 (0.69)	−0.021 (0.93)	−0.48 (0.05)

Pearson's correlation coefficient, *p*-values in parentheses.

3

On the Ambiguous Sign of the Optimal Utilitarian Marginal Income Tax

3.1 Introduction

In the last decade, the seminal paper by Saez (2002) has initiated a growing literature that aims at rationalizing the Earned Income Tax Credit (EITC), the largest tax/transfer program transferring resources towards the poor in the United States. For low-income workers, the EITC specifies a negative marginal income tax and a negative participation tax, i.e., a higher transfer than the one paid to the unemployed. Strikingly, both properties are at odds with the central result of optimal taxation theory due to Mirrlees (1971), according to which the optimal marginal income tax is strictly positive everywhere below the very top. Subsequent studies have shown the robustness of this result for all models in which, first, agents adjust their labor supply only at the intensive margin, i.e., choose how many hours or how hard to work, and second, the tax designer has a utilitarian desire for redistribution from rich (high-skill) to poor (low-skill) agents.[1]

Most prominently, two approaches have been brought forward to rationalize the EITC, each abandoning one of these basic assumption and explaining one

[1] Amongst others, see Seade (1977), Seade (1982), Diamond (1998), and Hellwig (2007). Note that, under certain assumptions, the optimal marginal tax is also zero at the very bottom.

property of the EITC. First, Saez (2002) shows that negative participation taxes might be optimal if agents adjust their labor supply only at the extensive margin, i.e., only take the binary decision whether or not to enter the labor market (see also Diamond 1980 and Choné and Laroque 2011). The basic intuition behind this result is that redistributing resources from the rich towards the working poor is less costly in efficiency terms than redistributing resources towards the unemployed. In particular, a negative participation tax for low-skill workers induces inefficient labor supply responses in this skill group only, while a rising unemployment benefit gives rise to labor supply distortions in all skill groups.

Second, Choné and Laroque (2010) show that negative marginal taxes can be rationalized in an intensive-margin model if the social planner prefers to redistribute resources from agents earning low incomes on the labor market to high-income earners. In this case, the social planner's anti-utilitarian desire to redistribute resources to high-skill agents is restricted by binding upward incentive compatibility constraints, which can only be relaxed through negative marginal taxes.[2]

These studies give rise to the questions whether an EITC-style tax scheme with negative marginal taxes and participation taxes can be optimal if, first, the social planner has a standard *utilitarian desire for redistribution* from high-skill to low-skill workers and, second, agents adjust their labor supply *at the intensive and the extensive margin*, which is arguably the most appropriate assumption from an empirical perspective.

In this case, marginal income taxes induce labor supply distortions at the intensive margin, which cannot occur in extensive-margin models by construction (Saez, 2002; Choné and Laroque, 2011). Relatedly, the social planner is restricted by incentive compatibility constraints as in the classical Mirrlees (1971) framework. As a consequence, it is unclear whether the simple intuition from the extensive models is still valid. If downward incentive compatibility constraints are binding in the optimal allocation, negative participation taxes for the working poor are associated with higher efficiency costs. Additional transfers to low-skill workers must then be accompanied either by stronger downward distortions at the intensive margin, or by similar transfers to workers of all higher skill types, which is at odds with the utilitarian objective. Moreover, negative marginal taxes can only be beneficial if upward incentive compatibility constraints are binding in the optimal allocation, i.e., if more resources are transferred to some group of workers than to a slightly less productive group of workers. The literature has not

[2]In the model by Choné and Laroque (2010), agents are heterogeneous with respect to skill and, additionally, some other cost-related parameter. The authors show that an anti-utilitarian desire to redistribute from low-income to high-income workers can arise if these two type parameters exhibit a sufficiently strong correlation.

yet provided an explanation for why this might be in the interest of a utilitarian planner.

To some extent, this skepticism is confirmed by Jacquet, Lehmann, and Linden (2013) in a recent paper on optimal income taxation with labor supply responses at both margins. In particular, the authors show that optimal marginal taxes are positive everywhere below the very top whenever some sufficient condition is met. However, this sufficient condition is expressed in terms of endogenous variables, i.e., endogenous social weights and properties of the optimal allocation itself. Moreover, the relation between this condition and common assumptions on the economic primitives and the social planner's objective function remains unclear.

Contributions The first contribution of this chapter is to show that the sign of the optimal marginal income tax is *in general ambiguous* even if the social planner holds a utilitarian desire for redistribution. For some utilitarian welfare functions, the optimal marginal tax is positive everywhere below the very top. But for other utilitarian welfare functions, the optimal marginal tax is zero throughout, or even negative at some low income levels. Complimenting these general insights, the analysis in this chapter is the first to provide sufficient conditions on the primitives such that an EITC-style tax scheme is indeed optimal, giving rise to upward distortions at both margins for some skill groups.

The second contribution of this chapter is to *explain why* negative marginal taxes can be optimal in the model with labor supply responses at both margins. In contrast to the Mirrlees (1971) model, the sign of the optimal tax rate is not pinned down by a standard tradeoff between equity and efficiency. Instead, an additional tradeoff between intensive efficiency and extensive efficiency aspects arises, which has not been discussed in the literature so far. In section 3.6, I show that both aspects of efficiency can be disentangled using an inverse elasticity rule. As will become clear below, this tradeoff between intensive efficiency and extensive efficiency drives the ambiguity of the optimal marginal tax: inducing upward distortions at the intensive margin through negative marginal taxes can be optimal if and only if this helps to reduce labor supply distortions at the extensive margin.

The final contribution of this chapter is to show that the potential optimality of the EITC depends crucially on the assumed information structure. Following the related literature, I study a model in which agents are heterogeneous with respect to two type dimensions, skills and fixed costs of working. I show that an EITC-style tax scheme can be optimal in this framework if and only if agents possess private information about both type dimensions. In contrast, optimal utilitarian marginal taxes and participation taxes are always non-negative if the planner is able to observe either skills or fixed costs of working directly. Put differently, the optimal directions of labor supply distortions at both margins are ambiguous in multi-dimensional screening problems, while they are pinned down uniquely in problems of one-dimensional screening.

The chapter proceeds as follows. I introduce the basic model in section 3.2 and impose a set of regularity conditions in section 3.3. Section 3.4 introduces the problem of optimal income taxation and some relevant terminology. In section 3.5, I first derive the main results on the ambiguous sign of the optimal marginal taxes and participation taxes. Then, I provide sufficient conditions for the optimality of specific non-standard tax schedules, including an EITC-style tax scheme that induces upward distortions at both margins for some skill groups. Section 3.6 studies an auxiliary problem that helps to develop an economic intuition for this ambiguity and work out the tradeoff between intensive efficiency and extensive efficiency. Section 3.7 studies optimal utilitarian taxation under the alternative assumptions that either skills or fixed costs are publicly observable. Section 3.8 discusses the relevance of the imposed assumptions. Section 3.9 reviews the related literature, and section 3.10 concludes. All formal proofs are relegated to the mathematical appendix.

3.2 Model

I study optimal Utilitarian income taxation in an economy with labor and one homogeneous good. There is a continuum of agents of mass one, each of whom is identified with a two-dimensional type (ω, δ). For reasons that will become clear below, I refer to $\omega \in \Omega$ as the skill type, and to $\delta \in \Delta$ as the fixed cost type. The skill type space Ω and the cost space Δ are compact sets, with $\underline{x}$ and $\bar{x}$ denoting the smallest and largest value of $x \in \{\omega, \delta\}$. Each agent's skill type ω and cost type δ are the realizations of two random variables $\tilde{\omega}$ and $\tilde{\delta}$ with joint probability distribution Ψ. The distribution Ψ is identical for all agents, and has full support on the type space $\Omega \times \Delta \in \mathbb{R}_+ \times \mathbb{R}$. Imposing a law of large numbers, I assume that Ψ also represents the cross-section distribution of types in the continuum of agents.[3]

The agents supply labor and consume the homogeneous good. If an agent with type (ω, δ) consumes c units and supplies labor to produce y units of this good, he receives a utility of $V(c, y, \omega, \delta)$. An allocation is given by two functions $c(\omega, \delta) \geq 0$ and $y(\omega, \delta) \geq 0$ that specify the consumption level and output level for each type

[3]For conditions justifying this approach, see Sun (2006).

in $\Omega \times \Delta$. It is feasible if and only if overall consumption does not exceed overall output, i.e.,

$$\int_{\Omega\times\Delta} c(\omega,\delta)d\Psi(\omega,\delta) \leq \int_{\Omega\times\Delta} y(\omega,\delta)d\Psi(\omega,\delta) \tag{3.1}$$

Each agent is privately informed about his skill ω and fixed cost δ. Thus, an allocation can only be implemented if it is incentive-compatible, i.e., if

$$V\left(c(\omega,\delta), y(\omega,\delta), \omega, \delta\right) \geq V\left(c(\omega',\delta'), y(\omega',\delta'), \omega, \delta\right) \tag{3.2}$$

for all types (ω,δ) and (ω',δ') in $\Omega \times \Delta$. Normative comparisons of allocations are enabled by the welfare function

$$\int_{\Omega\times\Delta} U\left[V\left(c(\omega,\delta), y(\omega,\delta), \omega, \delta\right)\right] d\Psi(\omega,\delta) \tag{3.3}$$

The welfare function integrates over all agents' utilities, subject to some positive-monotone transformation U. Its properties capture the planner's objective with respect to redistributive taxation, beyond the properties of the utility function V. Thus, the desirability of redistribution depends on both V and U. To guarantee existence of a solution, let $\lim_{z\to\infty} U'(z) \leq 1$.

3.3 Assumptions

Throughout the chapter, I will impose the following assumptions.

Regularity Conditions (RC): *The utility function $V : \mathbb{R}^4 \mapsto \mathbb{R}$ is twice continuously differentiable in c, ω, δ and, for $y > 0$, in y. It is strictly concave and increasing in c. For $y > 0$, it is strictly concave and decreasing in y, increasing in ω and decreasing in δ.*

Strict Single-Crossing (SSC): *For all $(c, y, \omega) \in \mathbb{R}^3_{+++}$, the utility function satisfies*

$$\frac{\partial}{\partial\omega}\left[\frac{V_c(c,y,\omega,\delta)}{V_y(c,y,\omega,\delta)}\right] < 0 \tag{3.4}$$

Assumptions *RC* and *SSC* are standard in the literature, and will not be discussed further.

Additive Fixed Costs (AFC): *The utility function consists of a gross utility component $\tilde{V}$ and an additively separable fixed cost component δ:*

$$V(c, y, \omega, \delta) = \tilde{V}(c, y, \omega) - 1_{y>0}\delta \tag{3.5}$$

Function U is twice continuously differentiable and strictly increasing in its argument, while $\tilde{V}$ inherits the properties of V with respect to c, y, and ω.

Assumption *AFC* is made for tractability, allowing to study the optimal tax problem with the random participation approach due to Rochet and Stole (2002). It has also been made in related papers on optimal taxation with labor supply responses at the extensive margin (Jacquet, Lehmann, and Linden, 2013; Choné and Laroque, 2011).

Quasi-Linearity in Consumption (QLC): *The gross utility component $\tilde{V}$ is quasi-linear in consumption:*

$$\tilde{V}(c, y, \omega) = c - h(y, \omega) \tag{3.6}$$

For $(y, \omega) \in \mathbb{R}^2_{++}$, the effort cost function is strictly increasing and convex in y, strictly decreasing in ω and has a strictly negative cross derivative $h_{y\omega}(y, \omega)$. For any $\omega \in \Omega$, the effort cost function satisfies $h(0, \omega) = 0$ and the Inada conditions $\lim_{y\to 0} h_y(y, \omega) = 0$ and $\lim_{y\to\infty} h_y(y, \omega) = \infty$.

Assumption *QLC* rules out income effects in labor supply, which considerably simplifies the analysis. For this reason, it has also been imposed in a number of related papers, including Diamond (1998). Moreover, it implies that the desirability of redistribution depends only on the properties of transformation U in the planner's objective function. For example, if transformation U were given by the identity function, welfare could not be increased through redistributive taxation.

Relevance of Extensive Margin (REM): *For any type (ω, δ), let*

$$y^{LF} = \arg\max_{y} V(y, y, \omega, \delta) \tag{3.7}$$

be the output level that an agent of type (ω, δ) would choose under laissez-faire. Heterogeneity in fixed costs is large enough to ensure that $y^{LF}(\underline{\omega}, \underline{\delta}) > 0$, and $y^{LF}\left(\bar{\omega}, \bar{\delta}\right) = 0$ hold.

By Assumption *REM*, every skill group would involve active workers and unemployed agents without redistributive income taxation. This guarantees that changes in the tax schedule will induce labor supply responses at the extensive margin by agents of all skill groups in some neighborhood of the laissez-faire allocation. The assumption is imposed to work out very clearly the differences to the standard Mirrleesian framework, where agents adjust their labor supply at the intensive margin only.

Discrete Skill Space (DSS): *The skill space Ω is given by the finite set $\{\omega_1, \omega_2, \ldots, \omega_n\}$ with $\omega_{j+1} > \omega_j$ for all natural numbers below n. The cost space Δ is given by some interval $\left[\underline{\delta}, \bar{\delta}\right]$ on the real line.*

By assumption *DSS*, the skill space is discrete, while the cost space is continuous. While this type space corresponds to the model studied by Saez (2002), it differs from Choné and Laroque (2011) and Jacquet, Lehmann, and Linden (2013) who consider models in which Ω and Δ are both given by an interval.

The next two assumptions restrict the joint type distribution Ψ, rewritten as $(F, G_1, \ldots, G_n)$. F denotes the cumulative distribution function of skills, with $f_j > 0$ representing the probability that an agent has skill type $\omega_j \in \Omega$. G_j denotes the cdf of fixed costs in the group of agents with skill type ω_j, and has a corresponding pdf g_j that is strictly positive if and only if $\delta \in \Delta$.

Log-Concave Fixed Cost Distributions (LC): *In all skill groups $\omega_j \in \Omega$, the distribution of fixed costs G_j is strictly log-concave, i.e., the inverse hazard rate $\frac{G_j(\delta)}{g_j(\delta)}$ is strictly increasing on the cost space Δ.*

This regularity assumption is satisfied for most commonly used distributions, including the uniform, normal, log-normal, exponential and Pareto distributions.

Ordered Fixed Cost Distributions (OFCD): *For each pair of skill levels ω_j and ω_{j+1} in Ω, the skill-dependent fixed cost distributions satisfy*

(i) $G_{j+1}(\delta) \geq G_j(\delta)$ *for all* $\delta \in \Delta$, *and*

(ii) $\frac{G_{j+1}(\delta)}{g_{j+1}(\delta)} \geq \frac{G_j(\delta)}{g_j(\delta)}$ *for all* $\delta \in \Delta$.

By the first part of Assumption *OFCD*, G_j weakly dominates G_{j+1} in the sense of first-order stochastic dominance. By the second part, the inverse hazard rate at any cost level δ is larger for high-skill groups than for low-skill groups. In general, both properties are closely related but not identical (for the uniform distribution, the second property is implied by first-order stochastic dominance). Note that *OFCD* covers the benchmark case of independence, in which $G_j(\delta) = G(\delta)$ for all $\delta \in \Delta$ and all $\omega_j \in \Omega$. As will become clear below, the results of this chapter depend crucially on this assumption.[4]

The final assumption restricts the social objective as captured by the positive-monotone transformation U. To simplify notation, define the *endogenous social weight* α_j of workers of skill levels ω_j in allocation (c, y) by

$$\alpha_j^U(c,y) \equiv \frac{1}{\bar{\alpha}(c,y)} \mathbb{E}_\delta \left[U' \left(V \left(c(\omega_j,\delta), y(\omega_j,\delta), \omega_j, \delta \right) \right) | \delta \in \Delta : y(\omega_j,\delta) > 0 \right] \tag{3.8}$$

and the endogenous social weight $\alpha_0^U(c, y)$ of unemployed agents by

$$\alpha_0^U(c,y) \equiv \frac{1}{\bar{\alpha}(c,y)} \mathbb{E}_{(\omega,\delta)} \left[U' \left(V \left(c(\omega,\delta), 0, \omega, \delta \right) \right) | \omega \in \Omega, \delta \in \Delta : y(\omega,\delta) = 0 \right], \tag{3.9}$$

where $\bar{\alpha}(c, y) = \mathbb{E}_{(\omega,\delta)} [U' (V (c(\omega, \delta), y(\omega, \delta), \omega, \delta))]$.

Economically, the social weight α_j measures the average welfare increase induced by a lump-sum transfer of a marginal unit to all workers with skill type ω_j, relative to the average welfare effect of a marginal lump-sum transfer to all agents in the economy, $\bar{\alpha}^U(c, y)$. Thus, the sequence of social weights measures the social planner's redistributive concerns.[5]

Desirability of Utilitarian Redistribution (DUR): *For every $\omega_j \in \Omega$, the following is true in every implementable allocation (c, y)*

$$0 < \alpha_{j+1}^U(c,y) < \alpha_j^U(c,y) < \alpha_0^U(c,y) \tag{3.10}$$

Condition *DUR* provides the rationale for optimal redistributive taxation. It implies that, if incentive considerations could be ignored, the social planner would unambiguously prefer redistributing resources from the workers within each skill group to each group of workers with lower skill type and to unemployed agents. It captures the same idea as condition *Desirability of Redistribution* in Hellwig (2007), which guarantees the optimality of positive marginal taxes in a model with labor supply responses at the intensive margin only.[6]

[4]In section 3.8, I discuss the effects of Assumption *OFCD* and the robustness of the results in more detail.

[5]By construction, the average social weight over all subgroups in the population equals unity.

[6]Note, however, that *DUR* is slightly stronger as it is assumed to hold for all implementable

In the following, I distinguish between the *economy* E and the *social objective* U as two separate parts of the optimal tax problem. I refer to the economy E as the collection of the type space $\Omega \times \Delta$, the type distribution Ψ and the utility function V.

Definition 3.1. Economy E is *regular* if and only if it satisfies assumptions *RC*, *SSC*, *AFC*, *QLC*, *REM*, *DSS*, *LC* and *OFCD*.

For any regular economy, the set of feasible and incentive-compatible allocations is uniquely pinned down. In contrast, the normative ranking of the allocations in this set is enabled by the planner's objective, in particular by transformation U.

Definition 3.2. For any regular economy E, the set of *utilitarian allocations* $\mathcal{U}(E)$ is given by all allocations that maximize some welfare function satisfying *DUR* over the set of feasible and incentive-compatible allocations.

These definitions allow to rephrase the research question of this chapter. In the following, I derive the properties of the income tax schedules that allow to decentralize utilitarian allocations. In particular, I shall show that some utilitarian allocations cannot be decentralized with positive marginal taxes.

3.4 The optimal taxation problem

The optimal taxation problem is given by the problem of maximizing social welfare by designing an income tax schedule T that maps gross income levels into tax payments, and letting each agent choose labor supply to solve the problem of household utility maximization:

Household Problem. Given individual type (ω, δ), maximize over $y \geq 0$ individual utility

$$y - T(y) - h(y, \omega) - 1_{y>0}\delta \tag{3.11}$$

Denote by $y_T(\omega, \delta)$ the gross income solving this problem for an agent with type (ω, δ), and by Y_T the set of all income levels solving this problem for some type in $\Omega \times \Delta$. I shall be interested in two key properties of the optimal utilitarian tax schedule for all income levels $y \in Y_T$. The effects of the tax schedule on individual labor supply decision depend on two characteristics.

allocations, while Hellwig (2007) assumes *Desirability of Redistribution* only for a subset of implementable allocations.

If the tax function T is continuously differentiable, the *marginal tax* $T'(y)$ is given by the derivative of T with respect to y. Under the imposed assumptions, every implementable allocation can indeed be decentralized through a continuously differentiable tax schedule.[7] As common in models with finite skill spaces, the set of implemented income levels Y_T in the optimal allocation will be finite. As a result, the optimal tax schedule might be increasing (or decreasing) over Y_T, even if the marginal tax is not positive (or negative) at any level y in Y_T.[8]

The *participation tax* $T^P(y) = T(y) - T(0)$ measures the increase in tax liabilities that an unemployed agent experiences if he enters the labor market and earns a gross income of y.[9] Depending on the sign of $T^P(y)$, the governmental budget constraint is constrained or relaxed if a positive mass of agents enter the labor market and earn gross income y.

Under the imposed assumptions, the taxation principle by Hammond (1979) and Guesnerie (1995) applies. Thus, the optimal tax problem is equivalent to the problem of maximizing the welfare function (3.3) subject to feasibility (3.1) and incentive compatibility (3.2). By standard arguments, the solution to this problem must be Pareto-efficient within the set of implementable, i.e., feasible and incentive-compatible, allocations.

Lemma 3.1. *Every implementable and Pareto-efficient allocation* $(c, y) : \Omega \times \Delta \to \mathbb{R}^2$ *can be characterized by two vectors* $\left(y_j\right)_{j=1}^n$, $\left(c_j\right)_{j=1}^n$ *and a scalar* $b \geq 0$ *such that*

- *within each skill group* $\omega_j \in \Omega$*, there is a threshold cost type* $\hat{\delta}_j \in \Delta$ *such that all*
- *all agents with types* (ω_j, δ) *such that* $\delta > \hat{\delta}_j = c_j - h(y_j, \omega_j) - b$ *provide zero output and enjoy the same consumption level* b*,*

---

[7]In the following, we thus assume that T is indeed continuously differentiable. For non-differentiable tax schedules, the implicit marginal tax $T_i'(y)$ can be defined for any consumption bundle $(y - T(y), y)$ with $y \in Y_T$. If this bundle is allocated to agents with skill type ω_j, the implicit marginal tax is given by one minus the marginal rate of substitution between output and consumption for this skill type, i.e., $T_i'(y) = 1 - h_1(y, \omega_j)$.

[8]Related to this issue, the marginal income tax is sometimes defined differently for models with discrete skill spaces. In particular, the marginal tax between two adjacent skill levels y_a and $y_b > y_a$ in Y_T can alternatively be defined as $T'(y_a, y_b) = \frac{[y_b - T(y_b)] - [y_a - T(y_a)]}{y_b - y_a}$ (see, e.g., Saez 2002). Defined this way, the marginal income tax does not convey information about the efficiency properties of implemented allocations. In contrast, the definition used here implies that the marginal tax is positive (negative) if and only if labor supply is upward distorted at the intensive margin.

[9]The term *participation tax* was first introduced by Choné and Laroque (2011). Referring to the same concept, Beaudry, Blackorby, and Szalay (2009) use the term *employment tax/subsidy*.

- *all agents with skill type ω_j and cost type $\delta \leq \hat{\delta}_j$ provide the same output y_j and enjoy the same consumption level c_j.*

By Lemma 3.1, every implementable allocation involves pooling of all unemployed agents and of all workers of the same skill group. The social planner's problem is thus reduced to choosing a universal unemployment benefit b and a consumption-output bundle for each skill type $\omega_j \in \Omega$. This simplification directly results from the additive separability of the fixed cost δ, imposed by assumption *AFC*. In the appendix, I demonstrate that the existence of a universal unemployment benefit b and identical gross utilities $c - h\left(y, \omega_j\right)$ in each skill group follow directly from implementability. In a second step, the Pareto criterion implies that all workers of the same skill group enjoy the same consumption level c_j and provide the same output level y_j.

Another implication of assumption *AFC* is that the value of employment is monotonically decreasing in δ within each skill group ω_j, while the outside option of unemployment has the same value for all types. Thus, there is at most one *threshold cost type* $\hat{\delta}_j \in \left[0, \bar{\delta}\right]$ for each skill group such that an agent with type $\left(\omega_j, \delta\right)$ weakly prefers labor market participation if and only if with $\delta \leq \hat{\delta}_j$.

Consequently, the social planner's problem can be formally defined much simpler.

Lemma 3.2. *The social planner's problem is equivalent to maximizing the utilitarian welfare function*

$$\sum_{j=1}^{n} f_j \left\{ \int_{\underline{\delta}}^{\hat{\delta}_j} g_j(\delta) U\left[c_j - h\left(y_j, \omega_j\right) - \delta\right] d\delta + \left[1 - G_j\left(\hat{\delta}_j\right)\right] U\left[b\right] \right\} \tag{3.12}$$

over $y = \left(y_j\right)_{j=1}^{n}$, $c = \left(c_j\right)_{j=1}^{n}$, *subject to the constraints*

$$b = \sum_{j=1}^{n} f_j G_j\left(\hat{\delta}_j\right)\left[y_j - c_j + b\right], \tag{3.13}$$

$$\hat{\delta}_j = \max\left\{\underline{\delta}, \min\left\{c_j - h\left(y_j, \omega_j\right) - b, \bar{\delta}\right\}\right\} \qquad \forall\, \omega_j \in \Omega, \tag{3.14}$$

$$c_{j+1} - c_j \geq h\left(y_{j+1}, \omega_{j+1}\right) - h\left(y_j, \omega_{j+1}\right) \qquad \forall\, \omega_j \in \Omega, \tag{3.15}$$

$$c_{j+1} - c_j \leq h\left(y_{j+1}, \omega_j\right) - h\left(y_j, \omega_j\right) \qquad \forall\, \omega_j \in \Omega \tag{3.16}$$

Constraint (3.13) represents the feasibility constraint. The incentive compatibility constraints along the fixed cost dimension are given by (3.14), boiled down to a set of indifference condition for the threshold cost types $\left(\omega_j, \hat{\delta}_j\right)$ in all skill groups. As argued above, the threshold worker type $\hat{\delta}_j$ and the set of active workers are uniquely determined by c_j, y_j and b for each skill level. Finally, (3.15) and (3.16) represent the set of local downward and upward incentive compatibility constraints along the skill dimension. By the single-crossing property, local incentive compatibility between all adjacent skill pairs ensures global incentive compatibility within each skill group. Note that the problem stated above does not explicitly take into account incentive-compatibility constraints between types that differ both along the skill dimension and along the fixed cost dimension. Due to the additive separability of the fixed cost component *AFC*, piece-wise incentive compatibility along each dimension guarantees global incentive compatibility between all types (ω, δ) and (ω', δ') in $\Omega \times \Delta$.

In the interest of readability, but with some abuse of terminology, I will refer to constraint (3.14) as participation constraint, and to (3.15) and (3.16) as incentive compatibility (IC) constraints. The social objective U does not appear in any of the constraints. Thus, it has no effect on the set of implementable and Pareto-efficient allocations, a subset of which is given by the set of utilitarian allocations.

The IC constraints have the same immediate implications as in the intensive model by Mirrlees (1971). First, both IC constraints can only simultaneously be satisfied if output is monotonically increasing in the skill type, $y_{j+1} \geq y_j$. Second, the single crossing property implies that the following inequality is true whenever $y_{j+1} > y_j > 0$:

$$h\left(y_{j+1}, \omega_j\right) - h\left(y_j, \omega_j\right) > h\left(y_{j+1}, \omega_{j+1}\right) - h\left(y_j, \omega_{j+1}\right) > 0$$

Thus, as long as there is no pooling across skill types with $y_{j+1} = y_j$, high-skill workers must enjoy strictly higher consumption than low-skill workers, and at most one IC constraint can be binding with respect to each pair of adjacent skill levels. In the model with labor supply responses at the intensive and extensive margin, the downward IC constraint has a third implication that does not apply in models with only one margin. The threshold cost types for high-skill groups must be strictly higher than for low-skill groups, $\hat{\delta}_{j+1} > \hat{\delta}_j$, as long as $\hat{\delta}_j$ is below the upper bound $\bar{\delta}$. This property holds because high-skill workers enjoy higher utility than low-skill workers with the same fixed cost type, whether or not there is pooling.

As argued above, Lemma 3.2 implies that any implementable allocation involves pooling of all active workers with the same skill type, and of all unemployed agents. With other words, the social planner can only vary the allocations and utility levels of agents in these $n+1$ (or less) sets simultaneously. The desirability of all viable changes is thus entirely captured by the sequence of endogenous social weights α^U, which varies over the set of implementable allocations. Assumption *DUR* requires this sequence to be strictly decreasing for any implementable allocation.

In the following, I will be interested in the efficiency properties of optimal allocations. For this purpose, it is convenient to introduce as an auxiliary function the (gross) employment surplus

$$s(y,\omega) = y - h(y,\omega). \tag{3.17}$$

By assumption *QLC*, function $s(y,\omega_j)$ has a well-defined maximizer in $\mathbb{R}_+$, which I denote as $\hat{y}_j = \arg\max_y s(y,\omega_j)$ in the following. Furthermore, denote by $\hat{s}_j = s\left(\hat{y}_j,\omega_j\right)$ the maximum level of employment surplus for an agent with skill type ω_j. The single-crossing property *SSC* implies that $\hat{y}_{j+1} > \hat{y}_j$ and $\hat{s}_{j+1} > \hat{s}_j$ for all $\omega_j \in \Omega$.

For any type (ω,δ), the efficient labor supply $y^*(\omega,\delta,v)$ and the efficient consumption level $c^*(\omega,\delta,v)$ are given as the pair of output and consumption that requires the lowest transfer of net resources to provide an agent of this type with utility level v, i.e., solves the problem

$$\min_{y,c}(c-y) \text{ subject to } V(c,y,\omega,\delta) \geq v$$

Lemma 3.3. *For any v in the domain of V, efficient labor supply is given by*

$$y^*(\omega_j,\delta) = \begin{cases} \hat{y}_j & \textit{for} \quad \delta \leq \hat{s}_j \\ 0 & \textit{for} \quad \delta > \hat{s}_j \end{cases} \tag{3.18}$$

By the quasi-linearity of V, the required utility level v does not affect the level of efficient labor supply. Using Lemma 3.1, distortions in labor supply can be defined as follows.

Definition 3.3. At the intensive margin, labor supply by workers of skill group ω_j is said to be undistorted if $y_j = \hat{y}_j$, downward distorted if $y_j \in \left(0,\hat{y}_j\right)$, and upward distorted if $y_j > \hat{y}_j$.

Definition 3.4. At the extensive margin, labor supply by workers of skill group ω_j is undistorted if $\hat{\delta}_j = \hat{s}_j$, downward distorted if $\hat{\delta}_j < \hat{s}_j$, and upward distorted if $\hat{\delta}_j > \hat{s}_j$.

3.5 Results

The results of this chapter are provided in the two following subsections. First, subsection 3.5.1 provides the main results of this chapter, which hold under the regularity assumptions imposed in section 3.3. The section mainly provides existence results, including an "anything-goes result" with respect to the sign of the optimal utilitarian marginal tax.

Second, subsection 3.5.2 provides sufficient conditions for the optimality of specific tax schedules, including an EITC-style tax schedule with negative marginal tax rates and negative participation tax rates. For this purpose, I impose further assumptions that allow me to focus on a smaller class of economies.

3.5.1 Main results

Proposition 3.1. *For every regular economy, labor supply by the workers of the highest skill group* ω_n *is undistorted at the intensive margin, and distorted downward at the extensive margin in any utilitarian allocation.*

Proposition 3.1 clarifies that the famous "no distortion at the top" result, a robust property of optimal tax schedules in intensive models á la Mirrlees (1971), continues to hold. However, it only applies to the intensive margin. At the extensive margin, labor supply of the highest skill group is strictly downwards distorted in any Utilitarian allocation.

Proposition 3.2. *For every regular economy, there is a utilitarian allocation in which labor supply of all skill groups is undistorted at the intensive margin throughout, and labor supply of some skill groups is distorted upward at the extensive margin.*

Proposition 3.3. *For some but not all regular economies, there is a utilitarian allocation in which labor supply is distorted downward at the intensive margin everywhere below the top, and distorted downward at the extensive margin everywhere.*

Proposition 3.4. *For some but not all regular economies, there is a utilitarian allocation in which labor supply of at least one skill group is distorted upward at both margins, and undistorted at the intensive margin for all other skill groups.*

Propositions 3.2 to 3.4 establish the indeterminacy of optimal marginal taxes in utilitarian redistribution programs. Propositions 3.2 and 3.3 cover extreme cases in which labor supply is either downward distorted at the intensive margin everywhere below the top, or undistorted throughout. Of course, there are also utilitarian allocations in which labor supply is downward distorted for some, and undistorted for other skill groups at this margin. Proposition 3.4 confirms the potential optimality of EITC-style tax-transfer schemes with upward distortions at both margins for some skill groups for some economies that satisfy the imposed regularity conditions. More precisely, it establishes the potential optimality of an extreme version of the EITC, in which labor supply is weakly upward distorted at the intensive margin for all skill groups. This result sharply contrasts with the unambiguous positivity of optimal marginal taxes in the intensive model (see, e.g., Mirrlees 1971 and Hellwig 2007.

The proofs of Propositions 3.1 to 3.4 are based on the analysis of a relaxed problem in which the incentive compatibility constraints between workers of different skill groups are not taken into account. In the solution to this relaxed problem, labor supply is generally undistorted at the intensive margin, because the social planner has no interest in slackening any IC constraints. In contrast to the intensive model, the solution to this relaxed problem satisfies any pair of IC constraints between skill levels ω_j and ω_{j+1} if the utilitarian welfare function is only mildly concave in the relevant range. For transformations U with sufficiently small second derivative $|U''|$, the solution to the relaxed problem actually also solves the full problem, including the complete set of IC constraints.

By Propositions 3.3 and 3.4, utilitarian allocations with downward or upward distorted labor supply at the intensive margin do not exist for all regular economies. Rather, the existence of both the "standard" case with downward distortions and of the "non-standard" case with upward distortions depend on details of the economic environment, in particular, on properties of the type set and type distribution. In the following subsection 3.5.2, I take a closer look at this issue by considering a class of economies with certain functional forms. Within this class of economies, I then provide sufficient conditions on the economic primitives–the type space $\Omega \times \Delta$, the joint type distribution Ψ, and the effort cost function h–under which utilitarian allocations with especially interesting properties exist.

3.5.2 Sufficient conditions

By Propositions 3.3 and 3.4, utilitarian allocations with labor supply distortions at the intensive margin exist for some, but not all regular economies. First, this is true for the standard constellation with downward distortions at the intensive margin everywhere below the top. Second, this also holds for extreme versions of EITC-style allocations with upward distortions at the intensive margin for some skill groups and no distortions for all other skill groups.

In this section, I provide sufficient conditions for the existence of utilitarian allocations with the discussed properties. For this purpose, I focus on a class of economies defined by the following assumption.

Assumption 3.1. *The economy satisfies the following conditions:*

(i) The effort cost function is given by $h(y, \omega) = \frac{1}{2}\frac{y^2}{\omega}$,

(ii) the skill space is given by $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ *with constant relative distances* $\frac{\omega_{j+1}}{\omega_j} = a > 1$, *and*

(iii) in each skill group $\omega_j \in \Omega$, *fixed costs are uniformly distributed on the interval* $\left[0, \bar{\delta}\right]$, *with* $\bar{\delta} > \frac{\omega_n}{2}$.

By assumption 3.1, we focus on a class of economies with simple functional forms that enable relatively simple expressions for the sufficient conditions in the remainder of this section. This includes the quadratic effort cost function, the constant relative distances between all adjacent skill levels, and the uniform distribution of fixed costs. The lower bound on $\bar{\delta}$ guarantees that agents with maximum skill and maximum fixed cost do not work under laissez-faire, as required by assumption *REM*. Note that assumption 3.1 also restricts attention to the benchmark case in which skills and fixed costs are independently distributed.

First, I provide necessary and sufficient conditions for the existence of a utilitarian allocation with standard properties, i.e., downward distortions at the intensive margin everywhere below the top.

Proposition 3.5. *If assumption 3.1 holds and* $a < 2$, *there is a utilitarian allocation in which labor supply is downward distorted at the intensive margin everywhere below the top, and downward distorted at the extensive margin everywhere.*

Proposition 3.6. *If assumption 3.1 holds,* $n = 2$ *and* $f_1 > \frac{1}{2}$, *there is a threshold* $\hat{a}(f_1) \in \left(2 + \sqrt{2}, \infty\right)$ *such that, if* $a > \hat{a}(f_1)$, *labor supply by workers of both skill groups is undistorted at the intensive margin in every utilitarian allocation.*

Note that the last result also extends to the Rawlsian (Maximin) welfare function. This is in contrast to the results by Jacquet, Lehmann, and Linden (2013) for a model with continuous skill space, according to which the Rawlsian allocation always involves downwards distortions at the intensive margin. The difference results only due to the assumed skill space with only two skill types, while all other assumption are nested in the model of Jacquet, Lehmann, and Linden (2013).[10]

[10] More precisely, maximizing the Rawlsian welfare function involves undistorted labor supply at the intensive margin under even slightly weaker conditions than those imposed in 3.6. In particular, labor supply by all workers is undistorted at the intensive margin in the Rawlsian allocation if $a > 2$ and assumption 3.1 holds.

Assumption 3.2. *The cardinality of the skill space is large enough to satisfy* $n > \inf\{z \in \mathbb{N} : z > 2 + \ln(a+1)/\ln(a)\}$. *The upper bound of the fixed cost space satisfies* $\bar{\delta} < \frac{\gamma_0 - \gamma_n}{\gamma_0 - 1}\frac{\omega_n}{2(2-\gamma_n)}$, *where* $\gamma_0 = 2 - \frac{1}{a}$ *and* $\gamma_n = 2 - \frac{a}{1+a^{2-n}(a^2-1)}$.

Assumption 3.3. *The share of agents with top skill level* ω_n *is high enough to satisfy*

$$f_n > \frac{\bar{\gamma}_{-n} - 1}{\bar{\gamma}_{-n} - \bar{\gamma}_n} \in [0, 1)$$

where $\bar{\gamma}_{-n} = \frac{\sum_1^{n-1} f_j \bar{\gamma}_j}{1-f_n}$ *and* $\bar{\gamma}_j = \gamma_0 - \frac{\omega_j}{2(2-\gamma_j)}\left(\gamma_0 - \gamma_j\right)$ *with* $\gamma_0 = \gamma_1 = \gamma_2 = 2 - \frac{1}{a}$ *and* $\gamma_j = 2 - \frac{a}{1+a^{2-j}(a^2-1)}$ *for* $j \geq 3$.

Assumption 3.2 excludes cases with particularly limited type heterogeneity. First, it requires sufficient heterogeneity in skills, depending on relative distance between two adjacent skill levels, $a = \frac{\omega_{j+1}}{\omega_j}$. Second, it imposes an upper limit on $\bar{\delta}$, so that a majority of agents with top skill ω_n participate on the labor market in the optimal allocations identified below.

Assumption 3.3 requires that the share of high-skill workers is sufficiently large. The exact level of the threshold share for f_n depends on the complete set of parameters, including the share f_j for all lower skill levels. However, it can be shown that this threshold is always below 1, and may even be negative. An increasing cardinality of the skill space, as measured by n, makes the assumption less demanding. Intuitively, assumption 3.3 seems more restrictive than Assumptions 3.1 and 3.2.

Proposition 3.7. *If assumptions 3.1, 3.2 and 3.3 hold, there is a utilitarian allocation in which labor supply by skill type* ω_2 *is upward distorted at both margins, and labor supply by all other agents is undistorted at the intensive margin.*

The economic mechanism behind this result is studied in more detail in the following section.

3.6 Intuition: The tradeoff between intensive and extensive efficiency

Propositions 3.2 to 3.4 imply that a utilitarian desire for redistribution does not pin down the direction of labor supply distortions at any margin, in contrast to the classical result in the Mirrlees (1971) model. This section aims at developing an economic intuition for the indeterminate sign of the optimal marginal tax, and its interdependence with the optimal participation tax.[11] First, I show that an elasticity rule helps to identify the optimal tax schedule. Second, I explain that, and why, labor supply responses at two margins can give rise to a tradeoff between intensive efficiency and extensive efficiency, which drives the indeterminacy of labor supply distortions. To work out this economic intuition, this section studies a simple auxiliary problem in which redistributive concerns are eliminated.

Consider an economy with only two skill levels, ω_1 and $\omega_2 > \omega_1$. The mass of low-skill agents is given by $f_1 > 0$, the mass of high-skill agents by f_2. In the social planner's objective function, let transformation U be given by the identity function. In contrast to assumption *DUR*, the social planner is thus interested in maximizing social surplus, i.e., the unweighted sum of individual utilities. Assume moreover that the planner is restricted by incentive compatibility and two additional constraints.[12] First, the allocation must satisfy a (positive or negative) exogenous revenue requirement A:

$$\sum_{j=1}^{2} f_j \int_{\underline{\delta}}^{\bar{\delta}} \left[y(\omega_j,\delta) - c(\omega_j,\delta)\right] dG_j(\delta) \geq A \tag{3.19}$$

Second, every unemployed agent with $y(\omega_j,\delta) = 0$ must receive an exogenously determined benefit b.[13]

As Lemma 3.1 applies, the social planner only has to consider allocations in which all workers with the same skill type ω_j provide identical output y_j and receive identical consumption $c_j = y_j - T_j^P + b$. Using the definition of the employment surplus $s(y_j,\omega_j) = y_j - h(y_j,\omega_j)$, the problem can be formally written as follows:

Auxiliary Problem. Maximize over y_1, y_2, T_1^P and T_2^P social surplus

$$\sum_{j=1}^{2} f_j \left[\int_{0}^{\hat{\delta}_j} \left(s\left(y_j,\omega_j\right) - T_j^P + b - \delta\right) dG_j(\delta) + \left[1 - G_j\left(\hat{\delta}_j\right)\right] b\right] \tag{3.20}$$

[11] So far, the literature has only studied the potential optimality of upward distortions at the extensive margin (Diamond, 1980; Saez, 2002; Choné and Laroque, 2011; Christiansen, 2012).

[12] Except *DUR*, all assumptions imposed in section 3.3 are taken to hold.

[13] In the following sense, the auxiliary problem can be interpreted as a part of the larger problem of optimal tax problem, rewritten as a two-step problem. In the first step, the social planner chooses (a) the amount of net resources A to be transferred from the group of high-skill agents with $\omega_j > \omega_2$ to the group of low-skill agents with skill types ω_1 and ω_2, and (b) the universal benefit to each unemployed agent b. In the second step, the planner decides how to redistribute resources within the low-skill and high-skill groups given A and b, subject to incentive-compatibility. This section focuses on the optimal amount of redistribution within the low-skill group only, and considers the benchmark case without equity concerns, i.e., $U'' \to 0$ on the relevant interval.

subject to the constraints

$$\tilde{A} = A + (f_1 + f_2)\, b \leq \sum_{j=1}^{2} f_j G_j \left(\hat{\delta}_j\right) T_j^P, \tag{3.21}$$

$$\hat{\delta}_j = s(y_j, \omega_j) - T_j^P \text{ for } j \in \{1, 2\}, \tag{3.22}$$

$$s\,(y_2, \omega_2) - s\,(y_1, \omega_2) \geq T_2^P - T_1^P, \text{ and} \tag{3.23}$$

$$s\,(y_2, \omega_1) - s\,(y_1, \omega_1) \leq T_2^P - T_1^P \tag{3.24}$$

Besides the existence of only two skill groups, there are two differences to the problem of optimal taxation defined above. First, the concave transformation U in the objective function is replaced by the identity function, which eliminates any redistributive concerns. Second, the feasibility constraint (3.21) contains the exogenous revenue requirement A. Participation constraints (3.22) and incentive compatibility constraints (3.23), (3.24) are given as before. To avoid irrelevant complications, I assume here that $\bar{\delta}$ is large enough to exceed $\hat{\delta}_j$ in every implementable allocation. Finally, recall that the unemployment benefit b is exogenously determined in the auxiliary problem.

The formal analysis of this auxiliary problem is presented in subsection 3.6.1, while subsection 3.6.2 illustrates the auxiliary problem and its solution graphically.

3.6.1 Formal analysis of the auxiliary problem

In the following, I refer to the solution of this problem, $\left(y_1^S, y_2^S, T_1^{P^S}, T_2^{P^S}\right)$, as the surplus-maximizing allocation. Lemmas 3.4 to 3.6 below imply that the level of the adjusted revenue requirement $\tilde{A} = A + (f_1 + f_2)\, b$ determines important properties of this solution, including the direction of labor supply distortions at both margins.

Lemma 3.4. *There are levels $A_{max} > 0$ and $A_{min} < 0$ such that*

(a) the auxiliary problem has a solution in $\mathbb{R}^4$ if and only if $\tilde{A} \leq A_{max}$, and

(b) this solution involves threshold worker types $\hat{\delta}_j < \bar{\delta}$ for $j \in \{1, 2\}$ if and only if $\tilde{A}\,(A_{min}, A_{max}]$.

On the one hand, the existence of unemployment as an outside option implies that the social planner can collect at most a tax revenue of A_{max}, which is realized if both skill groups are taxed at (incentive-compatible) Laffer rates. On the other hand, the auxiliary problem has a solution for any negative revenue requirement $\tilde{A} < 0$. For very negative levels of $\tilde{A}$, however, all agents of the high-skill group (or even of both groups) enter the labor market and the participation constraints are not binding anymore, i.e., labor supply becomes completely inelastic at the extensive margin. In the following, we focus on levels of the revenue requirement in the interval $[A_{min}, A_{max}]$.

Lemma 3.5. *For all $\tilde{A} \in [A_{min}, A_{max}]$, surplus maximization involves higher output by high-skill workers than by low-skill workers, $y_2^S > y_1^S$.*

(i) If $\tilde{A} \in [A_{min}, 0)$, high-skill workers receive higher participation subsidies than low-skill workers, $T_2^{P^S} < T_1^{P^S} < 0$.

(ii) If $\tilde{A} \in (0, A_{max}]$, high-skill workers pay higher participation taxes than low-skill workers, $T_2^{P^S} > T_1^{P^S} > 0$.

Lemma 3.6. *Let fixed costs in both skill groups be distributed uniformly on the interval $\left[0, \bar{\delta}\right]$, with $\bar{\delta}$ sufficiently large. There are values $A_U \in (A_{min}, 0)$ and $A_D \in (0, A_{max}]$ such that, in the surplus-maximizing allocation,*

- *high-skill labor is upward distorted at the intensive margin if $\tilde{A} \in [A_{min}, A_U)$, and*
- *low-skill labor is downward distorted at the intensive margin if and only if $\tilde{A} \in (A_D, A_{max})$.*

Thus, the relevant properties of surplus-maximizing participation taxes depend on the level of the revenue requirement $\tilde{A}$. Lemma 3.5 implies that optimal participation taxes for both skill groups are non-negative, inducing labor supply distortions at the extensive margin, whenever $\tilde{A}$ differs from zero. Moreover, high-skill workers always face either higher participation taxes or higher participation subsidies than low-skill workers. Lemma 3.6 focuses on the special case where fixed costs are distributed uniformly and identically across skill groups. For this case, the surplus-maximizing allocation also involves labor supply distortions at the intensive margin if the revenue requirement $\tilde{A}$ differs sufficiently from zero.

It will become clear below that these distortions at the intensive margin are optimal due to a tradeoff between two aspects of efficiency, in the following labeled *intensive efficiency* and *extensive efficiency*. I will then show that this tradeoff between *intensive efficiency* and *extensive efficiency* is the basis of the indeterminate sign of the optimal marginal tax established in Propositions 3.2, 3.3 and 3.4. In the remainder of this section, I focus on the case of a negative requirement $\tilde{A}$, for which surplus can be maximized through an EITC-style tax schedule inducing upward distortions in labor supply at both margins.[14]

The intuition behind both lemmas can be explained using an adapted version of the inverse elasticity (Ramsey) rule for optimal commodity taxation. Consider first a relaxed version of the auxiliary problem in which both IC constraints are ignored. For clarity, we denote by $\left(\tilde{y}_1, \tilde{T}_1^P, \tilde{y}_2, \tilde{T}_2^P\right)$ the relaxed problem's solution in the following. Without IC constraints, the social planner then treats high-skill and low-skill labor just as two separate varieties of labor, or two distinct tax bases. As there is no need to slacken an incentive constraint, optimal labor supply is undistorted at the intensive margin, $\tilde{y}_j = \hat{y}_j = \arg\max_y s(y, \omega_j)$, and the employment surplus equals its efficient level $\hat{s}_j = \max_y s(y, \omega_j)$. In the relaxed problem, *intensive efficiency* is consequently ensured. Thus, the social planner only needs to care about maximizing *extensive efficiency*, i.e., minimizing labor supply distortions at the extensive margin.

The mathematical structure of the relaxed auxiliary problem coincides with the classical Ramsey problem.[15] Consequently, the optimal pattern of taxes follows the familiar elasticity logic, according to which higher taxes or subsidies should be set for less elastic tax bases and vice versa. Formally, optimal participation taxes for both skill groups are characterized by the following version of the inverse elasticity rule

$$\tilde{T}_j^P = \frac{\lambda - 1}{\lambda} \frac{G_j(\hat{s}_j - T_j^P)}{g_j(\hat{s}_j - T_j^P)} \text{ for } j \in \{1, 2\} \tag{3.25}$$

This condition relates the optimal participation tax liability for each skill group to the semi-elasticity of its labor market participation, $\frac{g_j(\hat{\delta}_j)}{G_j(\hat{\delta}_j)} = \frac{\partial G_j(\hat{\delta}_j)/\partial\left(y_j - T_j^P\right)}{G_j(\hat{\delta}_j)}$.[16]

[14] The case of a negative $\tilde{A}$ is plausible, e.g., if the economy is additionally populated by workers with higher skill types $\omega_j > \omega_2$, from which the utilitarian planner redistributed resources towards workers with the lowest skill levels ω_1 and ω_2. More precisely, $\tilde{A}$ is negative if and only if the social planner prefers negative participation taxes, i.e., higher transfers to be paid to the working poor than to the unemployed. As Proposition 3.2 implies, there is always a well-behaved utilitarian welfare function such that negative participation taxes to the lowest skill levels are indeed optimal.

[15] A minor difference between the commodity tax and the labor tax setting is given by the elasticities of demand and supply functions. The assumptions on primitives taken here imply that labor demand is completely elastic, while labor supply is upward sloping for $r_j \in \left[0, \bar{\delta}\right]$ and completely inelastic otherwise.

[16] The semi-elasticity of participation measures the percentage increase in the participation rate that results if the net-of-tax labor income increases by one unit (instead of one percent as with the standard elasticity). In the framework of optimal commodity taxation, the inverse elasticity rule is usually expressed in terms of the standard elastity $\epsilon_j^P = \frac{y_j - T_j^P}{G_j(\hat{\delta}_j)} \frac{\partial G_j(\hat{\delta}_j)}{\partial\left(y_j - t_j^P\right)}$.

The intuition behind this rule rests on the social planner's desire to reduce distortions in labor supply as much as possible. Participation taxes differing from zero induce extensive margin responses, giving rise to labor supply distortions as agents enter (leave) the labor market which would stay unemployed (employed) in the first-best allocation. Thus, the surplus-maximizing social planner seeks to keep both participation rates T_1^P and T_2^P as close as possible to zero. This requires the optimal participation taxes for both skill groups to have the same sign, positive for $\tilde{A} > 0$ and negative for $\tilde{A} < 0$.[17] Otherwise, both participation taxes could be decreased in absolute terms, thereby also reducing labor supply distortions.

Moreover, the higher the semi-elasticity of participation for skill group ω_j, the larger is the extensive margin response induced by distributing an additional unit of resources to workers of this skill group. Consequently, it is optimal to set higher participation taxes (or higher subsidies) for the less responsive skill group.

Crucially, the relative sizes of these semi-elasticities are unambiguously pinned down by the imposed assumptions *LC* and *OFCD* on the type distribution. Recall that, in skill group ω_j, only workers with fixed cost types below some threshold $\hat{\delta}_j$ enter the labor market. As usually, incentive-compatibility implies that a high-skill worker enjoys a higher utility than a low-skill worker with the same fixed cost type δ in every implementable allocation. In contrast, the outside option of unemployment has the same value for all agents. Thus, agents in the high-skill group enter the labor market even with higher fixed costs than agents of the low-skill group, implying a higher cost threshold $\hat{\delta}_2 > \hat{\delta}_1$. The assumption of log-concavity implies that an increase in the threshold cost type $\hat{\delta}_j$ decreases the semi-elasticity $\frac{g_j(\hat{\delta}_j)}{G_j(\hat{\delta}_j)}$. Thus, the semi-elasticity of high-skill labor is smaller than the semi-elasticity of low-skill labor if skills and fixed cost are independently distributed, $G_1 = G_2$. Assumption *OFCD* also allows for some correlation, as long as the hazard rate for high-skill workers is larger, $\frac{g_2(\delta)}{G_2(\delta)} < \frac{g_1(\delta)}{G_1(\delta)}$ for all $\delta \in \Delta$. In this case, the difference between the semi-elasticities of high-skill participation and low-skill participation is even larger than in the case of independence.

By the inverse elasticity rule, the optimal ratio of high-skill to low-skill participation taxes thus exceeds unity in the solution to the relaxed problem whenever the revenue requirement differs from zero:

$$\frac{\tilde{T}_2^P}{\tilde{T}_1^P} = \frac{G_2(\hat{s}_2 - T_2^P)/g_2(\hat{s}_2 - T_2^P)}{G_1(\hat{s}_1 - T_1^P)/g_1(\hat{s}_1 - T_1^P)} > 1 \text{ for all } \tilde{A} \neq 0 \tag{3.26}$$

For any negative revenue $\tilde{A}$, equation 3.26 implies that *extensive efficiency* is maxi-

[17]Note that, for positive (negative) $\tilde{A}$, the Lagrange multiplier λ attains a level above (below) unity.

mized by setting a strictly higher participation subsidy for high-skill workers than for low-skill workers: $\tilde{T}_2^P < \tilde{T}_1^P < 0$.

This implies that the solution to the relaxed problem satisfies the downward IC constraint (3.24). Whether it also satisfies the upward IC constraint (3.23), however, is in general unclear. If this is indeed true, the relaxed problem's solution represents the surplus-maximizing allocation. Thus, it is possible to maximize extensive efficiency and intensive efficiency at the same time, and surplus-maximization does not give rise to labor supply distortions at the intensive margin.

If the relaxed problem's solution instead violates the upward IC constraint, a tradeoff between intensive efficiency and extensive efficiency arises. To maximize extensive efficiency according to the inverse elasticity rule, the social planner would like to redistribute more resources to the high-skill workers than compatible with the upward IC constraint. The upward IC constraint will consequently be binding. Moreover, the social planner can only increase extensive efficiency if he slackens this constraint by distorting labor supply of high-skill workers upwards. As this initially involves only negligible losses in intensive efficiency, surplus maximization gives rise to strict upward distortions in high-skill labor supply at the intensive margin, and strictly negative marginal taxes.[18]

The sign of the surplus-maximizing marginal tax thus depends on whether the relaxed problem's solution satisfies or violates the upward IC constraint. Without further assumptions on the revenue requirement $\tilde{A}$ and the properties of the fixed cost distributions G_1 and G_2, this can not be determined though.

Lemma 3.6 considers the simple case of identical uniform distributions of fixed costs in both skill groups. For this case, the solution to the relaxed problem violates upward incentive compatibility if the revenue requirement $\tilde{A}$ is below some threshold $A_U < 0$.[19] As argued above, the surplus-maximizing allocation consequently involves a binding upward IC and upward distorted labor supply at the

[18] The surplus-maximizing allocation then exactly balances marginal gains in extensive efficiency and marginal losses in intensive efficiency. Formally, this intuition can be captured by a generalized version of the inverse elasticity rule. For any $\tilde{A} \in [A_{min}, A_{max}]$, the surplus-maximizing participation tax $T_j^{P^S}$ for $j \in \{1, 2\}$ is characterized by

$$T_j^{P^S} = \left[\frac{\lambda - 1}{\lambda} + \frac{s_1\left(y_j, \omega_j\right)}{s_1\left(y_j, \omega_k\right) - s_1\left(y_j, \omega_j\right)}\right] \frac{G_j\left(\hat{\delta}_j\right)}{g_j\left(\hat{\delta}_j\right)} - \frac{f_k}{f_j} \frac{s_1\left(y_k, \omega_k\right)}{s_1\left(y_j, \omega_j\right) - s_1\left(y_k, \omega_k\right)} \frac{G_k\left(\hat{\delta}_k\right)}{g_j\left(\hat{\delta}_j\right)},$$

where $k \neq j$ refers to the other skill group, $\hat{\delta}_j = s\left(y_j, \omega_j\right) - T_j^P$ denotes the threshold worker type in skill group ω_j, and λ denotes the Lagrange multiplier associated with the planner's budget constraint (3.21). Note that $s_1\left(y_j, \omega_j\right) = 0$ if and only if $y_j = \hat{y}_j$, i.e., labor supply by workers of skill group ω_j is undistorted at the intensive margin.

[19] The same result holds if G_2 first-order stochastically dominates G_1, i.e., if high-skill agents have

intensive margin, $y_2 > \hat{y}_2$, in this case.

3.6.2 Graphical illustration of the auxiliary problem

Figures 3.6.2 and 3.6.2 on the following pages illustrate the tradeoff between intensive efficiency and extensive efficiency graphically for some negative revenue requirement $\tilde{A}$.

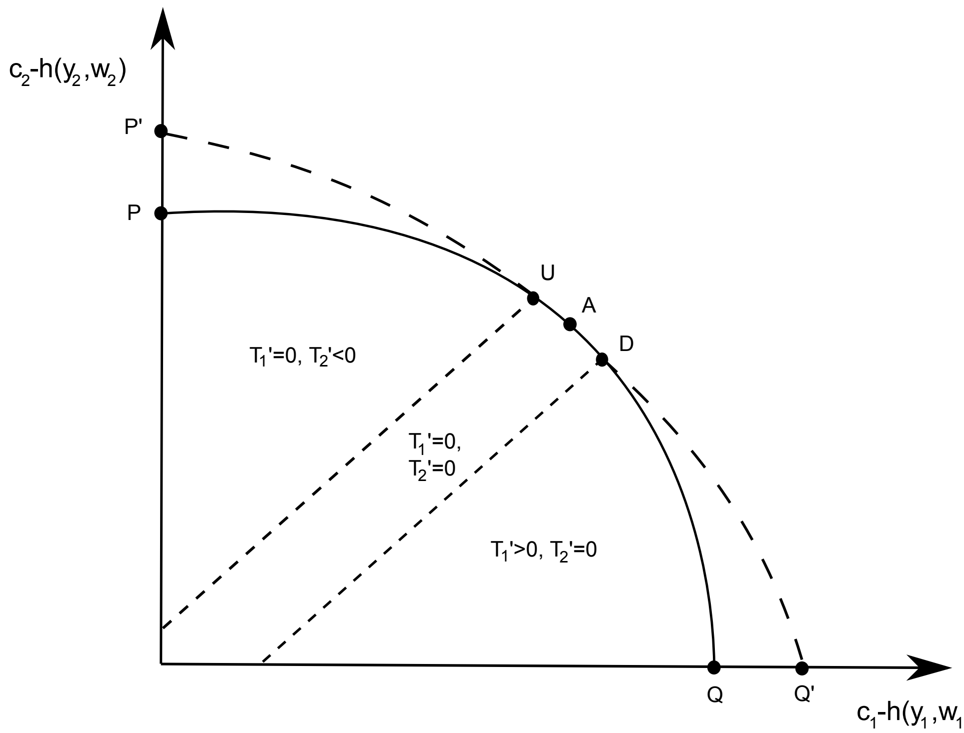
Figure 3.6.2 depicts the Pareto frontiers for the relaxed and the non-relaxed versions of the auxiliary problem. More precisely, it plots the gross utility levels $\tilde{V}_j \equiv \tilde{V}\left(c_j, y_j, \omega_j\right) = c_j - h\left(\hat{y}_1, \omega_1\right)$ of low-skill workers and high-skill workers corresponding to all (second-best) Pareto efficient allocations $\left(y_1, y_2, T_1^P, T_2^P\right)$. Recall that the utility level of a worker with type (ω_j, δ) is given by $V\left(c_j, y_j, \omega_j, \delta\right) = \tilde{V}_j - \delta$, so that $\tilde{V}_j$ represents the common element for all workers with the same skill type. In Figure 3.6.2, the gross utility $\tilde{V}_1$ of low-skill workers is depicted on the horizontal axis, while the gross utility $\tilde{V}_2$ of high-skill workers is on the vertical axis.

The dashed line $P'Q'$ represents the Pareto-frontier for the relaxed auxiliary problem, in which the social planner is not restricted by IC constraints. Moving this line down and to the right corresponds to reductions in the low-skill participation tax T_1^P, financed by an increasing level of the high-skill participation tax T_2^P. In the relaxed problem, these tax changes induce labor supply responses at the extensive margin, pulling some unemployed low-skill agents into employment and forcing high-skill workers out of the labor market. As the IC constraints can be ignored, labor supply by both skill groups is undistorted at the intensive margin in the allocation corresponding to all points on the dashed line $P'Q'$. Nevertheless, the Pareto frontier for the relaxed problem is strictly concave due to the extensive margin responses.

The solid line PQ represents the Pareto frontier for the non-relaxed auxiliary problem, which encloses the set of implementable allocations. Between points U and D, it coincides with the relaxed problem's Pareto frontier $P'Q'$. In the allocations corresponding to this interval, the participation taxes T_1^P and T_2^P are sufficiently close to each other to satisfy both IC constraints (3.23) and (3.24) even without distortions at the intensive margin. This necessarily includes point A, in which both participation taxes are identical $T_1^P = T_2^P$. The social planner can implement the allocations in this interval without distorting labor supply at the intensive margin. More generally, all combinations of $\tilde{V}_1$ and $\tilde{V}_2$ between both

overall lower fixed costs than low-skill agents. In this case, we find threshold values A'_U and A'_D that are even closer to zero, implying a larger propensity of intensive margin distortions in both directions.

Figure 3.1: The Pareto frontier



The figure shows the Pareto frontier for the auxiliary problem (solid line PQ) and the relaxed auxiliary problem (dashed line $P'Q'$). Horizontal axis: gross utility of low-skill workers, $c_1 - h(y_1, \omega_1)$. Vertical axis: gross utility of high-skill workers, $c_2 - h(y_2, \omega_2)$.

dashed lines can be implemented without intensive margin distortions, i.e., with marginal taxes $T_1' = T_2' = 0$.

To the left of point U, the solid Pareto frontier PQ for the non-relaxed problem is below the dashed line $P'Q'$. In this region, the gross utility $\tilde{V}_2$ of high-skill workers is so much higher than $\tilde{V}_1$ that the upward IC constraint would be violated without intensive margin distortions. Thus, all points on the solid Pareto frontier left of U correspond to allocations with a binding upward IC constraint, and upwards distorted high-skill labor y_2 at the intensive margin. These allocations can only be implemented with EITC-style tax schemes, involving negative marginal taxes T_2' for high-skill workers. Note also that movements along the Pareto frontier PQ thus involve labor supply responses at the intensive and the extensive margin. Moving up from point U, the upward IC constraint is tightened more and more, and can only be restored through stronger upwards distortions in y_2.

Symmetrically, the solid Pareto frontier PQ is below the dashed line $P'Q'$ to the right of point D, where the downward IC becomes binding. In the allocations below D and the lower dashed line, positive marginal taxes $T_1' > 0$ for low-skill worker induce downwards distortions in y_1, which are necessary to satisfy the downward IC constraint. Altogether, Figure 3.6.2 allows to distinguish three parts of the Pareto frontier with respect to the marginal effects on intensive efficiency. If agents would adjust their labor supply only at the intensive margin, surplus would be maximized in every point between U and D. But this is only one aspect of efficiency if labor supply also respond at the extensive margin.

Figure 3.6.2 allows to take into account extensive efficiency aspects as well. The dotted line EF depicts the locus of allocations maximizing extensive efficiency, i.e., satisfying the inverse elasticity condition[20]

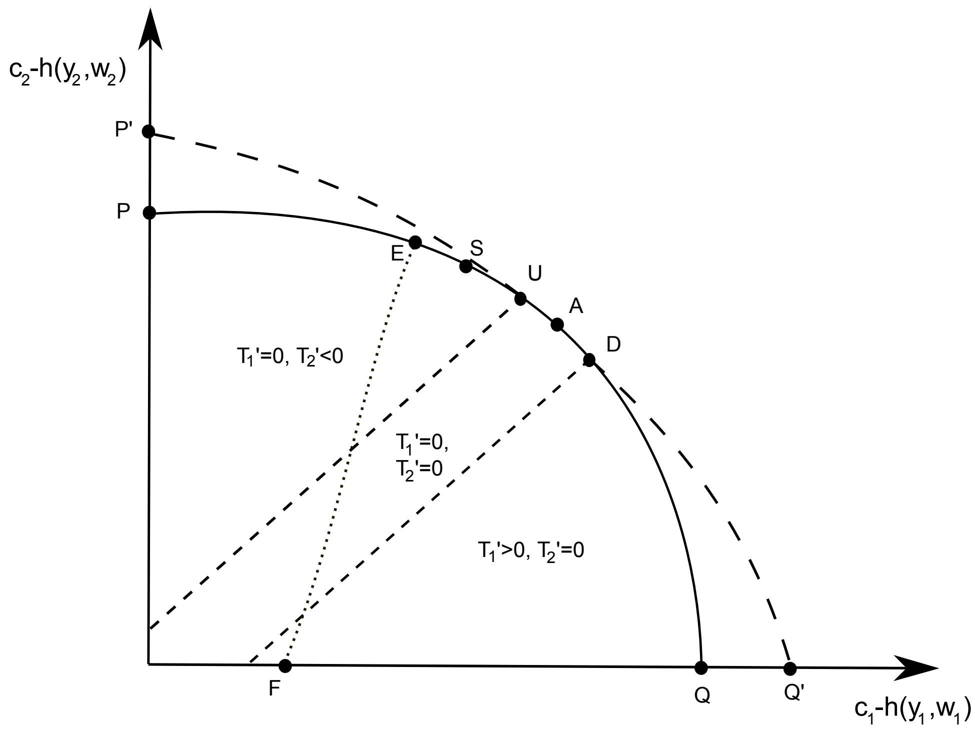
$$\frac{T_2^P}{T_1^P} = \frac{G_2\left(s\left(y_2, \omega_2\right) - T_2^P\right) / g_2\left(s\left(y_2, \omega_2\right) - T_2^P\right)}{G_1\left(s\left(y_1, \omega_1\right) - T_1^P\right) / g_1\left(s\left(y_1, \omega_1\right) - T_1^P\right)} > 1 \text{ for all } \tilde{A} \neq 0.$$

The intersection of this line with the Pareto frontier PQ is given by point E, which would be optimal if movements along the Pareto frontier would only induce labor supply responses at the extensive margin, but not at the intensive margin. Thus, the dotted line EF allows to distinguish two parts of the Pareto frontier with respect to the marginal effects on extensive margin. In particular, extensive efficiency is increased by movements down the Pareto frontier in the region left of point E, and decreased in the region right of E.

[20] Note that, for the non-relaxed problem, this condition does not necessarily involve the first-best workload levels $\hat{y}_1$ and $\hat{y}_2$. In contrast, it involves the levels of $y_1 \leq \hat{y}_1$ and $\hat{y}_2 \geq y_2$ that are required to ensure incentive-compatibility for the corresponding allocations.

The properties of the surplus-maximizing allocations thus depend on the location of point E, the intersection of this line with the Pareto frontier PQ. By Lemma 3.5, E must be located above the uniform-taxation point A for any negative revenue requirement $\tilde{A}$. Depending on the exact level of $\tilde{A}$ and the properties of G_1 and G_2, it may either lie to the left or to the right of point U, where the upward IC constraint becomes binding.

Figure 3.2: The surplus-maximizing allocation



The figure shows the allocations maximizing social surplus (S), extensive efficiency (E) and intensive efficiency (between U and D). Horizontal axis: gross utility of low-skill workers, $c_1 - h(y_1, \omega_1)$. Vertical axis: gross utility of high-skill workers, $c_2 - h(y_2, \omega_2)$.

Figure 3.6.2 illustrates the case in which this intersection is located to the left of U. Lemma 3.6 implies that this case indeed occurs under reasonable assumptions. For this case, a tradeoff between *intensive efficiency* and *extensive efficiency* arises between points E and U on the Pareto frontier. E maximizes extensive efficiency, but requires upward distortions in high-skill labor supply at the intensive margin.

In contrast, intensive efficiency is maximized at point U, which does not satisfy the inverse elasticity condition.

Starting from U and moving the Pareto frontier up towards E initially induces first-order gains in extensive efficiency, but only second-order losses in intensive efficiency. Starting instead from E and moving the Pareto frontier down towards U initially induces first-order gains in intensive efficiency, but only second-order losses in extensive efficiency. Consequently, the surplus-maximizing allocation must be located at some point S in the interior of this region, balancing marginal gains in intensive efficiency and marginal losses in extensive efficiency (see Figure 3.6.2).

Finally, the set of utilitarian allocations is given by the collection of all points on the Pareto frontier between points S and Q in Figure 3.6.2. By assumption *DUR*, the social planner would prefer to redistribute resources from high-skill workers to low-skill workers if he were not restricted by incentive considerations. Thus, any movement down the Pareto frontier induces a strict equity gain. At any point to the right of the surplus-maximizing allocation S, however, it also involves a loss in overall efficiency (combining intensive and extensive aspects). Thus, each point on the Pareto frontier below S corresponds to a utilitarian allocation. With respect to the intensive margin, this set contains allocations with upwards distortions in y_2 (between S and U), without distortions (between U and D) and with downward distortions in y_1 (between D and Q).

This clarifies that, and why, the *existence* of a utilitarian desire for redistribution does not pin down the direction of intensive margin distortions, nor the sign of the optimal marginal income tax. In cases as the one illustrated in Figure 3.6.2, this optimal sign instead depends on the *intensity* of the planner's local redistributive concerns. With a strong desire for redistribution between adjacent skill groups, he will typically prefer tax schedules with positive marginal taxes, implementing allocations in the lower right corner (between D and Q). If he instead values additional resources in the hands of workers of both skill groups almost equally, in contrast, an EITC-style tax scheme with negative marginal taxes is optimal, implementing an allocation between (S and U).[21]

In the latter case, the optimal upward distortion in labor supply cannot be understood in terms of the classical tradeoff between equity and intensive efficiency. Above U, moving down the Pareto frontier instead induces gains both in equity and intensive efficiency, which are counteracted by losses in extensive efficiency. Thus, the potential optimality of upward distortions at the intensive margin is not

[21] Proposition 3.6 however implies that for some regular economies, labor supply is undistorted in all utilitarian allocations. In these cases, the distance between points U and D is so large, that they enclose all points on the Pareto frontier. Put differently, the Pareto frontier for the non-relaxed problem coincides with the one for the relaxed problem.

driven, but rather reduced by local equity concerns, and can only be understood in terms of the efficiency-efficiency tradeoff studied in this section.

Along the same lines, it can be explained why low-skill labor y_1^S is downwards distorted in the surplus-maximizing allocation if and only if the revenue requirement A is large enough (above some threshold A_D), but still below the maximal tax revenue A_{max}. In this case, the surplus-maximizing allocation is located to the right of point D. The same is true for the complete set of utilitarian allocations, which implies that optimal marginal taxes are unambiguously positive in this case. One can conclude that negative participation taxes, which only arise for negative revenue requirements $\tilde{A}$, represent a necessary but not sufficient condition for the optimality of negative marginal taxes.

Summing up, I have shown that the problem of constrained surplus maximization gives rise to a tension between labor supply distortions at the intensive margin and labor supply distortions at the extensive margin, which has not been discussed in the literature so far. To minimize efficiency losses due to labor supply responses at the extensive margin, the social planner would prefer implementing an allocation that potentially violates upward incentive compatibility. Surplus-maximization then gives rise to a tradeoff between *intensive efficiency* and *extensive efficiency*, while welfare maximization involves a threeway-tradeoff between *equity*, *intensive efficiency* and *extensive efficiency*.[22]

3.7 One-dimensional private information

In the Mirrlees (1971) framework, agents differ in and are privately informed about their skill types only. In accordance with the recent literature on labor supply responses at the extensive margin, or at both margins, I have studied a model in which agents are heterogeneous with respect to skills and fixed costs of working (Saez, 2002; Choné and Laroque, 2011; Jacquet, Lehmann, and Linden, 2013). In the previous sections of this chapter as in all previous studies, it is moreover assumed that agents are privately informed about both dimensions of heterogeneity, so that the social planner can exclusively observe the gross income an agent earns on the labor market.

This gives rise to the question whether the derived results, in particular the potential optimality of the EITC, are driven by multi-dimensional heterogeneity or by multi-dimensional private information. From a theoretical perspective, this is important to understand the economic mechanism behind this result. From an applied perspective, one might argue that governments actually possess at least some information about these individual characteristics. Notice for example that

22

the US earned income tax credit (EITC) conditions tax liabilities on individual characteristics and life circumstances such as family status and the number of dependent children, which are commonly brought forward in the literature to motivate the assumption of heterogeneity in fixed costs of working. Similarly, tax authorities in many countries make use of tagging with respect to, e.g., disabilities or spatial distance between the place of residence and the workplace of tax payers.

This section aims at clarifying the importance of the imposed information structure. For this purpose, I study optimal taxation under the alternative assumptions that the social planner is able to observe one individual parameter directly.

3.7.1 Observable fixed costs

The first alternative to the information structure considered so far is to abandon the assumption of private information on fixed cost types. Instead, assume that the social planner is able to observe individual fixed cost types, while agents remain privately informed about their skill types. The information structure is thus similar to the one in the classical Mirrlees (1971) framework. In contrast to the latter, however, there is observable heterogeneity with respect to fixed costs, which can be used for tagging, i.e., to condition tax payments on individual fixed cost types. Nevertheless, changes in the tax schedule can give rise to labor supply responses at both margins.

With observable fixed costs, the planner only needs to take into account a limited set of incentive-compatibility constraints. In particular, only the incentive-compatibility constraints between agents with alternative skills, but the same cost parameter δ need to be satisfied:

$$c(\omega,\delta) - h\left(y(\omega,\delta),\omega\right) - 1_{y>0}\delta \geq c(\omega',\delta) - h\left(y(\omega',\delta),\omega'\right) - 1_{y>0}\delta \quad \text{for all } \omega,\omega' \in \Omega \text{ and } \delta \in \Delta \tag{3.27}$$

With observable fixed cost types, the social planner's problem is to maximize social welfare (3.3), subject to feasibility (3.1) and the reduced set of incentive compatibility constraints (3.27). However, this problem can be rewritten as a two-step problem. In the first step, the social planner maximizes overall welfare by redistributing resources between all fixed cost groups, without being constrained by any IC constraints. In the second step, the planner maximizes the group-specific welfare in each fixed cost group, subject to the group-specific IC constraints (3.27) and a group-specific feasibility constraint. Thus, he essentially solves separate optimal tax problems for each groups of agents with each fixed cost type $\delta \in \Delta$.

As Assumption *DUR* does not pin down the redistributive concerns of the social planner within the group of agents with fixed cost type δ, we need to replace it with the following assumption.

Desirability of Utilitarian Redistribution with Observable Costs (DUR δ): *For each fixed cost level $\delta \in \Delta$, the following is true in every implementable allocation (c, y):*

$$0 < {\alpha'}_{j+1}(c, y, \delta) < {\alpha'}_{j}(c, y, \delta) < {\alpha'}_{0}(c, y, \delta), \tag{3.28}$$

where ${\alpha'}_j(c, y, \delta) = U' \left[c(\omega_{j+1}, \delta) - h\left[y(\omega_{j+1}, \delta), \delta\right] - \delta\right]$ *and* ${\alpha'}_0(c, y, \delta) = \mathbb{E}_{\omega_k}\left[U'\left[c(\omega_k, \delta)\right] | y(\omega_k, \delta) = 0\right]$ *denote the endogenous weights associated to working agents with type* $\left(\omega_j, \delta\right)$ *and to unemployed agents, respectively.*

This assumption is clearly satisfied if function U is strictly concave on $\mathbb{R}$. Defining the set of utilitarian allocation based on Assumption *DUR* δ instead of *DUR*, the optimal structure of income tax schedule has similar effects on labor supply distortions as in the Mirrlees (1971) model.

Proposition 3.8. *With observable fixed cost types, labor supply in any utilitarian allocation is*

(i) undistorted at the intensive margin at the top skill, i.e., for all agents with skill type ω_n,

(ii) strictly downwards distorted at the intensive margin for all agents with lower skill types, and

(iii) weakly downwards distorted at the extensive margin for all types (ω, δ)

in all fixed cost groups for any regular economy.

Proposition 3.8 is closely related to the main results by Mirrlees (1971) and subsequent papers. In particular, parts (i) and (ii) correspond to the traditional results on optimal distortions at the intensive margin. These papers do not provide insights on optimal distortions at the extensive margin, though.[23] Nevertheless, similar arguments can be applied to show that all downward IC constraints in each fixed cost group must be binding in the optimal allocation. Distorting labor supply downwards at the intensive margin then helps to slacken these downward IC constraints, and to achieve further equity gains.

[23]Typically, Inada conditions ensure that all agents provide strictly positive output in the optimal allocation, thereby ruling out extensive margin distortions.

The crucial difference to the model with two-dimensional difference is directly related to the different information structure. With two-dimensional private information, there are agents in all skill groups that are indifferent between employment and unemployment. Thus, a small increase in the unemployment benefit induces unintended labor supply responses at the extensive margin in all skill groups. With observable fixed costs, only the least productive workers are indifferent between employment and unemployment, while all workers with higher skill types strictly prefer working. Consequently, a small increase in the benefit induces only extensive margin responses among the least productive workers, but does not drive high-skill workers out of the labor market.

3.7.2 Observable skill types

The second alternative to the information structure in the main part of this chapter involves observable skill types. In contrast, let the agents be privately informed about their fixed cost types. Thus, the social planner again faces a one-dimensional screening problem. Given this information structure, an allocation is incentive-compatible if and only if

$$c(\omega,\delta) - h\left(y(\omega,\delta),\omega\right) - 1_{y>0}\delta \geq (\omega,\delta') - h\left(y(\omega,\delta'),\omega\right) - 1_{y>0}\delta' \quad \text{for all } \omega \in \Omega \text{ and } \delta,\delta' \in \Delta \tag{3.29}$$

With observable fixed costs, the optimal tax problem is to maximize social welfare (3.3), subject to feasibility (3.1) and the new set of incentive compatibility constraints (3.29).

Again, Assumption *DUR* needs to be replaced with an assumption on the planner's redistributive concerns within the group of agents with each skill type ω_j.

Desirability of Utilitarian Redistribution with Observable Skills (DUR ω): *For every skill level $\omega_j \in \Omega$, the following is true in every implementable allocation (c, y)*

$$0 < \alpha_w^U(c, y, \omega_j) < \alpha_0^U(c, y, \omega_j), \tag{3.30}$$

where $\alpha_w^U(c,y,\omega_j) = \mathbb{E}_\delta\left[U'\left(c(\omega_j,\delta) - h\left[y(\omega_j,\delta),\omega_j\right] - \delta\right) | y(\omega_j,\delta) > 0\right]$ *and* $\alpha_0^U(c,y,\omega_j) = \mathbb{E}_\delta\left[U'\left[c(\omega_j,\delta]\right) | y(\omega_j,\delta) > 0\right]$ *denote the average weights associated to working agents and to unemployed agents, respectively.*

Again, this assumption is satisfied whenever function U is strictly concave on $\mathbb{R}$. The following proposition clarifies that optimal utilitarian income taxation cannot give rise to upward distortions in labor supply at any margin, as long as the social planner faces a one-dimensional screening problem.

Proposition 3.9. *With observable skill types, labor supply in any utilitarian allocation is*

- *undistorted at the intensive margin everywhere, and*
- *distorted downward at the extensive margin*

in all skill groups $\omega_j \in \Omega$ *for any regular economy.*

This insight and the logic behind it differ more strongly from the results by Mirrlees (1971) as well as Saez (2002) and Choné and Laroque (2011). Given this information structure, the social planner only needs to consider incentive compatibility constraints between agents with identical skill types, but different cost types. As there is no single-crossing condition imposed with respect to the fixed cost type δ, labor supply distortions at the intensive margin cannot help to slacken IC constraints and are thus never optimal. Because Assumption *AFC* imposes additive separability of the fixed cost component δ, every implementable allocation involves pooling by all workers with the same skill, and by all unemployed agents with the same skill. Thus, the social planner's problem is basically reduced to choosing a benefit level b_j for unemployed agents and a consumption-output bundle (c_j, y_j) for workers of each skill group $\omega_j \in \Omega$.

In any skill group, redistributing additional resources from workers to unemployed agents induces an equity gain, but also forces some previously indifferent workers out of the labor market. As long as labor supply is not downwards distorted at the extensive margin, this also implies an efficiency gain and, consequently, a strict increase in social welfare. Thus, the optimal allocation must involve a strict downward distortion at the extensive margin in each skill group.

To summarize, this section has clarified that neither two-dimensional heterogeneity nor the existence of labor supply responses at the intensive and the extensive margin per se alter the main insights of Mirrlees (1971). As long as the utilitarian planner is able to observe one dimension of heterogeneity, and needs to solve a one-dimensional screening problem, the optimal allocation will never involve upward distortions in labor supply. If agents are instead privately informed about skill types as well as fixed cost types, a utilitarian desire for redistribution does not pin down the optimal direction of labor supply distortions as implied by Propositions 3.2 to 3.4.

3.8 Discussion of assumptions

This chapter studies optimal utilitarian income taxation under a number of regularity assumptions imposed in section 3.3. In the following, I discuss the implications of these assumptions for the results of this chapter, in particular for the ambiguous sign of the optimal marginal income tax.

Assumption *AFC* and *QLC* restrict individual preferences. Assumption *AFC* follows Jacquet, Lehmann, and Linden (2013), the most prominent previous paper on optimal income taxation with labor supply responses at two margins. It imposes additive separability of the fixed cost component δ, which is required for reasons of tractability, as it allows to study the model using the random participation approach due to Rochet and Stole (2002). Under *AFC*, the fixed cost type δ only affects an agent's decision whether or not to enter the labor market. Conditional on entering the labor market, in contrast, the individually optimal level of workload y only depends on the skill type ω for any given tax schedule T. Thus, all workers with the same skill type react identically to changes in T. In mechanism perspective, assumption *AFC* implies that an allocation is implementable whenever it satisfies dimension-wise incentive compatibility, i.e., if no agent with type (ω, δ) prefers the allocations of types that differ in only one type parameter.

Assumption *QLC* follows the seminal paper by Diamond (1998) and has two implications. First, the imposed quasi-linearity in consumption considerably simplifies the optimal tax problem by eliminating income effects in labor supply. In particular, assumption *QLC* implies that individually optimal choices of workload y only depend on marginal income taxes, but are unaffected by lump-sum taxes. Thus, it simplifies the definition and analysis of labor supply distortions at the intensive margin.[24]

Second, the assumed quasi-linearity implies that the social planner's desire for redistribution only depends on the properties of the social objective function U (and the joint type distribution Ψ). This simplifies the analysis of sufficient conditions for condition *DUR* to be satisfied. In particular, the limit-case of a social planner without redistribute concerns is attained for U equaling the identity function.

By assumption *REM*, there would be unemployed as well as employed agents with each skill type under laissez-faire. This guarantees that variations in tax liabilities induce labor supply responses at the extensive margin in all skill groups, as long as the highest skill group faces a positive participation tax. By Proposition 3.1, this is always true for the optimal tax schedule. From a theoretical perspective,

[24] In a slightly weaker version of assumption *QLC*, income effects could also be ruled out by assuming that the utility function is given by $V(c, y, \omega, \delta) = \Phi\left[c - h(y, \omega) - 1_{y>0}\delta\right]$, where Φ is some strictly increasing function.

this assumption simplifies the comparison between the model studied here and the Mirrlees (1971) model, where no extensive margin responses occur.

The main results of this chapter survive, however, under the weaker condition that extensive margin responses occur for more than two skill groups. Consider an intermediate model in which extensive margin responses in labor supply only occur up to some threshold skill level $\omega_k < \omega_n$. Then, labor supply by all agents with skill types $\omega_j \in [\omega_k, \omega_{n-1}]$ is strictly downward distorted at the intensive margin in every utilitarian allocation, just as in the intensive model á la Mirrlees (1971). In contrast, the direction of optimal distortions at the intensive margin is ambiguous for all skill groups below ω_k, as in the model studied here.

Finally, assumption *LC* requires that the fixed cost distribution G_j for each skill group is strictly log-concave, i.e., has a strictly increasing reverse hazard rate, which is true for most commonly used distribution functions, including the uniform, normal, log-normal, Pareto and exponential distributions. Assumption *OFCD* imposes two conditions on the joint type distribution. By part (i), fixed costs must be larger among low-skill workers than among high-skill workers in the sense of first-order stochastic dominance. By part (ii), the hazard rate $G_j(\delta)/g_j(\delta)$ must be weakly lower for low-skill groups than for high-skill groups. Clearly, both properties are closely related, although not equivalent in general. However, they have two separate, crucial implications.

First, *LC* and Assumption *OFCD* (ii) jointly imply that the semi-elasticity of the participation rate is strictly lower for low-skill types than for high-skill types in every implementable allocation. As the analysis of the auxiliary problem in section 3.6 has revealed, this is a necessary condition for the ambiguous sign of the optimal marginal tax for the working poor, who receive employment subsidies. Crucially, strong empirical evidence confirms that low-skill workers indeed react more responsively on the extensive margin (see, e.g., Juhn et al. 1991; Immervoll et al. 2007; Meghir and Phillips 2010).[25] Thus, assumptions *LC* and *OFCD* guarantee the empirical relevance of the derived results.[26]

Second, *LC* and *OFCD* (i) jointly ensure that condition *DUR* does not restrict

[25]More precisely, these studies find that the elasticity of participation, $[y_j - T(y_j)]\,\frac{g_j[y_j - T(y_j)]}{g_j[y_j - T(y_j)]}$ is decreasing along the skill dimension. The same must be true for the semi-elasticities of participation, however, as all estimated elasticities are positive and $y_j - T(y_j)$ is strictly increasing in the skill type. Note, however, that these empirical studies only reveal relative semi-elasticities under the current tax schedules, which will typically differ from the optimal tax schedule.

[26]It is nevertheless interesting to note that the sign of the optimal marginal tax would even be ambiguous in the opposite case, in which high-skill groups would react more strongly at the extensive margin. Then, however, optimal marginal taxes would be strictly positive for low-skill workers and potentially negative for high-skill workers. I am not aware of real-world tax schedules with this property, however.

the analysis to the empty set. Although the planner's desire for redistribution from high-skill to low-skill workers is imposed directly through *DUR*, it actually represents a joint assumption on properties of the joint type distribution and the social objective U. It can be shown that the sequence of social weights is strictly decreasing whenever the social objective U is strictly concave and *LC* and *OFCD* (ii) hold. In contrast, concavity of U would neither be sufficient nor necessary if *LC* or *OFCD* (ii) would be violated.[27]

Intuitively, concavity of U implies that the planner prefers to redistribute from skill groups with high average utility to skill groups with lower average utility. With a strong positive correlation between skills and fixed costs, however, high-skill workers might be on average worse off than low-skill workers. Thus, the social planner might hold an anti-utilitarian desire to redistribute from low-skill to high-skill workers. More generally, there might exist joint type distributions Ψ such that Assumption *DUR* would not be satisfied for *any* strictly increasing function U.[28]

3.9 Related Literature

The chapter studies the implications of optimal utilitarian income taxation in a model with labor supply responses at two margins. Thus, it builds on the rich literature on optimal taxation with labor supply responses at the intensive margin only, starting with the seminal paper by Mirrlees (1971). Further important studies include Seade (1977) and Seade (1982) and Hellwig (2007). In their models, a utilitarian desire for redistribution leads to the optimality of strictly positive marginal taxes everywhere below the very top. In contrast, the optimal sign of marginal taxes is ambiguous in my chapter, which is a joint result of, first, the existence of two margins of labor supply responses, and second, individual heterogeneity in two dimensions that are both associated with private information.

Regarding the theoretical model, this chapter is more closely related to the literature on optimal taxation with labor supply responses at the extensive margin. This strand of the literature was initiated by Saez (2002), building on previous

[27] See Propositions 2 and 3 by Choné and Laroque (2011) for the same result in a model with labor supply responses at the extensive margin only. In the Mirrlees (1971) framework with one-dimensional heterogeneity, in contrast, concavity of U is a sufficient condition for a standard utilitarian desire for redistribution, irrespective of the properties of the type distribution.

[28] Choné and Laroque (2010) study the roots and effects of increasing social weight functions in a model with labor supply responses at the intensive margin only, also referring to settings with two-dimensional heterogeneity and strong correlation between both private parameters. In particular, they use this logic to rationalize an EITC-style income tax schedule with negative marginal taxes.

work by Diamond (1980). A rigorous theoretical treatment of the extensive model is provided by Choné and Laroque (2011). In these papers, agents differ in two individual parameter, interpreted as skills and fixed costs or opportunity costs of employment. Thus, the social planner faces a multi-dimensional screening problem. In contrast to this chapter, however, they focus on models in which the agents only face fixed costs of working, but no continuous cost of increasing their workload as in Mirrlees (1971).[29] Thus, agents only choose whether or not to work at all; if an agent enters the labor market, he always produces at full capacity. Consequently, distortions in labor supply can only occur at the extensive margin.

The main finding of these models is that negative participation taxes for low-skill workers are optimal if and only if the utilitarian planner associates to them a social weight above the population average. The intuition for this result rests on an efficiency argument, comparing the efficiency costs of two changes in the allocation: redistributing resources towards the working poor induces some upwards distortions in the labor supply of these groups, but redistributing resources towards the unemployed leads to adverse labor supply responses by workers of *all* skill groups.[30]

In the extensive models by Diamond (1980), Saez (2002), and Choné and Laroque (2011), the economic role of marginal income taxes differs strongly from the one in the Mirrlees (1971) model and in my model. First, non-zero marginal taxes do not induce labor supply distortions at the intensive margin. Second, labor supply distortions do not help to relax incentive compatibility contraints. In their models, there are no upward incentive compatibility constraints, and only degenerate downward incentive compatibility constraints.[31] In my model, negative (or positive) marginal taxes can in contrast only be optimal because they induce intensive margin distortions that help to relax incentive compatibility constraints. While Diamond (1980) and Choné and Laroque (2011) also provide examples under which negative marginal taxes for the working poor are optimal, the mathematical and theoretical arguments explaining these phenomena consequently differ from those provided above.[32]

[29]While focusing on models with one margin only, Saez (2002) also discusses the general model with labor supply responses at both margins. He simulates the optimal tax schedule for this general model, but does not study the properties of optimal tax schedule analytically.

[30]Christiansen (2012) studies in detail the economic mechanism giving rise to the optimality of negative participation taxes.

[31]In the paper by Choné and Laroque (2011), the optimal allocation always involves slack downward incentive compatibility constraints for all skill levels with relevant extensive margin, i.e., with positive shares of unemployed agents.

[32]The example provided in Choné and Laroque (2011) is based on the assumptions that first, the skill space is continuous and second, no agent of the lowest skill type would work under laissez-faire. Both assumptions do not hold in my chapter.

More generally, the social planner in my model needs to take into account labor supply responses at the intensive and the extensive margin responses. As shown above, the maximization of a utilitarian welfare function can give rise to a tradeoff between intensive efficiency and extensive efficiency, which is key to understand the ambiguous sign of optimal marginal taxes. This tradeoff is absent in the extensive models as well as the intensive models discussed above.

Most closely related to this chapter is the analysis by Jacquet, Lehmann, and Linden (2013), who also study optimal income taxation with labor supply responses at both margins. As in my model, agents face fixed costs of employment (as in the extensive model) as well as variable costs of providing effort in the job (as in the intensive model). The research questions of both papers differ strongly. This chapter contributes to the literature by showing that, and why, the optimal signs of marginal income taxes and participation taxes are ambiguous even if the social planner has a desire for utilitarian redistribution. Jacquet, Lehmann, and Linden (2013) focus on identifying conditions under which optimal marginal taxes are unambiguously positive. In particular, they provide a sufficient condition under which marginal taxes are throughout positive, expressed in terms of endogenous social weights and of the optimal allocation itself. They argue that this sufficient condition does not seem very restrictive, and provide some examples under which is is certainly satisfied. In contrast, I show that the optimal sign of marginal income taxes and participation taxes is in general ambiguous, and provide a sufficient condition for the optimality of negative marginal taxes, which is expressed in terms of the primitives, i.e., the type set, the type distribution and utility functions. One interpretation of my results is that, for a large class of economies, it mainly depends on the intensity of the planner's redistributive concerns whether or not the condition identified by Jacquet, Lehmann, and Linden (2013) is satisfied. As Jacquet, Lehmann, and Linden (2013) concentrate on cases in which the optimal marginal tax can be signed unambiguously, they are not concerned with working out the economic mechanism underlying the indeterminacy of this sign. Correspondingly, they do not discuss the tradeoff between intensive efficiency and extensive efficiency, which is identified as the source of ambiguity in my model.

There are two minor differences between this chapter and Jacquet, Lehmann, and Linden (2013). First, their model is more general as they allow for income effects in labor supply which are assumed away in this chapter. Second, the skill space in their model is given by an interval, while I study a finite set of skill types. Reflecting this difference, the mathematical proofs applied in both papers differ considerably.[33]

[33]In a related paper, Lorenz and Sachs (2011) study optimal income taxation with two margins

Two further papers aim at rationalizing negative marginal income taxes, both based on a desire to redistribute resources locally upwards. Choné and Laroque (2010) study a model á la Mirrlees (1971) with labor supply reponses at the intensive margin only, but with two-dimensional heterogeneity in individual characteristics. They argue that, if there is a specific correlation between both dimensions of heterogeneity, the social planner might want to redistribute resources locally upwards to the group of more skilled, but more disadvantaged (in the second dimension) agents. In this case, the anti-utilitarian desire to redistribute resources from low-skill to high-skill agents gives rise to a reversed equity-efficiency tradeoff, and to optimal upward distortions in labor supply by high-skill workers. In contrast, I assume the social planner to be a utilitarian who would strictly prefer to transfer resources from high-skill to low-skill workers, if he could ignore incentive considerations. In my framework, optimal upward distortions can thus result for efficiency reasons only, more precisely due to the tradeoff between intensive efficiency and extensive efficiency.

In the model by Beaudry, Blackorby, and Szalay (2009), agents differ in and are privately informed about their productivities in the formal sector as well as in the informal sector. Within each group of workers with identical productivity in the formal sector, the ones with highest informal productivity choose to stay officially unemployed in order to maximize their income. Thus, the social planner assigns lower social weights to the unemployed than to the employed within the same skill group, which again conflicts with the assumed desire for utilitarian redistribution in this chapter. In Beaudry, Blackorby, and Szalay (2009), the non-monotonic weight sequence implies that employment subsidies up to some threshold skill level are optimal. Their model differs in two further aspects from the classical Mirrlees (1971) setting. First, effort costs are linear so that all agents choose either to work at full capacity in the formal sector or to move towards the informal sector (except one threshold skill type). As in the extensive models discussed above, the optimal allocation cannot involve upward distortions at the intensive margin by construction. Second, they assume that the social planner can observe hours worked in the formal sector, thereby deviating from the conventional information structure.

of labor supply responses, where the extensive margin results from an minimum-hours constraint. While they do not study the sign of optimal marginal taxes, they provide a sufficient condition for the positivity of optimal participation taxes.

3.10 Conclusion

The largest US program transferring resources towards the poor, the Earned Income Tax Credit (EITC), involves negative marginal taxes and negative participation taxes for the working poor. Given a utilitarian desire for redistribution, this cannot be rationalized in a model in which agents adjust their labor supply only at the intensive margin as in the classical Mirrlees (1971) framework; the optimal *marginal tax* is then positive everywhere below the very top. In contrast, recent research finds that optimal *participation taxes* can be negative if agents adjust their labor supply at the extensive instead of the intensive margin (Saez, 2002; Choné and Laroque, 2011). This chapter is the first paper to show that, and explain why, EITC-style tax schemes with *negative marginal taxes and negative participation taxes* can indeed be optimal if labor supply responses take place at the intensive and the extensive margin, which is arguably the most appropriate assumption from an empirical perspective.

More generally, I show that the existence of a utilitarian desire to redistribute resources from high-skill to low-skill workers does neither pin down the optimal signs of marginal and participation taxes nor the optimal directions of labor supply distortions at both margins. Instead, the properties of the optimal tax scheme depend on the intensities of the social planner's concerns for redistribution, first, from the very rich to the very poor, and second, within the group of the working poor. The chapter works out the economic intuition behind this ambiguity, which is driven by an inherent, but yet undiscussed, tradeoff between intensive efficiency and extensive efficiency aspects. Negative marginal taxes create efficiency losses at the intensive margin; in certain situation, they can however help to increase extensive efficiency by slackening upward incentive compatibility constraints.

A number of questions remain unresolved. First, the theoretical analysis clarifies that the properties of the optimal tax scheme depend strongly on the relative (semi-)elasticities of labor market participation shares in different skill groups. While there is already some empirical evidence on this issue, future research should focus more strongly on the heterogeneity of labor supply responses, instead of mainly estimating average elasticities. Second, the analysis has been simplified considerably by a number of assumptions. In my view, the most restrictive of these assumptions are given by the quasi-linearity of preferences in consumption, which rules out any income effects in labor supply, and the discreteness of the skill type space. Although I conjecture that the basic insights would remain valid, relaxing these assumptions could improve the economic understanding of the mechanisms at work and complete the picture.

Appendix 3.A Proofs for Sections 3.4 to 3.6

Proof of Lemma 3.1 An allocation (c, y) is incentive compatible if it satisfies the following inequality for all pairs of (ω, δ) and (ω', δ') in $\Omega \times \Delta$:

$$c(\omega, \delta) - h\left[y(\omega, \delta), \omega\right] - 1_{y(\omega,\delta)>0}\delta \geq c(\omega', \delta') - h\left[y(\omega', \delta'), \omega\right] - 1_{y(\omega',\delta')>0}\delta$$

The proof of Lemma 3.1 requires to distinguish between several cases. First, consider two agents of types (ω, δ) and (ω', δ') such that both provide zero output. Incentive compatibility requires identical consumption $c(\omega, \delta) = c(\omega', \delta') \equiv b$.

Second, consider two agents with identical skill type ω_j and different cost types $\delta \neq \delta'$ such that both provide positive effort. As both IC constraints need to be satisfied, both agents need to receive the same gross (of fixed costs) utility level $c(\omega_j, \delta) - h\left[y(\omega_j, \delta), \omega_j\right] = c(\omega_j, \delta') - h\left[y(\omega_j, \delta'), \omega_j\right] = z_j$. In general, incentive compatibility does not imply $c(\omega_j, \delta) = c(\omega_j, \delta')$ and $y(\omega_j, \delta) = y(\omega_j, \delta')$, because different consumption bundles provide the same gross utility level z_j to workers with identical skill types. Incentive compatibility only requires that non of the bundles meant for some worker with skill ω_j is preferred by some worker with skill ω_k, i.e., that $c(\omega_j, \delta) - h\left[y(\omega_j, \delta), \omega_k\right] \leq z_k$ holds.

Second-best Pareto efficiency, however, requires identical bundles (c_j, y_j) for all workers of skill level ω_j. By the properties of effort cost function h, there is always a unique bundle (c_j, y_j) that minimizes the net transfer $c - y$ subject to incentive compatibility, i.e., subject to $c - h(y, \omega_j) = z_j$ and $c - h(y, \omega_k) < z_k$. This may either involve the efficient level $\hat{y}_j$ or the closest level to $\hat{y}_j$ that is still consistent with all IC constraints. If some agent with type (ω_j, δ) receives bundle $(c', y') \neq (c_j, y_j)$ with positive output $y' \neq y_j$, then net resources can be saved by changing his allocation to (c_j, y_j) without changing his utility level. But then, redistributing these resources lump-sum to all agents in the economy leads to a strict (incentive-compatible) Pareto improvement.

Finally, consider two agents with the same skill type ω_j and different cost types δ, δ' such that $y(\omega_j, \delta) > 0$ and $y(\omega_j, \delta') = 0$. By incentive compatibility, $\delta \leq c_j - h\left(y_j, \omega_j\right) - b \equiv \hat{\delta}_j$ and $\delta' \geq \hat{\delta}_j$.

Proof of Lemma 3.2 Assumption *DUR* ensures that the social planner associates positive weight to all skill groups. By standard arguments, any utilitarian allocation must then be Pareto-efficient, which implies identical bundles (c_j, y_j) for all workers of each skill group ω_j, and all unemployed agents. The welfare function (3.12) and the feasibility constraint (3.13) directly follow from inserting the skill-conditional levels c_j and y_j and the universal benefit b.

Incentive compatibility along the fixed cost dimension, i.e., between types with identical skills ω_j and different cost types δ, δ' is given if and only if the participation constraint (3.14) is satisfied. It takes into account the possibility of corner solutions, in which all agents of some skill groups are either unemployed, $\hat{\delta}_j = \bar{\delta}$, or employed, $\hat{\delta}_j = \underline{\delta}$. As all unemployed agents receive the same benefit, constraint (3.14) also ensures that no worker of skill group ω_j wants to mimic an unemployed agent of some other skill group.

Incentive compatibility between two workers with adjacent skill types ω_j, ω_{j+1} and arbitrary fixed cost types $\delta \leq \hat{\delta}_j$, $\delta' \leq \hat{\delta}_{j+1}$ is satisfied if and only if

$$\tilde{V}(c_{j+1}, y_{j+1}, \omega_{j+1}) - \delta' \geq \tilde{V}(c_j, y_j, \omega_{j+1}) - \delta' \text{ and}$$
$$\tilde{V}(c_j, y_j, \omega_j) - \delta \geq \tilde{V}(c_{j+1}, y_{+1}, \omega_j) - \delta \, ,$$

which is equivalent to constraints (3.15) and (3.16). By the single-crossing property, they also ensures incentive compatibility between non-adjacent skill types. Finally, (3.14), (3.15) and (3.16) jointly guarantee that no unemployed agent of skill type ω_j wants to mimic some worker with some other skill type ω_k, because $b > \tilde{V}(c_j, y_j, \omega_j) - \delta \geq \tilde{V}(c_k, y_k, \omega_j) - \delta$ for all unemployed agents with $\delta > \hat{\delta}_j$ and any $k \neq j$.

Proof of Lemma 3.3

Proof. For any type (ω_j, δ), efficient labor supply is given by the minimizer of the net transfer of resources $(c - y)$ subject to the constraint $V(c, y, \omega_j, \delta) \geq v$. This problem is equivalent to maximizing the following Lagrangian

$$\mathcal{L}(c, y) = y - c + \lambda \left[c - h(y, \omega_j) - 1_{j>0}\delta - v \right]$$

The discontinuity at $y = 0$ requires a case distinction. For the corner solution $y = 0$, the required net transfer trivially follows as $c_0(v) = v$.

For the interior solution $y > 0$, monotonicity and convexity of h ensure a unique solution, given by $y = \hat{y}_j$ and $c = \hat{c}_j(v) = h\left(\hat{y}_j, \omega_j\right) + \delta + v$, where $\hat{y}_j$ is implicitly defined by the first-order condition $1 - h_1\left(\hat{y}_j, \omega_j\right) = 0$. The net transfer is given by $\hat{c}_j(v) - \hat{y}_j = v - \hat{s}_j + \delta$. If and only if $\delta \leq \hat{s}_j$, the interior solution dominates the corner solution, so that $y^*(\omega_j, \delta) = \hat{y}_j$. □

Proofs of Propositions 3.1-3.2

Proposition 3.1 implies that the famous no-distortion-at-the-top result still holds with labor supply responses at the intensive margin, although only with respect to the intensive margin. At the extensive margin, labor supply is instead strictly downward distorted at the top. Proposition 3.2 derives the existence of utilitarian allocations without distortions at the intensive margin for any regular economy.

Both propositions are proven through a series of lemmas. To simplify notation in the following, I find it convenient to define the employment rent $r_j = c_j - h\left(y_j, \omega_j\right) - b$ as an auxiliary function. It measures the utility gain that a worker of skill level ω_j receives if he provides output $y_j > 0$ instead of staying unemployed, conditional on the mechanism (c, y, b) and gross of fixed costs.

First, consider a relaxed problem in which the incentive compatibility constraints (3.15 and (3.16) between active workers of different skill types are not taken into account. However, we still include the constraint that unemployed agents of all skill types must receive the same benefit b. Moreover, the planner is still restricted by the set of participation constraints (3.14), i.e., needs to take into account labor supply responses at the extensive margin. Note that this relaxed problem corresponds to the first-and-half problem in Jacquet, Lehmann, and Linden (2013), which is however studied under the assumption of a continuous skill space. Given the definition of the employment rent, the social planner's problem can be defined as follows

Relaxed Problem. Maximize over $y = \left(y_j\right)_{j=1}^n$, $r = \left(r_j\right)_{j=1}^n$, and b the welfare function

$$\sum_{j=1}^{n} f_j \left[\int_{\underline{\delta}}^{\hat{\delta}_j} U\left(r_j + b - \delta\right) dG_j(\delta) + \left[1 - G_j\left(\hat{\delta}_j\right)\right] U(b) \right]$$

subject to the constraints

$$b = \sum_{j=1}^{n} f_j G_j\left(\hat{\delta}_j\right) \left[s\left(y_j, \omega_j\right) - r_j\right],$$

$$\hat{\delta}_j = \max\left\{\underline{\delta}, \min\left\{r_j, \bar{\delta}\right\}\right\} \text{ for all } \omega_j \in \Omega$$

In this model, the planner's objective is not necessarily globally concave in all choice variables. The same problem arises in the model with labor supply responses at the extensive margin only, Choné and Laroque (2011) show that the Lagrangian can become convex in consumption levels if social weights are particularly high. The following assumption assumes away this irregularity in order to concentrate on the economic problem.

Assumption 3.4. *For any skill level ω_j, the social weight $\alpha_j^U(c, y)$ associated to workers with this skill type satisfies*

$$\alpha_j^U(c, y) < \chi_j(\delta) = \left(2 - \frac{G_j(\delta)g_j'(\delta)}{g_j(\delta)^2}\right) / \left(1 - \frac{G_j(\delta)g_j'(\delta)}{g_j(\delta)^2}\right)$$

for all $\delta \in \Delta$. Moreover, $\alpha_j^U(c, y)$ is weakly decreasing in c_j.

The log-concavity of G_j imposed by assumption *LC* ensures that $g_j(\delta)^2 > G_j(\delta)g_j'(\delta)$. Thus, the upper bound $\chi_j(\delta)$ exceeds unity for $\omega_j \in \Omega$ and $\delta \in \Delta$. For uniformly distributed fixed costs, $\chi_j(\delta) = 2$ for all δ and ω_j. All results derived in this chapter follow for utilitarian welfare functions that satisfy this assumption.

Lemma 3.7. *Let the relative social weight $\alpha_j^U(c^R, y^R)$ be defined as in equation 3.8 on page 98. In the solution to the relaxed problem, (r^R, y^R, b^R), all workers of skill type ω_j*

- *provide the efficient output level $y_j^R = \hat{y}_j$, and*
- *receive an employment rent that is implicitly defined by*

$$g(r_j^R)\left[r_j^R - \hat{s}_j\right] = G_j(r_j^R)\left(\alpha_j^U(c^R, y^R) - 1\right) .$$

The unemployment benefit is given by

$$b^R = \sum_{j=1}^{n} f_j G_j\left(r_j^R\right)\left(\hat{s}_j - r_j^R\right)$$

Proof. Assume that $r_j \in \left(\underline{\delta}, \bar{\delta}\right)$ for all skill levels $\omega_j \in \Omega$. Then, the Lagrangian of the relaxed problem is given by

$$\begin{aligned}\mathcal{L} = &\sum_{j=1}^{n} f_j \left[\int_{\underline{\delta}}^{r_j} g_j(\delta) U\left(r_j + b - \delta\right) d\delta + \left(1 - G_j(r_j)\right) U(b)\right] \\ &+ \lambda \left[\sum_{j=1}^{n} f_j G_j(r_j)\left(s\left(y_j, \omega_j\right) - r_j\right) - b\right]\end{aligned}$$

The first-order conditions with respect to r_j, y_j and b are given by

$$\mathcal{L}_{r_j} = f_j \left[\int_{\underline{\delta}}^{r_j^R} g_j(\delta) U'\left(r_j^R + b^R - \delta\right) d\delta - \lambda G_j(r_j^R) \right.$$
$$\left. + \lambda g_j(r_j^R) \left(s\left(y_j^R, \omega_j\right) - r_j^R \right) \right] = 0$$
$$\mathcal{L}_{y_j} = \lambda f_j G_j(r_j^R) s_1(y_j^R, \omega_j) = 0$$
$$\mathcal{L}_b = \underbrace{\sum_{j=1}^{n} f_j \left[\int_{\underline{\delta}}^{r_j^R} g_j(\delta) U'\left(r_j^R + b^R - \delta\right) d\delta + \left(1 - G_j(r_j^R)\right) U'(b^R) \right]}_{=\bar{\alpha}(c^R, y^R)} - \lambda = 0$$

By the last FOC, the value of multiplier λ equals the average marginal utility $\bar{\alpha}(c^R, y^R)$ in the optimal allocation. The same will be true for the full problem. By the FOC with respect to y_j, $s_1(y_j^R, \omega_j)$ must be zero in the solution to the relaxed problem. Thus, workers of all skill levels provide efficient output $y_j = \hat{y}_j$. Rearranging the FOC with respect to r_j and substituting in $\alpha_j^U(c^R, y^R) = \left[\int_{\underline{\delta}}^{r_j^R} g_j(\delta) U'\left(r_j^R + b^R - \delta\right) d\delta\right] / \bar{\alpha}(c^R, y^R)$ gives the expression in Lemma 3.7. By assumption *REM*, the first derivative is strictly positive for $r_j \to \underline{\delta}$. For $r_j \to \infty$, it is strictly negative by $\lim_{z\to\infty} U'(z) < 1$. By the continuity of the first-order condition in r_j, it must have at least one root. Assumption 3.4 guarantees concavity of the Lagrangian in r_j is for all $r_j \geq \underline{\delta}$. Thus, the first-order condition with respect to r_j has a unique root, which involves $r_j^R > \underline{\delta}$. □

The conditions defining the relaxed problem's solution have the same structure as those defining the optimal allocations in the extensive models by Saez (2002) and Choné and Laroque (2011), and the solution to the first-and-half problem in Jacquet, Lehmann, and Linden (2013). Due to the lack of IC constraints, labor supply is generally undistorted in the solution to the relaxed problem. The optimal vector of employment rents is determined by the sequence of endogenous social weights α^U. For $\alpha_j > 1$, workers of skill type ω_j receive an employment rent that exceeds the efficient surplus $\hat{s}_j = \max_y \left[y - h(y, \omega_j)\right]$. For $\alpha_j < 1$, workers of skill type ω_j receive an employment rent below $\hat{s}_j$. By assumption *REM*, this implies an interior solution $r_j < \hat{s}_j \leq \hat{s}_n < \bar{\delta}$. Note that $\alpha_n < 1$ is ensures for all utilitarian allocations.

Next, we identify conditions on the pair of social weights α_j and α_{j+1} such that the solution to the relaxed problem satisfies both IC constraints. For this purpose, I ignore the endogeneity of the weight sequence α^U for a while. In particular, assume that α^U equals some exogenous sequence $\beta = (\beta_0, \beta_1, \ldots, \beta_n)$, which determines the optimal employment rent $\tilde{r}_j(\beta_j)$ in the relaxed problem.

Lemma 3.8. *For any skill level ω_j, there is a threshold $\gamma_j^E > 1$ such that the function $Z_j(\beta_j, r) = g(r)\left[r - \hat{s}_j\right] - G_j(r)\left(\beta_j - 1\right)$ has a unique root $\tilde{r}_j\left(\beta_j\right) \in \left(\underline{\delta}, \bar{\delta}\right)$ in r if and only if $\beta_j < \gamma_j^E$. Moreover, $\tilde{r}_j\left(\beta_j\right)$ is strictly increasing in its argument for all $\beta_j < \gamma_j^E$.*

Proof. First, note that $\lim_{r\to\underline{\delta}} Z_j(r) < 0$ for all β_j. Second, for $r \in \left(\underline{\delta}, \bar{\delta}\right)$, the derivative of Z_j with respect to r is given by

$$\frac{\partial Z_j(r, \beta_j)}{\partial r} = g_j(r)\left(2 - \beta_j\right) + g'(r_j)\left(r - \hat{s}_j\right)$$

Assumption 3.4 ensures that this derivative is strictly positive at any root of Z in r. By continuity, there is consequently at most one root in the interval $\left(\underline{\delta}, \bar{\delta}\right)$. For $r \to \bar{\delta}$, Z approaches $1 + g_j\left(\bar{\delta}\right)\left(\bar{\delta} - \hat{s}_j\right) - \beta_j$. Thus, a unique root in r exists if β_j is smaller than the minimum of $1 + g_j\left(\bar{\delta}\right)\left(\bar{\delta} - \hat{s}_j\right) > 1$ and $\chi_j\left(\bar{\delta}\right) > 1$ as defined in Assumption 3.4. The derivative of $\tilde{r}_j$ with respect to β_j is given by

$$\frac{d\tilde{r}_j}{d\beta_j} = \frac{G_j\left(\tilde{r}_j\right)}{\left(\partial Z_j(\tilde{r}_j, \beta_j)\right)/(\partial r)} > 0.$$

The numerator is positive for all $\beta_j < \gamma_j^E$, where $\tilde{r}_j < \bar{\delta}$. As argued above, Assumption 3.4 ensures the same for the denominator. □

Lemma 3.9. *Consider any two adjacent skill groups ω_j and ω_{j+1} in Ω with weights β_j and β_{j+1}. There are a value $\gamma_j^D \in \left(1, \gamma_j^E\right)$ and a strictly increasing function $\beta_j^D : \left(-\infty, \gamma_j^D\right) \to \left(-\infty, \gamma_{j+1}^E\right)$ such that the solution to the relaxed problem satisfies the downward IC constraint if and only if $\beta_j < \gamma_j^D$ and $\beta_{j+1} \in \left[\beta_j^D\left(\beta_j\right), \gamma_{j+1}^E\right)$. There is a threshold level $\underline{\beta}_j < 1$ such that $\beta_j^D(x) < x$ for all $x \in \left(\underline{\beta}_j, \gamma_j^D\right)$.*

Proof. Using the definition of the employment rent, the downward IC constraint (3.15) reads $\tilde{r}_{j+1}\left(\beta_{j+1}\right) - \tilde{r}_j\left(\beta_j\right) \geq h(\hat{y}_j, \omega_j) - h(\hat{y}_j, \omega_{j+1})$. Note that the right-hand side does not depend on the weights β_j, β_{j+1}. By Lemma 3.8, $\tilde{r}_j$ and $\tilde{r}_{j+1}$ are defined and below $\bar{\delta}$ if and only if $\beta_j < \gamma_j^E$ and $\beta_{j+1} < \gamma_{j+1}^E$.

First, define γ_j^D implicitly by $\tilde{r}_j\left(\gamma_j^D\right) = \bar{\delta} - h(\hat{y}_j, \omega_j) + h(\hat{y}_j, \omega_{j+1}) < \bar{\delta}$. We have $\gamma_j^D > 1$ due to $\tilde{r}_j(1) < \hat{s}_{j+1} - h(\hat{y}_j, \omega_j) + h(\hat{y}_j, \omega_{j+1})$ and $\hat{s}_{j+1} < \bar{\delta}$. By the monotonicity of $\tilde{r}_k$ in β_k for $k \in \{j, j+1\}$, the downward IC can only be satisfied for $\beta_{j+1} \to \gamma_{j+1}^E$ if $\beta_j < \gamma_j^D$. For any $\beta_j < \gamma_j^D$, there is moreover a unique level $\beta_j^D(\beta_j) < \gamma_{j+1}^E$ such that $\tilde{r}_{j+1}(x) - \tilde{r}_j\left(\beta_j\right) \geq h(\hat{y}_j, \omega_j) - h(\hat{y}_j, \omega_{j+1})$ is satisfied with equality if and only if $x = \beta_j^D(\beta_j)$, and with strict inequality if and only if $x \in \left[\beta_j^D(\beta_j), \gamma_{j+1}^E\right)$. Moreover, β_j^D is strictly increasing in β_j due to the monotonicity of $\tilde{r}_k$ in β_k.

For the threshold $\underline{\beta}_j$, consider the case where $\beta_j = \beta_{j+1} = x$. As the derivative of $\tilde{r}_j(x)$ with respect to ω_j is strictly positive, we have $\tilde{r}_{j+1}(x) > \tilde{r}_j(x)$. Combining both first-order conditions, we have

$$\tilde{r}_{j+1}(x) - \tilde{r}_j(x) = \hat{s}_{j+1} - \hat{s}_j + (x-1)\left[\frac{G_{j+1}(\tilde{r}_{j+1}(x))}{g_{j+1}(\tilde{r}_{j+1}(x))} - \frac{G_j(\tilde{r}_j(x))}{g_j(\tilde{r}_j(x))}\right]$$

By assumptions *LC* and *OFCD*, the inequality $\tilde{r}_{j+1} > \tilde{r}_j$ implies that the last term in brackets is strictly positive. Moreover, note that

$$\hat{s}_{j+1} - \hat{s}_j > s\left(\hat{y}_j, \omega_{j+1}\right) - s\left(\hat{y}_j, \omega_j\right) = h(\hat{y}_j, \omega_j) - h(\hat{y}_j, \omega_{j+1}).$$

Thus, the downward IC constraint is satisfied with strict inequality for all $x \geq 1$. On the other hand, $\lim_{x\to-\infty} \tilde{r}_{j+1}(x) = \lim_{x\to-\infty} \tilde{r}_j(x) = \underline{\delta}$. Thus, the downward IC constraint is violated for sufficiently small $x \to -\infty$. By continuity, there must be at least one value $x < 1$ such that the downward IC is satisfied with equality for $\beta_{j+1} = \beta_j = x$. The threshold $\underline{\beta}_j$ is given by the highest value with this property. Thus, the downward IC constraint is satisfied with strict inequality for all $\beta_j = \beta_{j+1} > \underline{\beta}_j$. By the continuity of $\tilde{r}_{j+1}$ in β_{j+1}, $\beta_j^D(\beta_j) < \beta_j$ must be satisfied for all $\beta_j \in \left(\underline{\beta}_j, \gamma_j^D\right)$. □

Lemma 3.10. *Consider any two adjacent skill groups ω_j and ω_{j+1} in Ω with weights β_j and β_{j+1}. There are a value $\gamma_j^U < \gamma_j^E$ and a strictly increasing function $\beta_j^U : \left(-\infty, \gamma_j^U\right) \to \left(-\infty, \gamma_{j+1}^E\right)$ such that the solution to the relaxed problem violates the upward incentive compatibility constraint between groups ω_j and ω_{j+1} if and only if $\beta_j < \gamma_j^U$ and $\beta_{j+1} \in \left[\beta_j^U\left(\beta_j\right), \gamma_{j+1}^E\right)$. There is a threshold level $\bar{\beta}_j \in \left(1, \gamma_j^D\right]$ such that $\beta_j^U(x) > x$ for all $x < \bar{\beta}_j$.*

Proof. The upward IC constraint (3.16) can be rewritten $\tilde{r}_{j+1}\left(\beta_{j+1}\right) - \tilde{r}_j\left(\beta_j\right) \leq h(\hat{y}_{j+1}, \omega_j) - h(\hat{y}_{j+1}, \omega_{j+1})$. By the monotonicity of $\tilde{r}_j$ in β_j, there is a unique $\gamma_j^U < \gamma_j^E$ such that $\tilde{r}_j\left(\gamma_j^D\right) = \bar{\delta} - h(\hat{y}_{j+1}, \omega_j) + h(\hat{y}_{j+1}, \omega_{j+1}) < \bar{\delta}$. The upward IC is satisfied for all $\beta_j \geq \gamma_j^U$ and $\beta_{j+1} < \gamma_{j+1}^E$. For all $\beta_j < \gamma_j^U$, there is in contrast a unique level β_j^U such that $\tilde{r}_{j+1}(x) - \tilde{r}_j\left(\beta_j\right) \leq h(\hat{y}_j, \omega_j) - h(\hat{y}_j, \omega_{j+1})$ is satisfied with equality if and only if $x = \beta_j^U(\beta_j)$, and violated if and only if $x \in \left[\beta_j^U(\beta_j), \gamma_{j+1}^E\right)$. Moreover, β_j^U is strictly increasing in β_j due to the monotonicity of $\tilde{r}_j$ and $\tilde{r}_{j+1}$.

For the threshold $\bar{\beta}_j$, consider the case where $\beta_j = \beta_{j+1} = x$. Combining both first-order conditions, we have

$$\tilde{r}_{j+1}(x) - \tilde{r}_j(x) = \hat{s}_{j+1} - \hat{s}_j + (x-1)\left[\frac{G_{j+1}(\tilde{r}_{j+1}(x))}{g_{j+1}(\tilde{r}_{j+1}(x))} - \frac{G_j(\tilde{r}_j(x))}{g_j(\tilde{r}_j(x))}\right]$$

Recall that the last term in brackets is strictly positive. Moreover, note that

$$\hat{s}_{j+1} - \hat{s}_j < s\left(\hat{y}_{j+1}, \omega_{j+1}\right) - s\left(\hat{y}_{j+1}, \omega_j\right) = h(\hat{y}_j, \omega_j) - h(\hat{y}_j, \omega_{j+1}).$$

Thus, the upward IC constraint is satisfied with strict inequality for all $x \leq 1$. Depending on parameters and the properties of G_j and G_{j+1}, it is possible that either the upward IC is satisfied for all levels of $x < \gamma_j^U$ so that $\bar{\beta}_j = \gamma_j^U$, or the upward IC is violated for some $x \in \left(1, \gamma_j^U\right)$. In the latter case, $\bar{\beta}_j < 1$ is given by the lowest level $x < \gamma_j^U$ such that the upward is satisfied with equality for $\beta_j = \beta_{j+1} = x$ and violated for $\beta_j = \beta_{j+1} = x + \varepsilon$ with ε approaching 0 from above. By continuity, $\beta_j^U(\beta_j) > \beta_j$ for all $\beta_j < \bar{\beta}_j$. □

Proof of Proposition 3.1

Proof. The sequence of endogenous social weights must be strictly decreasing in every implementable allocation for every Utilitarian objective as defined in *DUR*. Thus, $\alpha_n^U(c, y) < 1$ must be true in every utilitarian allocation. By Lemma 3.7, this implies $r_n < \hat{s}_n$ in the solution to the relaxed problem.

First, consider the intensive margin. By assumption *REM*, $\hat{s}_n < \bar{\delta}$ so that the extensive margin is relevant in each utilitarian allocation. By Lemma 3.10, $\alpha_n^U < 1$ then implies that the relaxed problem's solution cannot violate the upward IC constraint. The same is true for the solution to a semi-relaxed problem in which all IC constraints below skill level ω_{n-1} are taken into account. By standard arguments, labor supply y_n is undistorted at the intensive margin, whether or not the downward IC constraint between skill types ω_{n-1} and ω_n is binding.

Second, consider the extensive margin. Labor supply by workers with skill ω_n is downward distorted at the extensive margin if and only if $r_n < \hat{s}_n$. Because $y_n = \hat{y}_n$ as argued above, this is equivalent to $T(y_n) > -b$, where $b \geq 0$. If the downward IC between skill types ω_n and ω_{n-1} is not binding, $\alpha_n < 1$ and $\hat{s}_n < \bar{\delta}$ jointly imply that $\hat{\delta}_j = r_j < \hat{s}_j$.

Assume instead that the downward ICs between ω_n and some skill type ω_k with $k \in [1, n-1]$ are binding, while the downward IC between ω_k and ω_{k-1} is not binding. By standard arguments, this implies that $r_k < \tilde{r}_k = \hat{s}_k + \frac{G_k(\tilde{r}_k)}{g_k(\tilde{r}_k)}(\beta_k - 1)$. If $r_k < \hat{s}_k$, then this implies that $T(y_k) > -b$. The binding downward ICs imply that $T(y_n) > T(y_{n-1}) > \cdots > T(y_k)$. Thus, labor supply is downwards distorted at skill level ω_n.

If instead $r_k > \hat{s}_k$, this requires that $\beta_j > 1$ for all $j \leq k$. Then, workers of all skill levels $\omega_j < \omega_k$ upward distorted labor supply at the extensive margin, and $T(y_j) < -b < 0$. This would directly be true for all skill levels for which either no IC is binding, and for which the downward IC is binding. Assume finally that there are skill types ω_a and ω_b between all upward IC constraints are binding. Then, r_a must exceed $\tilde{r}_a = \hat{s}_a + \frac{G_a(\tilde{r}_a)}{g_a(\tilde{r}_a)}(\beta_a - 1) > \hat{s}_a$, so that $T(y_a) < -b$. Moreover, the binding upward ICs imply that $T(y_b) < T(y_{b-1} < \cdots < T(y_a) < -b$. Altogether, this implies that $T(y_n)$ can only be below $-b$ if $T(y_j) < -b$ is also true for all other skill levels. But this is clearly not consistent with the feasibility constraint. More concretely, budget balance requires that $T(y_n) > 0$ so that labor supply is strictly downward distorted at the extensive margin. □

Lemma 3.11. *There is a strictly decreasing sequence $\beta = (\beta_0, \beta_1, \ldots, \beta_n)$ such that the following conditions are satisfied*

1. $\sum_{j=1}^{n} f_j \left[G_j\left(x_j\right)\beta_j + \left(1 - G_j\left(x_j\right)\right)\beta_0\right] = 1$ *with x_j implicitly defined by* $x_j - \hat{s}_j = \frac{G_j(x_j)}{g_j(x_j)}\left(\beta_j - 1\right)$ *for all $\omega_j \in \Omega$,*
2. $\beta_{j+1} \in \left[\beta_j^D(\beta_j), \beta_j^U(\beta_j)\right]$ *for all $\omega_j \in \Omega$, and*
3. $\beta_0 > \beta_1 > 1$.

Proof. Consider the following family of sequences: Let $\tilde{\beta}_1(\varepsilon, \phi) = 1 + \phi$, while $\tilde{\beta}_j(\varepsilon, \phi) = 1 - (j-1)\varepsilon$ for all $j \in [2, n]$, and

$$\tilde{\beta}_0(\varepsilon, \phi) = \frac{1 - \sum_{j=1}^{n} f_j G_j(x_j)\tilde{\beta}_j(\varepsilon, \phi)}{\sum_{j=1}^{n} f_j\left[1 - G(x_j)\right]}$$

For any ε and ϕ, the sequence has average 1. For any $\varepsilon > 0$ and $\phi > 0$, the sequence is strictly decreasing from α_1 on, and $\alpha_1 > 1$. If ϕ is small enough compared to ε, the sequence satisfies $\alpha_0 > \alpha_1$. Lemmas 3.9 and 3.10 imply that $\beta_j^U(x) > x > \beta_j^D(x)$ for all x close enough to 1 and all $\omega_j \in \Omega$. This implies that there is some threshold $\varepsilon_1 > 0$ such that $\tilde{\beta}_{j+1} \in \left[\beta_j^D(\tilde{\beta}_j), \beta_j^U(\tilde{\beta}_j\right]$ for all $j \in [2, n-1]$ for any $\varepsilon \in (0, \varepsilon_1]$. If $\varepsilon > 0$ is small enough, there is moreover a threshold $\phi_1 > 0$ such that $\tilde{\beta}_2 \in \left[\beta_1^D(\tilde{\beta}_1), \beta_1^U(\tilde{\beta}_1\right]$ for all $\phi \in (0, \phi_1]$. If ϕ is small enough compared to ε, the sequence finally satisfies $\tilde{\beta}_0 > \tilde{\beta}_1$. □

Proof of Proposition 3.2

Proof. By Lemma 3.11, there exists a strictly decreasing sequence β such that (a) the solution to the relaxed problem also solves the full problem because $\beta_j \in \left[\alpha_j^D(\beta_{j-1}), \alpha_j^U(\beta_{j-1})\right]$ for all $\omega_j \in \Omega$, and (b) the social weight associated to workers of skill level ω_1 is above the population average of 1. By Lemma 3.7, labor supply is undistorted at the intensive margin for all workers in the relaxed problem's solution. Moreover, $\alpha_j^U > 1$ implies that $\hat{\delta}_1 = r_1 > \hat{s}_1$. Thus, labor supply by workers of skill group 1 is strictly upward distorted at the extensive margin. By construction, any strictly decreasing weight sequence satisfies Assumption *DUR*.

For an example with endogenous social weights, assume that the social objective is given by a member of some family of functions K such that $U(x) = K(a, x)$, where K is twice continuously differentiable in both arguments and satisfies for all $x \in \mathbb{R}$ the following properties: a) $K'(a, x) > 0$ for all $a \geq 0$ and $x \in \mathbb{R}$, and b) $K''(a, x) < 0$ for all $a > 0$ and $\lim_{a \to 0} K''(a, x) = 0$. If assumptions *LC* and *OFCD* hold and $G_{j+1}(\delta) \geq G_j(\delta)$ for all skill levels, the endogenous weight sequence α^U is strictly decreasing for all $a > 0$ (see Proposition 3 in Choné and Laroque 2011). Moreover, there exists again some $\bar{a} > 0$ such that the optimal utilitarian allocation involves no distortions at the intensive margin at any skill level for all $a \in (0, \bar{a})$. If the curvature of $K(a, x)$ is sufficiently small on the interval $x \in [0, \hat{s}_1]$ relative to the interval $[\hat{s}_1, \hat{s}_n]$, then the resulting social weight α_1^U will certainly be below unity, giving again rise to upward distortions at the extensive margin. □

Propositions 3.3 and 3.4 are proven by example in section 3.5.2.

Proofs of Propositions 3.5, 3.6 and 3.7 Proposition 3.5 is proven by a series of lemmas. In particular, a redistributive weight sequence is constructed for which the upward incentive compatibility constraint between skill groups 1 and 2 is binding and y_2 is upwards distorted in the optimal second-best allocation, while labor supply by all other skill groups is undistorted at the intensive margin. The strategy taken is, first, to solve a relaxed problem in which all incentive compatibility constraints are ignored, and second, to construct a sequence of decreasing exogenous weights such that the solution to the relaxed problem also solves the full problem in which the local incentive compatibility constraints are taken into account if and only if Assumptions 3.1 and 3.2 are met. However, the average weight implied by this weight sequence will generally differ from unity. In the third step, we prove that a redistributive weight sequence (with unity average) with the same properties exists, if additionally f_n exceeds some threshold level $\hat{f}_n$. The final step is then to construct a redistributive weight sequence for which y_2 is upwards distorted in the second-best allocation.

Lemma 3.12. *Under assumption 3.1, the solution to the relaxed problem involves*

- *an employment rent of* $r_j^R = \frac{\omega_j/2}{2-\alpha_j(r_j,b)}$ *if* $\bar{\delta} > r_j^R$*, and*
- *the efficient output level* $y_j^R = \omega_j$ *for all skill levels* $\omega_j \in \Omega$*.*

Proof. For the quadratic effort cost function, the efficient levels of output and employment surplus are given by $\hat{y}_j = \omega_j$ and $\hat{s}_j = \hat{y}_j - \frac{\hat{y}_j^2}{2\omega_j} = \frac{\omega_j}{2}$. For the uniform distribution on some interval $\left[0, \bar{\delta}\right]$, we have $\frac{G_j(r_j)}{g_j(r_j)} = r_j$ for any $r_j < \bar{\delta}$. For all $r_j^R < \bar{\delta}$, the first-order condition with respect to r_j can thus be rearranged to have $r_j^R - \hat{s}_j = r_j^R \left(\alpha_j^U - 1\right)$. Solving for r_j^R then gives the equation in Lemma 3.12. □

Lemma 3.13. *Under assumptions 3.1, the thresholds introduced in Lemmas 3.9 and 3.10 are given by* $\underline{\beta}_j = 2 - a < 1$ *and* $\bar{\beta}_j = 2 - \frac{1}{a} > 1$ *for all* $\omega_j \in \Omega$. *Furthermore,* $\beta_j^D(\beta) = 2 - \frac{a}{\frac{1}{2-\beta}+1-\frac{1}{a}}$ *and* $\beta_j^U(\beta) = 2 - \frac{a}{\frac{1}{2-\beta}+a(a-1)}$ *for all skill levels in* Ω. *This implies that* $\beta_j^D(x) > x$ *for all* $x < \underline{\beta}_j$ *and* $\beta_j^U(x) < x$ *for all* $x > \bar{\beta}_j$.

Proof. First, note that under Assumption 3.1, $\frac{\hat{s}_{j+1}}{\hat{s}_j} = \frac{\hat{y}_{j+1}}{\hat{y}_j} = \frac{\omega_{j+1}}{\omega_j} = a > 1$. Thus, the IC constraints are given by

$$r_{j+1} - r_j \geq \frac{y_j^2}{2}\left(\frac{1}{\omega_j} - \frac{1}{\omega_{j+1}}\right), \text{ and}$$
$$r_{j+1} - r_j \leq \frac{y_{j+1}^2}{2}\left(\frac{1}{\omega_j} - \frac{1}{\omega_{j+1}}\right)$$

Plugging in the solution to the relaxed problem gives

$$\frac{\omega_j}{2}\left(\frac{a}{2-\beta_{j+1}} - \frac{1}{2-\beta_j}\right) \geq \frac{\omega_j}{2}\frac{a-1}{a}, \text{ and}$$
$$\frac{\omega_j}{2}\left(\frac{a}{2-\beta_{j+1}} - \frac{1}{2-\beta_j}\right) \leq \frac{a\omega_j}{2}(a-1)$$

Solving for β_{j+1} in both constraints gives the functions β_j^D and β_j^U. Setting $\beta_{j+1} = \beta_j = \beta$, the downward IC is satisfied if $\beta \in \left[2 - a, \gamma_D^j\right]$, and violated for all $\beta < 2 - a = \underline{\beta}_j$. The upward IC is satisfied for all $\beta \leq \min\left\{2 - \frac{1}{a}, \gamma_j^U\right\}$, and violated for all $\beta \in \left(2 - \frac{1}{a}, \gamma_j^U\right)$, if the latter interval is non-empty. □

Proof of Proposition 3.5

Proof. If assumption 3.1 holds, the downward IC constraint is binding whenever $\beta_{j+1} < \beta_j \leq 2 - a$. As $a \in (1, 2)$, this is compatible with strictly positive weights for all skill types. Consider for example weight function β' with $\beta_1' = 2 - a \in (0, 1)$ and $\beta_{j+1}' = \beta_j - \frac{2-a}{n}$ for all $j \in \{2, \ldots, n\}$. Given these weight function, there is a unique weight $\beta_0' > 1$ associated to the unemployed such that average social weight is one. By construction, the social objective corresponding to weight function β' satisfies assumption *DUR*. The social weight of all worker groups is below unity, thus giving rise to downward distortions at the extensive margin. In particular, the optimal level of r_1 will be below $r_1^R < \hat{s}_1$ because the downward IC between skill levels ω_1 and ω_2 is binding. Thus, workers of skill level ω_1 will pay positive participation taxes and have downward distorted labor supply at the extensive margin. As all downward IC constraints are binding, workers of all higher skill levels pay even higher participation taxes and have downward distortions at the extensive margin, too. Thus, there is a utilitarian allocation in which labor supply is distorted downward at the intensive margin everywhere below the very top, and at the extensive margin everywhere. Note that, with $a < 2$, the same pattern of distortions arises in the Rawlsian allocation, which results for social weights $\alpha_j = 0$ for all worker types and $\alpha_0 > 1$ for unemployed agents. □

Proof of Proposition 3.6

Proof. In the following, I proof that there is a threshold $\hat{a}(f_1)$ such that, if $a > \hat{a}(f_1)$ both IC constraints are slack for all welfare functions satisfying Assumption *DUR*. First, note that $\alpha_2^U < 1 < \bar{\beta}_2 = 2 - \frac{1}{a}$ for all $a > 1$ and all utilitarian welfare functions. Thus, the upward IC cannot be violated by the relaxed problem's solution (this is a corollary of Proposition 3.1).

Second, the downward IC is slack for all utilitarian welfare function. By the monotonicity of $\tilde{r}_j$ in β_j, it suffices to show that the downward IC is still slack if α_1^U is at the highest possible level and α_2^U is at the lowest possible level. The lower bound of α_2^U is clearly given by 0. For the upper bound of α_1^U, $\alpha_0^U > \alpha_1^U$ implies that

$$f_1\left[G_1(r_1)\alpha_1^U + [1 - G_1(r_1)]\,\alpha_0^U\right] + f_2\left[G_2(r_2)\alpha_2^U + [1 - G_2(r_2)]\,\alpha_0^U\right] = 1$$
$$\Leftrightarrow \alpha_1^U < \frac{1}{f_1 + f_2\,[1 - G_2(r_2)]} \le \frac{1}{f_1}.$$

Using function $\beta_j^D(\beta)$ as given in Lemma 3.13, the downward IC constraint is satisfied for all combinations $\left(\alpha_1^U, \alpha_2^U\right)$ compatible with *DUR* if

$$\beta_j^D(\zeta) = 2 - \frac{a}{\frac{1}{2-1/f_1} + 1 - \frac{1}{a}} < 0$$
$$\Leftrightarrow a^2 - 2\frac{3f_1 - 1}{2f_1 - 1}a + 2 > 0$$
$$\Rightarrow a > \frac{3f_1 - 1}{2f_1 - 1} + \sqrt{\left(\frac{3f_1 - 1}{2f_1 - 1}\right)^2 - 2} = \hat{a}\,(f_1)$$

Note that the lower root of this quadratic function is below 1, and thus irrelevant due to $a > 1$. Finally, note that $\hat{a}(f_1)$ goes to ∞ for f_1 approaching $1/2$ (from above) and to $2+\sqrt{2}$ for f_1 approaching 1. □

Lemma 3.14. *Under assumptions 3.1 and 3.2, and with the social weight associated to unemployed agents and workers of the highest skill type ω_n given by $\gamma_0 = 2 - \frac{1}{a}$ and $\gamma_n = 2 - \frac{a}{1+a^{2-n}(a^2-1)}$, respectively, the average weight $\bar{\gamma}_n = G_n(\tilde{r}_n)\gamma_n + [1 - G_n(\tilde{r}_n)]\gamma_0$ associated to agents of skill type ω_n is below unity.*

Proof. First, the average $G_n(\tilde{r}_n)\gamma_n + [1 - G_n(\tilde{r}_n)]\gamma_0$ can only be below 1 if γ_n is below 1. Given the definition of n, this is true if and only if

$$\begin{aligned}\gamma_n = 2 - \frac{a}{1+a^{2-n}(a^2-1)} &< 1 \\ \Leftrightarrow a^{2-n}\left(a^2-1\right) &< a-1 \\ \Leftrightarrow (n-2)\ln(a) &> \ln(a+1) \\ \Leftrightarrow n &> 2 + \frac{\ln(a+1)}{\ln(a)},\end{aligned}$$

which is identical to the lower bound imposed on n. Then, the average is negative if the share of workers $G_n(\tilde{r}_n)$ is above $\frac{\gamma_0 - 1}{\gamma_0 - \gamma_n} < 1$. By Lemma 3.12, $G_n(r_n) = \frac{1}{\delta}\frac{\omega_n}{2(2-\gamma_n)}$ for $\alpha_n^U = \gamma_n$. Solving for $\bar{\delta}$ gives the upper bound imposed on the length of the fixed cost space $\left[0, \bar{\delta}\right]$. □

Lemma 3.15. *Under assumptions 3.1 and 3.2, and with the social weight sequence β equaling sequence γ as defined in Assumption 3.3, the upward IC constraint between skill types ω_1 and ω_2, and the downward IC constraints between all other skill types are satisfied with equality. Moreover, $\gamma_j > \underline{\beta}_j = 2 - a$ for all $\omega_j \in \Omega$.*

Proof. The elements of sequence γ are defined as $\gamma_0 = \gamma_1 = \gamma_2 = \bar{\beta}_j = 2 - \frac{1}{a}$ and $\gamma_j = 2 - \frac{a}{1+a^{2-j}(a^2-1)}$ for all $j \geq 3$. This sequence is designed in such a way that, by functions β_j^D and β_j^U defined in Lemma 3.13, $\gamma_2 = \beta_1^U(\gamma_1)$ and $\gamma_j = \beta_j^D(\gamma_{j-1})$ for all $j \geq 3$. As long as $\tilde{r}_j\left(\gamma_j\right) < \bar{\delta}$ for all skill types, this implies that the relaxed problem's solution satisfies with equality one of the IC constraints for each pair ω_j, ω_{j+1} in Ω. By the construction of sequence γ, $\tilde{r}_n(\gamma_n) > \tilde{r}_j(\gamma_j)$ for all $j < n$ (otherwise, the downward ICs could not be satisfied). As $\gamma_n < 1$ by Lemma 3.14, $\tilde{r}_n(\gamma_n) < \hat{s}_n < \bar{\delta}$, where the last inequality follows from assumption *REM*.

Thus, if the weight sequence α^U would be identical to γ, then the upward IC constraint between skill types ω_1 and ω_2 would be satisfied with equality. Moreover, it would be violated for any $\alpha_1^U = \alpha_2^U > \gamma_1$. Furthermore, if $\alpha^U = \gamma$, the downward IC constraints between all pairs ω_j and ω_{j+1} in Ω with $j \geq 2$ are satisfied with equality.

Finally, $\beta_j^D(\beta) \in (2-a, \beta)$ holds if and only if $\beta > 2-a$. Thus, $\gamma_j > \underline{\beta}_j = 2-a$ for all j and the sequence is strictly decreasing with $\gamma_{j+1} < \gamma_j$ for all $j \geq 2$. □

Proof of Proposition 3.7

Proof. Under Assumption 3.3, the population average over sequence γ in the relaxed problem's solution is below unity:

$$\bar{\gamma} = \sum_{j=1}^{n} f_j \left[G_j(\tilde{r}_j)\gamma_j + \left[1 - G_j(\tilde{r}_j)\right] \gamma_0 \right] = \sum_{j=1}^{n} f_j \bar{\gamma}_j = (1-f_n)\,\bar{\gamma}_{-n} + f_n \bar{\gamma}_n$$
$$<1$$

Thus, the sequence γ as defined above cannot be a sequence of social weights. It is however possible to construct a similar sequence $\tilde{\gamma}$ with $\tilde{\gamma}_0 > \tilde{\gamma}_1 > \gamma_1$, $\tilde{\gamma}_2 \in \left(\beta_1^U(\tilde{\gamma}_1), \tilde{\gamma}_1\right)$, $\tilde{\gamma}_3 > \beta_2^D(\tilde{\gamma}_2)$ and $\tilde{\gamma}_{j+1} = \beta_j^D(\gamma_j)$ for all $j \geq 3$ such that the average weight is given by 1. Recall that $\beta_1^U(x) < x$ for all $x > \gamma_1 = 2 - \frac{1}{a}$ by Lemma 3.13.

By construction, the sequence $\tilde{\gamma}$ is strictly decreasing throughout and thus satisfies assumption *DUR*. Furthermore, the relaxed problem's solution satisfies all downward ICs between skill types ω_2 and ω_n, but violates the upward IC constraint between skill types ω_1 and ω_2. The solution to the optimal tax problem thus involves an upward distortion in y_2 at the intensive margin (the proof of Lemma 3.6 below shows in more detail that a binding upward IC constraint gives rise to an upward distortions at the intensive margin). Labor supply by all other skill groups is undistorted at the intensive margin. In particular, as $\tilde{\gamma}$ is designed so that the downward IC between skill types ω_2 and ω_3 is slack, it clearly is still slack with a small upward distortion in y_2.

Furthermore, labor supply by skill types ω_1 and ω_2 is upward distorted at the extensive margin, as $\tilde{r}_1 > \hat{s}_1$ by $\alpha_1 > 1$ and $r_1 > \tilde{r}_1$ due to the binding upward IC, while $r_2 = r_1 + h(y_2, \omega_1) - h(y_2, \omega_2) > \tilde{r}_1 + h(\hat{y}_2, \omega_1) - h(\hat{y}_2, \omega_2) > \hat{s}_2$. □

Under Assumption 3.1, a social objective U giving rise to social weight sequence $\alpha^U = \tilde{\gamma}$ can be derived explicitly. Given the uniform distribution of fixed costs, the social weights are given by $\alpha_j^U(c,y) = \frac{1}{r_j}\left[U\left(r_j + b\right) - U(0)\right]$ for all $j \in [1, n]$, while $\alpha_0^U(c,y) = U'(b)$. Thus, $\alpha^U = \tilde{\gamma}$ if and only if $U\left(r_j + b\right) = U(b) + r_j\tilde{\gamma}_j$ and $U'(b) = \tilde{\gamma}_0$, where $\left(r_j\right)_{j=1}^{n}$ and b solve the set of first-order conditions of the optimal tax problem setting weights according to sequence $\tilde{\gamma}$.

Proof of Lemma 3.4 For $\tilde{A} = 0$, the auxiliary problem is solved by setting $T_1^P = T_2^P = 0$. Consider a relaxed problem in which both IC constraints (3.23) and (3.23) are ignored. Then, $y_j = \hat{y}_j$ for $j \in \{1,2\}$. For $\hat{\delta}_j < \bar{\delta}$, the first-order condition with respect to T_j^P is given by

$$\mathcal{L}_{T_j^P} = f_j \left[G_j(\hat{\delta}_j) \left(-1 + \lambda\right) - \lambda g_j(\hat{\delta}_j) T_j^P \right] = 0,$$

where λ is the Lagrange parameter associated with the feasibility condition. Combining the FOCs with respect to T_1^P and T_2^P, both need to have the same sign. Thus, the feasibility constraint (3.21) can only be satisfied if $T_1^P = T_2^P = 0$, which also satisfies the second-order condition. This solution satisfies both IC constraints (3.23) and (3.23), and involves $\hat{\delta}_1 < \hat{\delta}_2 < \bar{\delta}$ by Assumption *REM*. Thus, the solution to the relaxed problem is given by $T_j^P = 0$ and $y_j = \hat{y}_j$ for both skill groups.

With respect to the upper bound A_{max}, consider first the problem of maximizing revenue from participation taxes, $\sum_{j=1}^{2} f_j G_j \left[s\left((y_j, \omega_j\right) - T_j^P \right] T_j^P$, if the social planner is not restricted by IC constraints. The first-order conditions with respect to T_j^P is given by

$$\begin{aligned} f_j \left[G_j \left(s\left(y_j, \omega_j\right) - T_j^P \right) - g_j \left(s\left(y_j, \omega_j\right) - T_j^P \right) T_j^P \right] &= 0 \\ \Leftrightarrow T_j^P &= \frac{G_j \left(s\left(y_j, \omega_j\right) - T_j^P \right)}{g_j \left(s\left(y_j, \omega_j\right) - T_j^P \right)} \end{aligned}$$

While the left-hand side is increasing in T_j^P, the right-hand side is strictly decreasing by the log-concavity of G_j. As the left-hand side is smaller than the right-hand side for $T_j^P = 0$ and larger for $T_j^P \leq s(y_j, \omega_j) - \underline{\delta}$, the tax maximization problem has a unique maximizer $\left(T_1^{P^*}, T_2^{P^*}\right)$ and a unique maximum $\bar{A} < \sum_{j=1}^{2} f_j \left(s(y_j, \omega_j) - \underline{\delta} \right)$. Taking the IC constraints into account, this maximum is weakly lower, given by some level $A_{max} \leq \bar{A}$. By the construction of A_{max}, the auxiliary problem has no solution in reals for revenue requirements $\tilde{A} > A_{max}$.

With respect to the lower bound A_{min}, note first that, for any level $\tilde{A}$, one Pareto efficient allocation involves uniform taxation $T_1^P = T_2^P = T_E$. In this point, no IC is binding, so that $y_j = \hat{y}_j$ for both groups of workers. If $T_E < \hat{s}_1 - \bar{\delta} < \hat{s}_2 - \bar{\delta}$, then all workers of both skill levels would work under uniform taxation, i.e., $\hat{\delta}_j = \bar{\delta}$ for $j \in \{1,2\}$. Then, the negative revenue created by tax level T_E is given by $[f_1 + f_2]\, T_E$. In every other allocation on the Pareto frontier, workers of one skill group must be better off. Thus, $s(y_j, \omega_j) - T_j^P > \hat{\delta}_j = \bar{\delta}$ for at least one skill group and any $\tilde{A} < [f_1 + f_2] \left(\hat{s}_1 - \bar{\delta} \right) \equiv \underline{A}$. One can conclude that there must be some $A_{min} \in (\underline{A}, 0)$ such that $\hat{\delta}_j < \bar{\delta}$ for both skill levels is true in the surplus-maximizing allocation only if $\tilde{A} < A_{min}$.

It remains to show the *if* part, i.e., uniqueness of the threshold A_{min} satisfying $s(y_j, \omega_j) - T_j^P = \hat{\delta}_j = \bar{\delta}$ for one group and $\hat{\delta}_k < \bar{\delta}$ for the other group. By the downward IC constraint, $\hat{\delta}_2 > \hat{\delta}_1$ as long as both are below $\bar{\delta}$. Reducing $\tilde{A}$ further requires either reducing T_1^P or T_2^P. While the former induces further distortions at the extensive margin, the latter has no effect on labor market participation. Thus, the social planner will choose $\hat{s}_2 - T_j^2$ strictly above $\bar{\delta}$ for any $\tilde{A} < A_{min}$. This implies that $\hat{\delta}_2 = \bar{\delta}$ for all $\tilde{A} < A_{min}$, which is consequently unique.

Proof of Lemma 3.5

Proof. Again, I first solve the auxiliary problem in terms of employment rents (r_1, r_2) and workloads (y_1, y_2). Then, I substitute in the participation tax levels $T_j^P = s(y_j, \omega_j) - r_j$. By Lemma 3.4, $r_j = \hat{\delta}_j = s(y_j, \omega_j) - T_j^P < \bar{\delta}$ for all $\tilde{A} \in (A_{min}, A_{max})$. Thus, the Lagrangian of the auxiliary problem can be written

$$\begin{aligned}\mathcal{L} = &\sum_{j=1}^{2} f_j \left[\int_{\underline{\delta}}^{r_j} g_j(\delta) \left(r_j + b - \delta \right) d\delta + \left(1 - G_j(r_j) \right) b \right] \\ &+ \lambda \left[\sum_{j=1}^{2} f_j G_j(r_j) \left[s\left(y_j, \omega_j \right) - r_j \right] - A - (f_1 + f_2) b \right] \\ &+ \mu_D \left[r_2 - r_1 - h\left(y_1, \omega_1 \right) + h\left(y_1, \omega_2 \right) \right] \\ &+ \mu_U \left[r_1 - r_2 + h\left(y_2, \omega_1 \right) - h\left(y_2, \omega_2 \right) \right],\end{aligned}$$

where $\mu_D > 0$ ($\mu_D = 0$) if the downward IC is binding (not binding), and $\mu_U > 0$ ($\mu_U = 0$) if the upward IC is binding (not binding). The first-order conditions with respect to T_1^P, T_2^P, y_1, y_2 are given as

$$\begin{aligned}\mathcal{L}_{r_1} &= f_1 \left[G_1(r_1) \left(1 - \lambda \right) - \lambda g_1(r_1) \left[s\left(y_1, \omega_1 \right) - r_1 \right] \right] - \mu_D + \mu_U &&= 0 \\ \mathcal{L}_{r_2} &= f_2 \left[G_2(r_2) \left(1 - \lambda \right) - \lambda g_2(r_2) \left[s\left(y_2, \omega_2 \right) - r_2 \right] \right] + \mu_D - \mu_U &&= 0 \\ \mathcal{L}_{y_1} &= \lambda f_1 G_1(r_1) s_y \left(y_1, \omega_1 \right) - \mu_D \left[h_y(y_1, \omega_1) - h_y(y_1, \omega_2) \right] &&= 0 \\ \mathcal{L}_{y_2} &= \lambda f_2 G_2(r_2) s_y \left(y_2, \omega_2 \right) + \mu_U \left[h_y(y_2, \omega_1) - h_y(y_2, \omega_2) \right] &&= 0\end{aligned}$$

By the Lagrange theorem, the multipliers μ_1 and μ_2 are positive if the corresponding IC constraint is binding. If the downward IC constraint is not binding, $\mu_D = 0$, then the first-order condition with respect to Y_1 implies that $s_y\left(y_1, \omega_1 \right) = 0$, i.e., labor supply by low-skill workers is undistorted at the intensive margin with $y_1 = \hat{y}_1$. If the downward IC constraint is instead binding, $\mu_D > 0$, then the single-crossing condition implies that $s_y(y_1, \omega_1) > 0$ must be true, i.e., y_1 is strictly downward distorted. By the corresponding arguments, high-skill labor supply is undistorted if the upward IC is not binding, $\mu_U = 0$, and strictly upward distorted if it is binding.

Thus, we have $y_2 \geq \hat{y}_2 > \hat{y}_1 \geq y_1$ in every solution to this problem, implying that there cannot be pooling of high-skill workers and low-skill workers. The single-crossing condition then implies that $h(y_2, \omega_1) - h(y_2, \omega_2) > h(y_1, \omega_1) - h(y_1, \omega_2)$ holds. Consequently, at most one IC constraint is binding in any implementable allocation.

The main question then is which, if any, of the IC constraints is actually binding in the surplus-maximizing allocation. I first study a relaxed problem in which both incentive compatibility constraints are ignored, and then check explicitly whether the solution to this relaxed problem violates one of the ignored constraints. For clarity, we denote the relaxed problem's solution for variable x by $\tilde{x}$.

The FOCs of this relaxed problem equal the ones of the auxiliary problem, setting $\mu_1 = \mu_2 = 0$. As argued above, the FOC with respect to y_j then requires $s_y(y_j, \omega_j) = 0$. Thus, labor supply is undistorted at the intensive margin, with $y_j = \hat{y}_j$ and $s(y_j, \omega_j) = \hat{s}_j$ for both skill groups. Second, rearranging the first-order conditions with respect to r_j gives

$$\hat{s}_j - \tilde{r}_j = \frac{\lambda - 1}{\lambda} \frac{G_j(\tilde{r}_j)}{g_j(\tilde{r}_j)}$$

Recall that $r_j = c_j - h\left(y_j, \omega_j\right) - b$, so that $\frac{\partial G_j(r_j)}{\partial c_j} = g_j(r_j)$. Thus, the semi-elasticity of the participation share of type ω_j workers with respect to the net labor income c_j is given by the fraction $\frac{g_j(r_j)}{G_j(r_j)}$. Replacing r_j by $s\left(y_j, \omega_j\right) - T_j^P$ gives the inverse elasticity rule (3.25).

The inverse elasticity rule has the following two implications for the relaxed auxiliary problem. First, both participation taxes have the same sign as the semi-elasticities of both skill groups are strictly positive for any $\tilde{A} \in [A_{min}, A_{max}]$. To satisfy the feasibility constraint (3.21), they have to be positive (negative) if $\tilde{A} = A + (f_1 + f_2)b$ is positive (negative). For $\tilde{A}$, both participation taxes have to equal zero.

Second, the higher the semi-elasticity of participation, the lower is the absolute value of the surplus-maximizing participation tax $\tilde{T}_j^P$. Thus, the surplus-maximizing taxes satisfy

$$\frac{\tilde{T}_2^P}{\tilde{T}_1^P} = \frac{G_2(\tilde{r}_2)}{g_2(\tilde{r}_2)} \frac{g_1(\tilde{r}_1)}{G_1(\tilde{r}_1)}.$$

Thus, the optimal participation taxes depend crucially on the relative sizes of both semi-elasticities. For any allocation with $r_2 > r_1$, Assumptions *LC* and *OFCD* imply that the semi-elasticity for low-skill workers must be larger than the one for high-skill workers. More precisely, Assumption *LC* ensures that $\frac{G_2(\tilde{r}_2)}{g_2(\tilde{r}_2)} > \frac{G_2(\tilde{r}_1)}{g_2(\tilde{r}_1)}$ if $r_2 > r_1$. Assumption *OFCD* implies that $\frac{G_2(\tilde{r}_1)}{g_2(\tilde{r}_1)} \geq \frac{G_1(\tilde{r}_1)}{g_1(\tilde{r}_1)}$.

For the non-relaxed auxiliary problem, the inequality $r_2 > r_1$ is ensured for all levels of $\tilde{A}$ by the downward IC constraint. For the relaxed problem, this is immediately clear only for $\tilde{A} = 0$, where $T_2^P = T_1^P = 0$ ensures $r_2 = \hat{s}_2 > r_1 = \hat{s}_1$. It can be shown, however, that there is no level $\tilde{A}$ for which $\frac{G_2(\tilde{r}_2)}{g_2(\tilde{r}_2)} = \frac{G_1(\tilde{r}_1)}{g_1(\tilde{r}_1)}$. By the inverse elasticity rule, it would then be optimal to set identical taxes, $T_2^P = T_1^P$. But then, we would again have $r_2 = \hat{s}_2 - T_2^P > r_1 = \hat{s}_1 - T_1^P$, which implies $\frac{G_2(\tilde{r}_2)}{g_2(\tilde{r}_2)} > \frac{G_1(\tilde{r}_1)}{g_1(\tilde{r}_1)}$. Because the solution $(\tilde{r}_2, \tilde{r}_1)$ is continuous in $\tilde{A}$, this also rules out $\frac{G_2(\tilde{r}_2)}{g_2(\tilde{r}_2)} < \frac{G_1(\tilde{r}_1)}{g_1(\tilde{r}_1)}$ for any levels of $\tilde{A}$. We can conclude that the semi-elasticity of low-skill workers is larger than the one of high-skill workers in every solution to the relaxed auxiliary problem as well.

Thus, the optimal ratio of participation taxes always satisfies $\frac{\tilde{T}_2^P}{\tilde{T}_1^P} > 1$. For all $\tilde{A} < 0$, this implies $\tilde{T}_2^P < \tilde{T}_1^P < 0$. Thus, the relaxed problem's solution satisfies the downward IC $T_2^P - T_1^P \leq s\left(y_2, \omega_2\right) - s\left(y_1, \omega_2\right)$, where the right-hand side is strictly positive. Without further assumptions on the properties of G_1 and G_2, it cannot be determined whether relaxed problem satisfies the upward IC constraint. If it does, the relaxed problem's solution $(\tilde{T}_1^P, \tilde{T}_2^P, \hat{y}_1, \hat{y}_2)$ also solves the non-relaxed problem. Then, no IC constraint is binding in the surplus-maximizing allocation, which furthermore involves $T_2^{P^S} < T_1^{P^S} < 0$ and $r_j > \hat{s}_j$ for both skill levels, as claimed in Lemma 3.5. This will certainly be true in some neighborhood of $\tilde{A} = 0$, where $T_2^P \approx T_1^P$.

If the relaxed problem's solution instead violates the upward IC constraint, this constraint will be binding, and its Lagrange multiplier μ_U will be strictly positive in the surplus-maximizing allocation. Then, the first-order condition with respect to y_2 implies

$$s_y(y_2, \omega_2) = -\frac{\mu_U}{\lambda f_2 G_2(r_2)} \left[h_y(y_2, \omega_1) - h_y(y_2, \omega_2)\right] < 0 \,,$$

where the term $\left[h_y(y_2, \omega_1) - h_y(y_2, \omega_2)\right]$ is strictly positive by the single-crossing property. By the strict concavity of s in y, we have $y_2 > \hat{y}_2 > \hat{y}_1$ and, by standard arguments, $T_1^{P^S} < \tilde{T}_1^P < 0$. Jointly, this implies $T_2^{P^S} = T_1^{P^S} + s(y_2, \omega_1) - s(\hat{y}_1, \omega_1) < T_1^{P^S} < 0$.

In the second case, $\tilde{A} > 0$, a similar argument implies that either the relaxed problem's solution also solves the non-relaxed problem, so that $T_2^{P^S} = \tilde{T}_2^P >$

$T_1^{P^S} = \tilde{T}_1^P > 0$, or the downward IC is binding, $y_1 < \hat{y}_1 < \hat{y}_2$. Then, we have $T_1^{P^S} > \tilde{T}_1^P > 0$, and $T_2^{P^S} = T_1^{P^S} + s(\hat{y}_2, \omega_2) - s(y_1, \omega_2) > T_1^{P^S} > 0$. □

Proof of Lemma 3.6

Proof. First, consider again the relaxed problem. With the assumed uniform distribution on $[0, \bar{\delta}]$, we have $\frac{G_j(r_j)}{g_j(r_j)} = r_j = \hat{s}_j - T_j^P$. Inserting this into the inverse elasticity formulas for optimal tax rates (3.25), the optimal ratio of participation tax rates is given by

$$\frac{\tilde{T}_2^P}{\tilde{T}_1^P} = \frac{\hat{s}_2 - T_2^P}{\hat{s}_1 - T_1^P} = \frac{\hat{s}_2}{\hat{s}_1} > 1$$

This implies that both participation tax levels are strictly increasing in $\tilde{A}$, and that $\frac{dT_2^P}{d\tilde{A}} > \frac{dT_1^P}{d\tilde{A}}$ on the interval $[A_{min}, A_{max}]$. Thus, if there is some level of $\tilde{A}$ at which the upward (downward) IC is violated by the relaxed problem's solution, then the same is also true for all lower (higher) levels.

For ease of notation, define the auxiliary parameter $q \equiv \frac{\hat{s}_1}{\hat{s}_2} < 1$. Thus, the difference in participation taxes is given by $\tilde{T}_2^P - \tilde{T}_1^P = (1-q)\tilde{T}_2^P$. For any level $\tilde{A} < 0$, this difference is negative by Lemma 3.5. The relaxed problem's solution violates the upward IC if

$$\begin{aligned} \tilde{T}_2^P - \tilde{T}_1^P = (1-q)\tilde{T}_2^P &< s(y_2, \omega_1) - s(y_1, \omega_1) \\ \Leftrightarrow \tilde{T}_2^P &< \frac{s(\hat{y}_2, \omega_1) - \hat{s}_1}{1-q} \equiv z_U \end{aligned}$$

Note that term z on the right-hand side of this inequality only depends on exogenous parameters, while the left-hand side is strictly increasing in $\tilde{A}$. On the Pareto-frontier, the feasibility condition holds with equality. Substituting in the optimal ratio of participation tax levels then gives

$$\tilde{A} = f_1 G_1\left(\hat{\delta}_1\right) T_1^P + f_1 G_2\left(\hat{\delta}_2\right) T_2^P = \left(f_1 q^2 + f_2\right) \frac{\hat{s}_2 - T_2^P}{\bar{\delta}} T_2^P.$$

By Lemma 3.6, if the highest fixed cost type $\bar{\delta}$ is sufficiently large, there is a threshold $A_U \in (A_{min}, 0)$ such that the surplus-maximizing allocation involves upward distortions in y_2 for all $\tilde{A} \in (A_{min}, A_U)$. In particular, $A_U > A_{min}$ holds if and only if $\bar{\delta} > \hat{s}_2 - z_U$ is true.

First, the solution to the relaxed problem involves $\hat{\delta}_2 < \bar{\delta}$ if and only if $\tilde{T}_2^P > \hat{s}_2 - \bar{\delta}$. Second, it violates the upward IC constraint if and only if $\tilde{T}_2^P < z_U$. If $\bar{\delta} < \hat{s}_2 - z$, both conditions cannot hold at the same time. Then, the lower bound of $\tilde{A}$ for interior solutions is given by $A_{min} = \left(f_1 q^2 + f_2\right)\left[\hat{s}_2 - \bar{\delta}\right]$, and the upward IC constraint is satisfied for all $\tilde{A} > A_{min}$.

If instead $\bar{\delta} > \hat{s}_2 - z_U$, then both conditions can hold simultaneously. In this case, the upward IC constrained is satisfied by the relaxed problem's solution, and is slack in the surplus-maximizing allocation if and only if $\tilde{A} \geq A_U = \left(f_1 q^2 + f_2\right)\frac{\hat{s}_2 - z}{\bar{\delta}} z$. If $\tilde{A}$ is between A_U and $\left(f_1 q^2 + f_2\right)\left[\hat{s}_2 - \bar{\delta}\right]$, the relaxed problem has an interior solution with $\hat{\delta} = q\hat{\delta}_2 < \hat{\delta}_2 < \bar{\delta}$ and violated the upward IC constraint.

In the non-relaxed problem, the upward IC is thus binding and high-skill labor supply is upwards distorted at the intensive margin, $y_2 > \hat{y}_2$. Moreover, $T_2^P > \tilde{T}_2^P$ because further reductions in T_2^P would require even stronger upward distortions in y_2. Thus, the threshold A_{min} for an interior solution with $\hat{\delta}_2 < \bar{\delta}$ is given by some level $A_{min} < \left(f_1 q^2 + f_2\right)\left[\hat{s}_2 - \bar{\delta}\right] < A_U$.

Similar arguments can be made with respect to the threshold A_D above which the downward IC becomes binding. With uniformly distributed taxes, the downward IC constraint is given by

$$T_2^P - T_1^P \leq s\left(y_2, \omega_2\right) - s\left(y_1, \omega_2\right)$$

For the relaxed problem, the Laffer rates are given by $\tilde{T}_2 = \frac{\hat{s}_2}{2}$ and $\tilde{T}_1 = \frac{\hat{s}_1}{2} = q\tilde{T}_2 < \tilde{T}_2$. Inserting the optimal ratio of taxes, the downward IC constraint then follows as

$$(1-q)\frac{\hat{s}_2}{2} \leq \hat{s}_2 - s\left(\hat{y}_1, \omega_2\right)$$
$$\Leftrightarrow (1+q)\frac{\hat{s}_2}{2} \geq s\left(\hat{y}_1, \omega_2\right)$$

Both sides of this inequality contain only exogenous variables. Whether the downward IC is satisfied or violated for Laffer rates in the relaxed problem thus only depends on properties of the variable cost function h and the difference between skill levels ω_1 and ω_2. If the inequality above is satisfied, then the downward IC is slack in the surplus-maximizing allocation for all levels $\tilde{A}$ in the interval (A_{min}, A_{max}). If is is instead violated, then there is a threshold $A_D \in (0, A_{max})$ such that the downward IC is binding, and y_1 is downward distorted in the surplus-maximizing allocation for all levels of $\tilde{A} \in (A_D, A_{max})$.

This result seems to contrast with the result for threshold A_U, which is above A_{min} if and only if $\bar{\delta}$ is sufficiently large. Allowing for $\underline{\delta} \neq 0$, however, one can also show that A_D is below A_{max} if and only if $\underline{\delta}$ is sufficiently small. □

Appendix 3.B Proofs for Section 3.7

Proof of Proposition 3.8

In the following, I assume that the social planner observes fixed cost types, while the agents are privately informed about their skill types only. Proposition 3.8 studies optimal utilitarian income taxation given this information structure. Then, observable fixed costs types can be used for tagging, i.e., the social planner is able to design specific tax schedules for each fixed cost group. For example, he might choose different benefit payments for unemployed agents with different fixed costs types.

For readability, I denote in the following the consumption-output bundle allocated to agents of type (ω_j, δ) by $c_j(\delta) = c(\omega_j, \delta)$, and $y_j(\delta) = y(\omega_j, \delta)$. Furthermore, I rewrite the joint type distribution Ψ using the functions $G(\delta)$ and $F(\delta)$. $G(\delta)$ denotes the unconditional cdf of fixed costs, with pdf $g(\delta) > 0$ if and only if $\delta \in \Delta$. $F(\delta)$ represents the cdf of skill types ω in the group of agents with fixed cost type δ, while the share of agents with skill type ω_j is denoted by $f_j(\delta)$.

Lemma 3.16. *With observable fixed cost types, an allocation is incentive compatible if and only if, in each group of agents with fixed cost type $\delta \in \Delta$,*

(i) there is a unique threshold type $k(\delta) \in \mathbb{N}$ such that all agents with skill type $\omega_j < \omega_{k(\delta)}$ are unemployed and receive the same cost-specific benefit $b(\delta) \in \mathbb{R}$, while all agents with skill type $\omega_j \geq \omega_{k(\delta)}$ provide positive output $y_j(\delta) > 0$,

(ii) if $\omega_{k(\delta)} > \omega_1$, the allocation of the threshold worker type $(\omega_{k(\delta)}, \delta)$ satisfies

$$c_{k(\delta)}(\delta) - h\left(y_{k(\delta)}, \omega_{k(\delta)}\right) \geq b(\delta) + \delta \geq c_{k(\delta)}(\delta) - h\left(y_{k(\delta)}, \omega_{k(\delta)-1}\right), \text{ and}$$

(iii) if $\omega_{k(\delta)} < \omega_n$, the allocations of all workers with skill types $\omega_j \geq \omega_{k(\delta)}$ satisfy

$$h\left(y_{j+1}(\delta), \omega_j\right) - h\left(y_j(\delta), \omega_j\right) \geq c_{j+1}(\delta) - c_j(\delta) \geq h\left(y_{j+1}(\delta), \omega_{j+1}\right) - h\left(y_j(\delta), \omega_{j+1}\right).$$

Proof. For part (i), consider first two types (ω_i, δ) and (ω_j, δ) such that $y_i(\delta) = y_j(\delta) = 0$. Incentive compatibility requires that $c_i(\delta) = c_j(\delta) = b(\delta)$, which is the benefit receives by all unemployed agents with fixed cost type δ. Second, consider some employed type (ω_j, δ) with $y_j(\delta) > 0$. Incentive compatibility requires $c_j(\delta) - h\left(y_j(\delta), \omega_j\right) - \delta \geq b(\delta)$. By single-crossing, all agents with higher skill type prefer bundle $\left(c_j(\delta), y_j(\delta)\right)$ strictly to bundle $(b(\delta), 0)$, and must thus provide positive output in any incentive-compatible allocation. Symmetrically, if there is some type (ω_i, δ) that weakly prefers unemployment, then all agents with lower skill type will strictly prefer unemployment. Thus, there is a unique threshold $\omega_{k(\delta)} \in [\omega_1, \omega_n]$ for each fixed cost level.

For parts (ii) and (iii), note that we only need to consider incentive compatibility constraints between agents with identical fixed cost δ. The inequalities given in part (ii) guarantee that $\omega_{k(\delta)}$ is indeed the threshold skill level. The inequalities in part (iii) represent standard IC constraints between adjacent skill types. As usual, the single-crossing property implies that global incentive-compatibility holds if and only if all local IC constraints are satisfied. □

Lemma 3.17. *At any utilitarian allocation, the downward IC constraint for the threshold worker type $\omega_{k(\delta)}$ is binding in each group of agents with fixed cost type $\delta \in \Delta$, i.e., $c_{k(\delta)}(\delta) - h\left(y_{k(\delta)}, \omega_{k(\delta)}\right) = b(\delta) + \delta$ holds.*

Proof. Given Lemma 3.16, the planner's objective can be written

$$W(c, y) = \int_{\underline{\delta}}^{\bar{\delta}} W_\delta(c(\delta), y(\delta)) dG(\delta),$$

where the cost-group welfare level $W_\delta(c(\delta), y(\delta))$ for each $\delta \in \Delta$ is given by

$$W_\delta(c(\delta), y(\delta)) = F_{k(\delta)-1}(\delta) U\left[b(\delta)\right] + \sum_{j=k(\delta)}^{n} f_j(\delta) U\left[c_j(\delta) - h\left(y_j(\delta), \omega_j\right) - \delta\right].$$

The feasibility constraint can be divided into a global constraint $\int_{\underline{\delta}}^{\bar{\delta}} A(\delta) dG(\delta) \geq 0$ and a set of cost-dependent constraints $\sum_{j=k(\delta)}^{n} f_j(\delta)\left[y_j(\delta) - c_j(\delta) + b(\delta)\right] \geq b(\delta) + A(\delta)$. The set of incentive-compatibility constraints is given as in parts (ii) and (iii) of Lemma 3.16.

By standard arguments, any utilitarian allocation satisfies the feasibility constraints with equality. The function of cost-specific revenues $A(\delta)$ is chosen to equate average marginal utilities (and average endogenous weights) in all fixed cost groups, which typically implies redistribution from low-cost groups to high-skill groups. Within each fixed cost group, the functions $c(\delta)$, $y(\delta)$ and the benefit $b(\delta)$ are chosen to maximize cost-specific welfare $W_\delta(c(\delta), y(\delta))$ subject to the cost-specific revenue requirement $A(\delta)$ and the cost-specific IC constraints.

A proof by contradiction demonstrates that the threshold worker type $\left(\omega_{k(\delta)}, \delta\right)$ must be indifferent between employment and unemployment, i.e., the downward IC between types $\left(\omega_{k(\delta)}, \delta\right)$ and $\left(\omega_{k(\delta)-1}, \delta\right)$ must be binding in any utilitarian allocation. Assume this were not the case, i.e., there is an incentive compatible and feasible allocation that maximizes welfare and involves $c_{k(\delta)}(\delta) - h\left(y_{k(\delta)}, \omega_{k(\delta)}\right) >$

$b(\delta)+\delta$. Then, leaving $y(\delta)$ unchanged, reducing $c_j(\delta)$ uniformly by a small amount $\varepsilon > 0$ for all workers with $\omega_j \geq \omega_{k(\delta)}$ and increasing the unemployment benefit $b(\delta)$ by $\varepsilon \left[1 - F_{k(\delta)-1}(\delta)\right] / F_{k(\delta)-1}(\delta)$ would be possible without violating feasibility or incentive-compatibility. The marginal welfare effect of this variation is given by

$$\frac{dW_\delta}{d\varepsilon} = \left[1 - F_{k(\delta)-1}(\delta)\right] \alpha_0(\delta) - \sum_{j=k(\delta)}^{n} f_j(\delta)\alpha_j(\delta) > 0$$

This is positive as Assumption *DUR* δ implies $\alpha_0'(c,y,\delta) > \alpha_j'(c,y,\delta)$ for all $j \geq k(\delta)$. Thus, the original allocation cannot be a utilitarian allocation.

Note that, with observable fixed costs, increasing $b(\delta)$ induces extensive margin responses if and only if it conflicts with the IC constraint for type $(\omega_{k(\delta)-1}, \delta)$. Thus, an equity-efficiency tradeoff can arise if and only if the downward IC of type $\left(\omega_{k(\delta)}, \delta\right)$ is binding. □

Lemma 3.18. *At any utilitarian allocation, all downward IC constraints between active workers with* $\omega_j \geq \omega_{k(\delta)}$ *are binding in each group of agents with fixed cost type* $\delta \in \Delta$*:*

$$\begin{aligned} c_{j+1}(\delta) - h\left(y_{j+1}(\delta), \omega_{j+1}\right) &= c_j(\delta) - h\left(y_j(\delta), \omega_{j+1}\right) \\ &= b(\delta) + \delta + \sum_{l=k(\delta)}^{j} \left[h\left(y_l(\delta), \omega_l\right) - h\left(y_l(\delta), \omega_{l+1}\right)\right] \quad . \end{aligned}$$

Proof. I only provide a sketch of the proof, because it is based on standard arguments that are familiar from the literature on optimal income taxation with labor supply responses at the intensive margin only (see, e.g., Mirrlees 1971). Consider some feasible and incentive-compatible allocation in which the downward IC constraint between types $\left(\omega_j, \delta\right)$ and $\left(\omega_{j+1}, \delta\right)$ is not binding, where $\omega_j \geq \omega_{k(\delta)}$. Then, it is possible to reduce consumption uniformly for all agents with skill type $\omega_i \geq \omega_{j+1}$, and using these resources for uniform transfers towards all agents with skill types $\omega_l \leq \omega_j$, until the downward IC constraint between agents with skill types ω_j and ω_{j+1} becomes binding. This is consistent with incentive-compatibility and feasibility, and yields a marginal welfare increase of

$$\frac{dW_\delta}{d\varepsilon} = \frac{1 - F_j(\delta)}{F_j(\delta)} \left[F_{k(\delta)-1}(\delta)\alpha_0(\delta) + \sum_{l=k(\delta)}^{j} f_l(\delta)\alpha_l(\delta)\right]$$

$$- \sum_{l=j+1}^{n} f_l(\delta)\alpha_l(\delta) > 0$$

As social weights are strictly decreasing in ω by Assumption *DUR* δ, this induces a strict welfare gain.

Thus, the downward IC must be binding between all pairs of skill types above $\omega_{k(\delta)}$, as well as for the threshold skill type $\omega_{k(\delta)}$. Consequently, $c_j(\delta)$ follows as a function of δ, $b(\delta)$ and the output levels $y_i(\delta)$ of all skill types $\omega_i \leq \omega_j$. □

Lemma 3.19. *At the intensive margin, labor supply is undistorted at the top skill level ω_n and strictly downwards distorted everywhere below the top for all workers in each group of agents with fixed cost type $\delta \in \Delta$.*

Proof. In the following, we write $x_j^\delta = x_j(\delta)$ for $x \in \{y, b, f, \lambda, A\}$ for reasons of readability. By Lemmas 3.17 and 3.18, the group-specific Lagrangian can be written

$$\begin{aligned}\mathcal{L}^\delta =& F_{k(\delta)-1}^\delta U\left[b^\delta\right] + \sum_{j=k(\delta)}^{n} f_j^\delta U \left[b^\delta + \sum_{l=k(\delta)}^{j-1} \left[h\left(y_l^\delta, \omega_l\right) - h\left(y_l^\delta, \omega_{l+1}\right)\right]\right] \\ &+ \lambda^\delta \left\{ \sum_{j=k(\delta)}^{n} f_j^\delta \left[y_j^\delta - h\left(y_j^\delta, \omega_j\right) - \delta - \sum_{l=k(\delta)}^{j-1} \left[h\left(y_l^\delta, \omega_l\right) - h\left(y_l^\delta, \omega_{l+1}\right)\right]\right] \right. \\ &\left. - b^\delta - A^\delta \right\}\end{aligned}$$

Taking the derivative with respect to $b(\delta)$ implies that $\lambda(\delta)$ equals the cost-specific average weight $\bar{\alpha}(\delta)$. The derivative with respect to $y_j(\delta)$ is given by

$$\begin{aligned}\mathcal{L}_{y_j} =& \underbrace{\left[h_1\left(y_j(\delta), \omega_j\right) - h_1\left(y_j(\delta), \omega_{j+1}\right)\right]}_{>0} \underbrace{\sum_{l=j+1}^{n} f_l(\delta)\left[\alpha_l(\delta) - \lambda(\delta)\right]}_{<0} \\ &+ \lambda(\delta) f_j(\delta) \underbrace{\left[1 - h_1(y_j(\delta), \omega_j\right]}_{>0} = 0.\end{aligned}$$

By the single-crossing property, the term in the first bracket is strictly positive. As the social weights are decreasing with ω, the second term is strictly negative. Thus, the first-order condition can only be satisfied if $h_1\left(y_j(\delta), \omega_j\right) < 1$. In other words, labor supply is strictly downward distorted for all worker types below ω_n, $y_j(\delta) < \hat{y}_j$, in any utilitarian allocation. For the top skill level, the familiar "no-distortion-at-the-top" result prevails. Intuitively, the downward distortion in $y_j(\delta)$ slackens the downward IC constraint between types (ω_{j+1}, δ) and (ω_j, δ), allowing to redistribute more resources to lower skill types. Starting from $y_j(\delta) = \hat{y}_j$, this has negligible efficiency costs, but allows to achieve first-order equity gains. Again, the crucial difference to the model with two-dimensional private information is that changes in y_j do not involve labor supply responses at the extensive margin. □

Lemma 3.20. *At the extensive margin, labor supply is weakly downward distorted in each group of agents with fixed cost type $\delta \in \Delta$, and strictly downward distorted for some fixed cost levels $\delta \in \Delta$.*

Proof. Again, the Lemma can be proven by contradiction. Assume that a utilitarian allocation involves, for workers with skill type ω_j, some output requirements $\left(y_j(\delta)\right)_{j=1}^n$ and $\hat{s}_{k(\delta)} < \delta$, i.e., upward distortions in labor supply at the extensive margin. By Lemmas 3.17 and 3.18, all downward IC constraints must be binding in any utilitarian allocation. Thus, an agent with threshold skill type $\omega_{k(\delta)}$ must be indifferent between employment and unemployment. In this allocation, the level of the unemployment benefit $b(\delta)$ is pinned down by the feasibility constraint:

$$\begin{aligned} b(\delta) = \sum_{j=k(\delta)}^{n} f_j(\delta) \left[y_j(\delta) - h(y_j(\delta), \omega_j\right] - \delta - A(\delta) \\ - \sum_{l=k(\delta)}^{j-1} \left[h(y_l(\delta), \omega_l) - h(y_l(\delta), \omega_{l+1})\right] \end{aligned}$$

If $\hat{s}_{k(\delta)} \leq \delta$, welfare can be increased by removing agents of type $(\omega_{k(\delta)}, \delta)$ from the labor market by setting $y_{k(\delta)}(\delta) = 0$, while keeping the workloads and consumption levels of all agents with $\omega_j > \omega_{k(\delta)}$ constant. Because the former agents were indifferent between working and staying unemployed before, this is possible without violating any IC constraint. All else equal, the feasibility constraint is relaxed by

$$-f_{k(\delta)}(\delta) \left[y_{k(\delta)}(\delta) - h\left(y_{k(\delta)}(\delta), \omega_j\right) - \delta\right] > -f_{k(\delta)}(\delta) \left[\hat{s}_{k(\delta)} - \delta\right] \geq 0.$$

The first inequality follows due to the downward distortion in $y_{k(\delta)}(\delta)$ at the intensive margin (see Lemma 3.19), the second one by assumption. As the feasibility constraint is slack after this deviation, the consumption levels of all agents in the skill group can be increased uniformly, inducing a Pareto improvement. Consequently, the initial allocation with upward distortions at the extensive margin cannot represent a utilitarian optimum.

By the same argument, labor supply is strictly downward distorted at the intensive margin in all fixed costs groups such that $\delta = \hat{s}_j$ for some $\omega_j \in \Omega$. For skill groups with $\delta \in \left(\hat{s}_j, \hat{s}_{j+1}\right)$, in contrast, labor supply is strictly downward distorted if and only if the social planner has a sufficiently strong desire for redistribution. □

Proof of Proposition 3.9

In the following, I assume that the social planner observes skill types, while the agents are privately informed about their fixed cost types only. Proposition 3.9 studies optimal utilitarian income taxation given this information structure. Then, the social planner can use skill types for tagging, i.e., can condition unemployment benefits as well as tax payments directly on an agent's skill type. Proposition 3.9 is proven by a series of lemmas.

Lemma 3.21. *In every implementable allocation, there is a unique fixed cost threshold type $\tilde{\delta}_j \in \Delta$ for each skill level $\omega_j \in \Omega$ such that each agent with skill type ω_j and*

(i) fixed cost type $\delta > \tilde{\delta}_j$ is unemployed and consumes a skill-specific benefit $b_j \in \mathbb{R}$,

(ii) fixed cost type $\delta \leq \tilde{\delta}_j$ provides positive output $y(\omega_j, \delta) > 0$ and enjoys a gross (of the fixed cost) utility $c(\omega_j, \delta) - h\left[y(\omega_j, \delta), \omega_j\right] = z_j = b_j + \tilde{\delta}_j$.

Proof. For part (i), consider agents with two fixed cost types δ and $\delta' \neq \delta$ such that $y(\omega_j, \delta) = y(\omega_j, \delta') = 0$. Incentive compatibility requires that $c(\omega_j, \delta) = c(\omega_j, \delta') = b_j$, which represents the unemployment benefit. For part (ii), consider agents with two fixed cost types δ and $\delta' \neq \delta$ such that $y(\omega_j, \delta) > 0$ and $y(\omega_j, \delta') > 0$. Incentive compatibility requires that $c(\omega_j, \delta) - h\left[y(\omega_j, \delta), \omega_j\right] = c(\omega_j, \delta') - h\left[y(\omega_j, \delta'), \omega_j\right] = z_j$. Note that incentive compatibility does not imply pooling of all workers with skill type ω_j. For the threshold type $\tilde{\delta}_j$, a worker with type (ω_j, δ) prefers his bundle to $(b_j, 0)$ if and only if $c(\omega_j, \delta) - h\left[y(\omega_j, \delta), \omega_j\right] - \delta = z_j - \delta \geq b_j$, i.e., if $\delta \leq z_j - b_j = \tilde{\delta}_j$. Symmetrically, unemployed agents prefer bundle $(b_j, 0)$ to the bundle of any worker if and only if $\delta \geq z_j - b_j = \tilde{\delta}_j$. □

Lemma 3.22. *An allocation is Pareto efficient in the set of implementable allocations if and only if, for each skill type $\omega_j \in \Omega$, all workers are allocated the same bundle $(c_j, \hat{y}_j)$ with undistorted labor supply at the intensive margin.*

Proof. By Lemma 3.21, each worker with type (ω_j, δ) is indifferent between his bundle $\left(c(\omega_j, \delta), y(\omega_j, \delta)\right)$ and the bundles of all other types (ω_j, δ) such that

$\delta \leq \tilde{\delta}_j$. With observable skills, the social planner does not have to satisfy incentive compatibility constraints between agents with different skill types. Thus, the social planner can allocate to all workers with skill type ω_j the bundle (c, y) which minimizes $(c-y)$ subject to $c-h(y, \omega_j) \geq z_j$. By Lemma 3.3, the solution to this problem is given by $\hat{y}_j$, i.e., undistorted labor supply at the intensive margin. The consumption level c_j follows as $c_j = z_j + h\left(\hat{y}_j, \omega_j\right)$. If a positive measure of agents would provide some positive output $y \neq \hat{y}_j$, then giving them instead bundle $(c_j, \hat{y}_j)$ and redistributing the saved resources lump-sum to all agents without violating any IC constraint would lead to a Pareto improvement. □

Lemma 3.23. *In any utilitarian allocation, labor supply is strictly downward distorted at the extensive margin with $\tilde{\delta}_j \in \left(\underline{\delta}, \hat{s}_j\right)$ in all skill groups.*

Proof. By Lemmas 3.21 and 3.22, the Lagrangian for the problem of optimally redistributing resources within skill group ω_j can be written as

$$\begin{aligned}\mathcal{L}_j = & \int_{\underline{\delta}}^{\tilde{\delta}_j} g_j(\delta) U\left[c_j - h\left(\hat{y}_j, \omega_j\right) - \delta\right] d\delta + \left[1 - G_j(\tilde{\delta}_j)\right] U\left[b_j\right] \\ & + \lambda_j \left[G_j(\tilde{\delta}_j)\left(y_j - c_j + b_j\right) - b_j - A - j\right],\end{aligned}$$

with $\tilde{\delta}_j = c_j - h\left(\hat{y}_j, \omega_j\right) - b_j$ if $\tilde{\delta}_j \in \left(\underline{\delta}, \bar{\delta}\right)$. Assume for the moment that the latter is true. Combining the first-order conditions with respect to b_j and c_j, the Lagrange multiplier associated with the feasibility constraint equals the average social weight in skill group ω_j, given by

$$\lambda_j = \int_{\underline{\delta}}^{\tilde{\delta}_j} g_j(\delta) U'\left[c_j - h\left(\hat{y}_j, \omega_j\right) - \delta\right] d\delta + \left[1 - G_j(\tilde{\delta}_j)\right] U'\left[b_j\right].$$

The first-order condition with respect to b_j reads

$$\frac{\partial \mathcal{L}_j}{\partial b_j} = \left[1 - G - j(\tilde{\delta}_j)\right]\left[U'\left(b_j\right) - \lambda_j\right] - \lambda_j g_j\left(\tilde{\delta}_j\right)\left[\hat{y}_j - c_j + b_j\right] = 0.$$

For $\tilde{\delta}_j \in \left(\underline{\delta}, \bar{\delta}\right)$, the second bracket in this equation is positive by Assumption *DUR* ω. The same is true for the second bracket. Thus, the optimal level of c_j must be smaller than $\hat{y}_j + b_j$ to satisfy the first-order condition. For the threshold cost type, this implies $\tilde{\delta}_j = c_j - h\left(\hat{y}_j, \omega_j\right) - b_j < \hat{y}_j - h\left(\hat{y}_j, \omega_j\right) = \hat{s}_j$.

For $\tilde{\delta}_j = \underline{\delta}$, the first-order condition with respect to b_j cannot be satisfied. In this case, all agents in this skill group would be unemployed so that $\lambda_j = U'(b_j)$. Then, $y_j - c_j + b_j = 0$ would have to be true, implying $\tilde{\delta}_j = \hat{s}_j$. By Assumption *REM*, this is however inconsistent with $\tilde{\delta}_j = \underline{\delta}$. Similarly, the FOC with respect to

c_j cannot be satisfied for the corner solution $\tilde{\delta}_j = \bar{\delta}$. Thus, labor supply is strictly downward distorted with $\tilde{\delta}_j \in \left(\underline{\delta}, \hat{s}_j\right)$ in all skill groups. □

Bibliography

Adams, J. (1999). “Policy Divergence in Multicandidate Probabilistic Spatial Voting”. *Public Choice* 100, 103–122.

Ansolabehere, S. and J. M. Snyder (2000). “Valence Politics and Equilibrium in Spatial Election Models”. *Public Choice* 103 (3), 327–336.

Aragones, E. and T. R. Palfrey (2002). “Mixed Equilibrium in a Downsian Model with a Favored Candidate”. *Journal of Economic Theory* 103, 131–161.

Armingeon, K. (2002). “The Effects of Negotiation Democracy: A Comparative Analysis”. *European Journal of Political Research* 41, 81–105.

Armingeon, K. et al. (2011). *Comparative Political Data Set III 1990-2009.* Institute of Political Science, University of Berne.

Atkinson, A. (1973). “How Progressive Should Income Tax Be?” In: *Essays on Modern Economics.* Ed. by M. Parkin and R. Nobay. London: Longman.

Barro, R. J. (1991). “Economic Growth in a Cross Section of Countries”. *Quarterly Journal of Economics* 106, 407–43.

Beaudry, P., C. Blackorby, and D. Szalay (2009). “Taxes and Employment Subsidies in Optimal Redistribution Programs”. *American Economic Review* 99, 216–42.

Beck, T. et al. (2001). “New Tools in Comparative Political Economy: The Database of Political Institutions”. *World Bank Economic Review* 15, 165–176.

Besley, T. (2005). “Political Selection”. *Journal of Economic Perspectives* 19, 43–60.

Besley, T. and S. Coate (1997). “An Economic Model of Representative Democracy”. *Quarterly Journal of Economics* 112, 85–114.

Besley, T. and M. Smart (2007). “Fiscal Restraints and Voter Welfare”. *Journal of Public Economics* 91, 755–773.

Blume, L. et al. (2009). “The Economic Effects of Constitutions: Replicating-and Extending-Persson and Tabellini”. *Public Choice* 139, 197–225.

Cadigan, J. and E. Janeba (2002). “A Citizen-Candidate Model with Sequential Elections”. *Journal of Theoretical Politics* 14, 387–407.

Campante, F. R. (2011). “Redistribution in a Model of Voting and Campaign Contributions”. *Journal of Public Economics* 95, 646–656.

Caselli, F. and M. Morelli (2004). “Bad Politicians”. *Journal of Public Economics* 88, 759–782.

Cho, I.-K. and D. M. Kreps (1987). "Signaling Games and Stable Equilibria". *Quarterly Journal of Economics* 102, 179–222.

Choné, P. and G. Laroque (2010). "Negative Marginal Tax Rates and Heterogeneity". *American Economic Review* 100, 2532–47.

——— (2011). "Optimal Taxation in the Extensive Model". *Journal of Economic Theory* 146, 425–453.

Christiansen, V. (2012). *Optimal Participation Taxes.* CESifo Working Paper Series. http://ssrn.com/abstract=2155601.

Dhillon, A. (2004). *Political Parties And Coalition Formation.* Warwick Economics Research Paper Series. http://ideas.repec.org/p/wrk/warwec/697.html.

Diamond, P. (1980). "Income Taxation with Fixed Hours of Work". *Journal of Public Economics* 13, 101–110.

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

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

Eguia, J. (2007). "Citizen Candidates Under Uncertainty". *Social Choice and Welfare* 29, 317–331.

Enikolopov, R. and E. Zhuravskaya (2007). "Decentralization and Political Institutions". *Journal of Public Economics* 91, 2261–2290.

Feld, L. P. and S. Voigt (2003). "Economic Growth and Judicial Independence: Cross-Country Evidence Using a New Set of Indicators". *European Journal of Political Economy* 19, 497–527.

Gerber, A. and I. Ortuño-Ortin (1998). "Political Compromise and Endogenous Formation of Coalitions". *Social Choice and Welfare* 15, 445–454.

Gomberg, A., F. Marhuenda, and I. Ortuño-Ortin (2004). "A Model of Endogenous Political Party Platforms". *Economic Theory* 24, 373–394.

Grunewald, A., E. Hansen, and G. Pönitzsch (2013). *Political Selection and the Concentration of Political Power.* Working Paper. http://ssrn.com/abstract=2239979.

Guesnerie, R. (1995). *A Contribution to the Pure Theory of Taxation.* Cambridge: Cambridge University Press.

Hammond, P. (1979). "Straightforward Individual Incentive Compatibility in Large Economies". *Review of Economic Studies* 46, 263–282.

Hellwig, M. F. (2005). "A Utilitarian Approach to the Provision and Pricing of Excludable Public Goods". *Journal of Public Economics* 89, 1981–2003.

——— (2007). "A Contribution to the Theory of Optimal Utilitarian Income Taxation". *Journal of Public Economics* 91, 1449–1477.

Herrera, H., D. K. Levine, and C. Martinelli (2008). "Policy Platforms, Campaign Spending and Voter Participation". *Journal of Public Economics* 92, 501 –513.

Iaryczower, M. and A. Mattozzi (2008). *Ideology and Competence in Alternative Electoral Systems.* Levine's Working Paper Archive. http://ideas.repec.org/p/cla/levarc/122247000000002387.html.

IDEA (2004). *Handbook of Electoral System Design.* International Institute for Democracy and Electoral Assistance.

Immervoll, H. et al. (2007). "Welfare Reform in European Countries: A Microsimulation Analysis". *The Economic Journal* 117, 1–44.

ISSP Research Group (2012). *International Social Survey Programme 2004: Citizenship I (ISSP 2004).* GESIS Data Archive, Cologne, ZA3950 Data file Version 1.3.0, doi: 10.4232/1.11372.

Jacquet, L., E. Lehmann, and B. Van der Linden (2013). "Optimal redistributive taxation with both extensive and intensive responses". *Journal of Economic Theory* 148, 1770–1805.

Juhn, C. et al. (1991). "Why Has the Natural Rate of Unemployment Increased over Time?" *Brookings Papers on Economic Activity* 1991, pp. 75–142.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2009). *Governance Matters VIII : Aggregate and Individual Governance Indicators 1996-2008.* World Bank Policy Research Working Paper No. 4978.

Keefer, P. and D. Stasavage (2003). "The Limits of Delegation: Veto Players, Central Bank Independence and the Credibility of Monetary Policy". *American Political Science Review* 97, 407–423.

King, A. (2002). "Do Leaders' Personalities Really Matter?" In: *Leaders' Personalities and the Outcomes of Democratic Elections.* Ed. by A. King. Oxford: Oxford University Press, 1–42.

Krasa, S. and M. Polborn (2011). *A Political-Economy Model of Taxation and Government Expenditures with Differentiated Candidates.* Working Paper. http://works.bepress.com/polborn/20.

Levy, G. (2004). "A Model of Political Parties". *Journal of Economic Theory* 115, 250–277.

Lijphart, A. (1999). *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries.* New Haven: Yale University Press.

——— (2012). *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries.* 2nd ed. New Haven: Yale University Press.

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

Lizzeri, A. and N. Persico (2001). "The Provision of Public Goods under Alternative Electoral Incentives". *American Economic Review* 91, 225–239.

Lizzeri, A. and N. Persico (2005). "A Drawback of Electoral Competition". *Journal of the European Economic Association* 3, 1318–1348.

Lorenz, N. and D. Sachs (2011). *Optimal Nonlinear Taxation, Minimum Hours, and the Earned Income Tax Credit.* University of Trier: Research Papers in Economics. http://ideas.repec.org/p/trr/wpaper/201111.html.

Ma, Y., M. Genton, and E. Parzen (2011). "Asymptotic Properties of Sample Quantiles of Discrete Distributions". *Annals of the Institute of Statistical Mathematics* 63, 227–243.

Madison, J. (1788a). *Federalist No. 51, The Structure of the Government Must Furnish the Proper Checks and Balances Between the Different Departments.* http://thomas.loc.gov/home/histdox/fed_51.html.

——— (1788b). *Federalist No. 57, The Alleged Tendency of the New Plan to Elevate the Few at the Expense of the Many Considered in Connection with Representation.* http://thomas.loc.gov/home/histdox/fed_57.html.

Marshall, M. G. and K. Jaggers (2010). *Polity IV Project: Dataset Users' Manual.* http://www.systemicpeace.org/inscr/p4manualv2010.pdf.

Maskin, E. and J. Tirole (2004). "The Politician and the Judge: Accountability in Government". *American Economic Review* 94, 1034–1054.

Mattozzi, A. and A. Merlo (2008). "Political Careers or Career Politicians?" *Journal of Public Economics* 92, 597–608.

Meghir, C. and D. Phillips (2010). "Labour Supply and Taxes". In: *Dimensions of Tax Design: the Mirrlees Review.* Ed. by J. Mirrlees et al. Oxford University Press. Chap. 3, 90–201.

Messner, M. and M. Polborn (2004). "Paying Politicians". *Journal of Public Economics* 88, 2423–2445.

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

Morelli, M. (2004). "Party Formation and Policy Outcomes under Different Electoral Systems". *Review of Economic Studies* 71, 829–853.

Ortuño-Ortin, I. and C. Schultz (2005). "Public Funding of Political Parties". *Journal of Public Economic Theory* 7, 781–791.

Osborne, M. J. and A. Slivinski (1996). "A Model of Political Competition with Citizen-Candidates". *Quarterly Journal of Economics* 111, 65–96.

Osborne, M. J. and R. Tourky (2008). "Party Formation in Single-Issue Politics". *Journal of the European Economic Association* 6, 974–1005.

Pancer, S. M., S. D. Brown, and C. W. Barr (1999). "Forming Impressions of Political Leaders: A Cross-National Comparison". *Political Psychology* 20, 345–368.

Persson, T. and G. Tabellini (2003). *The Economic Effects of Constitutions.* Cambridge: The MIT Press.

——— (2004). "Constitutional Rules and Fiscal Policy Outcomes". *American Economic Review* 94, 25–45.

Poutvaara, P. (2003). "Party Platforms with Endogenous Party Membership". *Public Choice* 117, 79–98.

Poutvaara, P. and T. Takalo (2007). "Candidate Quality". *International Tax and Public Finance* 14, 7–27.

Rivière, A. (1999). *Citizen Candidacy, Party Formation and Duverger's Law*. Working Paper. URL: http://idei.fr/doc/conf/eco/riviere.pdf.

Rochet, J.-C. and L. A. Stole (2002). "Nonlinear Pricing with Random Participation". *Review of Economic Studies* 69, 277–311.

Roemer, J. E. (2006). "Party Competition under Private and Public Financing: A Comparison of Institutions". *Advances in Theoretical Economics* 6, 1229.

Saez, E. (2002). "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses". *The Quarterly Journal of Economics* 117, 1039–1073.

Sahuguet, N. and N. Persico (2006). "Campaign Spending Regulation in a Model of Redistributive Politics". *Economic Theory* 28, 95–124.

Sala-i-Martin, X. (1994). "Cross-Sectional Regressions and the Empirics of Economic Growth". *European Economic Review* 38, 739–747.

——— (1997). "I Just Ran Two Million Regressions". *American Economic Review* 87, 178–83.

Seade, J. K. (1977). "On the Shape of Optimal Tax Schedules". *Journal of Public Economics* 7, 203–235.

Seade, J. (1982). "On the Sign of the Optimum Marginal Income Tax". *Review of Economic Studies* 49, 637–43.

Smart, M. and D. M. Sturm (2013). "Term Limits and Electoral Accountability". *Journal of Public Economics* 107, 93–102.

Snyder, J. M. and M. M. Ting (2002). "An Informational Rationale for Political Parties". *American Journal of Political Science* 46, 90–110.

Stokes, D. E., A. Campbell, and W. E. Miller (1958). "Components of Electoral Decision". *The American Political Science Review* 52, pp. 367–387.

Sun, Y. (2006). "The Exact Law of Large Numbers via Fubini Extension and Characterization of Insurable Risks". *Journal of Economic Theory* 126, 31–69.

Tsebelis, G. (2002). *Veto Players. How Political Institutions Work*. Princeton University Press/Russell Sage Foundation.

Voigt, S. (2011). "Positive Constitutional Economics II – A Survey of Recent Developments". *Public Choice* 146, 205–256.

World Bank (2012). *World Development Indicators.* July 2012, http://data.worldbank.org/indicator/NY.GDP.PCAP.KD.

WVS (2009). *World Values Survey 1981-2008 official aggregate v.20090902, 2009.* World Values Survey Association (http://www.worldvaluessurvey.org), Aggregate File Producer: ASEP/JDS Data Archive, Madrid, Spain.