THE UNIVERSITY of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

Essays on Political Economy

Safter Burak Darbaz

Doctor of Philosophy

School of Economics

University of Edinburgh

2015

To my mother Müren and my father Tufan who supported me until the very end.

Contents

Declaration

I hereby declare that this thesis has been composed by me and that the work contained in it is my own unless clearly stated and referenced. It has not been submitted for any other degree or professional qualification.

Safter Burak Darbaz

Acknowledgements

Abstract

This thesis consists of three stand-alone chapters studying theoretical models concerning a range of issues that take place within the context of political delegation: tax enforcement, political selection, electoral campaigning.

First chapter studies the problem of a small electorate of workers who cannot influence tax rates but can influence their local politicians to interfere with tax enforcement. It develops a two-candidate Downsian voting model where voters are productivity-heterogenous workers who supply labor to a local firm that can engage in costly tax evasion while facing an exogenously given payroll tax collected at the firm level. Two purely office motivated local politicians compete in a winner-takes-all election by offering fine reductions to take place if the firm gets caught evading. Two results stand out. First, equilibrium tax evasion is (weakly) increasing in the productivity of the median voter as a result of the latter demanding a weaker enforcement regime through more aggressive fine reductions. Second, if politicians were able to propose and commit on tax rates as well, then the enforcement process would be interference-free and the tax level would coincide with the median voter's optimal level. These two results underline the fact that from voters' perspective, influencing enforcement policy is an imperfect substitute for influencing tax policy in achieving an optimal redistribution scheme due to tax evasion being costly. In other words, a lax enforcement pattern in a given polity can be indicative of a political demand arising as an attempt to attain a redistributive second-best when influencing tax policy is not a possibility.

Second chapter turns attention to the role and incentives of media in the context of *ex ante* political selection, i.e. at the electoral participation level. It constructs a signalling model with pure adverse selection where a candidate whose quality is private information decides on whether to challenge an incumbent whose quality is common knowledge given an electorate composed of voters who are solely interested in electing the best politician. Electoral participation is costly and before the election, a benevolent media outlet which is assumed to be acting in the best interest of voters decides on whether to undertake a costly investigation that may or may not reveal challenger's quality and transmit this information to voters. The focus of the chapter is on studying the selection and incentive effects of changes in media's information technology. The setting creates a strategic interaction between challenger entry and media activity, which gives rise to two main results. First, an improvement in media's information technology, whether due to cost reductions or gains in investigative strength always

(weakly) improves *ex ante* selection by increasing minimum challenger quality in equilibrium. Second, while lower information costs always (weakly) make the media more active, an higher media strength may reduce its journalistic activity, especially if it is already strong. The intuition behind this asymmetry is simple. While both types of improvements increase media's expected net benefits from journalism, a boost to its investigative strength also makes the media more threatening for inferior challengers at a given level of journalistic activity. Combining this with the first result implies that the media can afford being more passive without undermining selection if it is sufficiently strong to begin with. In short, a strong media might lead to a relatively passive media, even though the media is "working as intended".

Third chapter is about electoral campaigns. More precisely, it is a theoretical investigation into one possible audience-related cause for diverging campaign structures of different candidates competing for the same office: state of political knowledge in an electorate. Electorate is assumed to consist of a continuum of voters heterogenous along two dimensions: policy preferences and political knowledge. The latter is assumed to partition the set of voters into ignorant and informed segments, with the former consisting of voters who are unable to condition their voting decisions on the policy dimension. Political competition takes place within a probabilistic voting setting with two candidates, but instead of costless policy proposals as in a standard probabilistic voting model, it revolves around campaigning. Electoral campaigning is modelled as a limited resource allocation problem between two activities: policy campaigning and valence campaigning. The former permits candidates to relocate from their initial policy positions (reputations or legacies), which are assumed to be at the opposing segments of the policy space (i.e. left and right). The latter allows them to generate universal support via a partisanship effect and can be interpreted as an investment into non-policy campaign content such as impressionistic advertising, recruitment of writers capable of producing emotionally appealing speeches, etc. The chapter has two central results. First, a candidate's resource allocation to valence campaigning increases with the fraction of ignorant voters, ideological (non-policy) heterogeneity of informed voters and proximity of candidate's initial position to the bliss point of the informed pseudo-swing voter.[1] The last one results from decreasing relative marginal returns for politicians from converging to pseudo-swing voter's ideal position. Second, even if candidates are otherwise symmetric,

---

[1] Pseudo-swing voter refers to the position of the voter who would be the (expected) swing voter if valence campaigning was not allowed and candidates were able to freely announce any policy.

a monotonic association between policy preferences and political knowledge can induce divergence into campaign structures. For instance, if ignorance and policy preferences are positively correlated (e.g. less educated preferring more public good) then the left candidate would conduct a campaign with a heavier valence focus and vice versa.[2] Underlying this result is again the decreasing relative marginal returns argument: a candidate whose initial position is already close to that of the informed pseudo-swing voter would benefit more from a valence oriented campaign. An implication of this is that a party that is known having a relatively more ignorant voter base can end up conducting a much more policy focused campaign compared to a party that is largely associated with politically aware voters.

---

[2]Left is utilized in a purely spatial sense, devoid of real life context.

x

Chapter 1

Political Economy of Tax Enforcement - Voting on Evasion Fines

Abstract

Motivated by frequent fine reductions, pardons and other tax enforcement issues widespread in many developing countries, this chapter studies under what conditions a political demand for a weak tax enforcement regime might arise in an electorate in the context of payroll taxation. For this purpose, I study a two-candidate Downsian policy competition model where candidates can promise to reduce evasion fines but cannot influence tax rates. It is shown that in an electorate composed of voters who differ with respect to their productivities, firm-level payroll tax enforcement becomes subject to political interference when the given tax policy diverges from the redistributive tastes of the median voter. It is also shown that the enforcement regime's credibility gap, i.e. the difference between *de jure* and *de facto* fine rates is increasing in the productivity of the median voter and if multidimensional voting was allowed, the political pressure on the enforcement regime would be completely eliminated and the tax rate would coincide with median voter's bliss point. The main point is that from the perspective of voters, enforcement policy is an imperfect substitute for tax policy. As long as the tax policy is non-negotiable, a political demand for a lax enforcement regime may arise as an attempt to attain a second-best outcome in terms of redistribution.

1 Introduction

1.1 Motivation and Summary

One of the persistent issues in developing countries is the prevalence of high tax evasion and weak tax enforcement. While tax collection issues also exist in developed countries, what distinguishes the former is the significant difficulties they

face in mobilizing revenues via direct taxes, which is reflected by the disproportional share of indirect taxes in their tax mixes.[1] A related pattern, especially relevant for countries in Latin America, Central and Eastern Europe (CEE) and Asia is the widespread under-collection of labor taxes.[2] According to International Labor Organization (ILO) estimates, between 20% to 30% of contributions remained uncollected in CEE countries throughout late 90's (Gillian et al.(2000)), while in some Latin American countries the rate of collection remained as low as 40% (Mesa-Lago (1998)). In another study, Saunders and Shang (2001) reports that nearly 70% of firms in Shanghai engage in some form of payroll tax evasion. Turkey is one country that should also be mentioned, as its unregistered employment rate is almost 40% (Karadeniz (2013)) and where until recently payroll tax collection was the exception rather than the rule (Bailey and Turner (2001)).

There are a variety of reasons leading to tax evasion but two necessary conditions should be satisfied if it were to occur in the first place. First, the potential punishments should not be fully deterrent, either due to their scale or credibility issues.[3] Second, there should be a positive probability of not getting caught. These are the twin pillars of tax enforcement: sticks and probability of sticks. The most usual suspects to lead to a weak enforcement regime would be cost-related issues and agency problems such as rent-seeking bureaucrats. Both can affect audit probabilities faced by tax evaders, as well as potential punishments. For instance, limited human resources can constrain the frequency of audits, or litigation costs can introduce a gap between *de jure* and *de facto* penalties. Similarly, corruptible bureaucrats can provide a limited shelter against audits to some firms or use discretionary powers to obtain them fine reductions in return for monetary gains. Normative or extractive considerations can also lead fiscal authorities to design optimal audit rules or penalty schemes responding to demographic differences, industry characteristics and firm behaviour, leading to regional or industry-wide differences in the strictness of enforcement policies. Another possibility, which have been largely ignored in the theoretical literature until recently, is political interference. Lobbying mechanisms can play an important role. But as Sandmo (2005) remarked, an electorate's resistance against stricter enforcement can also

---

[1]Two striking examples of this are Mexico and Turkey who raise nearly 60 percent of their revenues via indirect taxation, in contrast to an OECD average of 38 percent. This doesn't mean that they are good at collecting indirect taxes; just that they are less bad at it.

[2]By labor taxes, I mean earnings taxes collected at the employee level, i.e. payroll taxes. These usually consist of social security contributions, but they might also include other country-specific elements such as TV tax being collected as part of payroll taxes in Turkey.

[3]Even very high *de jure* punishments wouldn't be fully deterrent if the agents did not believe that they would be strictly carried out.

create political incentives for such interference and create favourable conditions for lax enforcement.[4] Indeed, casual observation suggests that tax enforcement and particularly the administration of penalties for tax evaders is quite sensitive to local election cycles in Turkey, where tax pardons and fine reductions are a customary part of its politics.

In this chapter, I develop a political economy model of labor tax enforcement where the strictness of the enforcement regime is determined by voting. There is a small electorate composed of workers with different productivities supplying labor and a large local firm who hires them. The firm pays payroll taxes at a level imposed by the central government, but it can evade some of these by incurring evasion costs and facing some probability of getting caught and paying fines in case of an audit. Two purely office motivated politicians compete in a winner-takes-all election by promising a stricter or a weaker enforcement regime. The way they do that is by making credible promises of fine reductions, or even tax pardons in the event that the firm gets caught, through leveraging their contacts with fiscal authorities. These can be interpreted as formal fine reductions, as well as informal ones obtained thanks to colluding bureaucrats misreporting the evaded amount. The enforcement regime that prevails from this picture accommodates the median voter's redistributive tastes, based on a trade-off between labor revenues and transfers.

The chapter has two main results. First, by voting on their preferred *de facto* penalty scheme, voters are essentially choosing an effective enforcement rate: a measure comprised of a *de facto* penalty rate and an *ex ante* probability of an audit, which is decreasing in the productivity of the median voter. This is a familiar redistribution story, where rich voters (ones with a higher labor productivity in this case) wish less taxation and poor wants more redistribution. It underlines enforcement policy as a taxation substitute which is used by voters to have a say in tax policy even if they don't have sufficient clout in the policy circles of the central government. Second, if voters were allowed to vote on both the tax level and the enforcement policy, then a fully deterrent enforcement regime would prevail and tax level would coincide with median voter's optimal choice under no evasion opportunity. As a corollary, political demand for lax enforcement is always associated with a divergence from voters' optimal taxation preferences. This

---

[4]The opposite can be true as well, especially if an electorate is divided between evaders and compliers and the latter is a majority. In a recent paper, Casaburi and Troiano (2013) show a positive causal relationship between reelection likelihood in local elections and the regional intensity of a nationwide anti-tax evasion intervention program in Italy, especially in areas with low tolerance for evasion.

underlines enforcement as an imperfect substitute for taxation and that if they could, voters would want to minimize deadweight losses associated with it due to costly evasion.

1.2 Previous Literature and Contribution

This chapter builds on the tax evasion literature when formulating the firm's evasion problem, so this is a good starting point. The seminal paper in this literature is Allingham and Sandmo (1972), who consider individual income tax evasion in a very simple setting inspired by Becker's (1968) classical paper on the economics of crime. They treat income as fixed and enforcement as a lottery characterized by a fixed pair of audit probability and penalty rate, and consider optimal reporting/evasion decisions of a risk-averse agent. Most of the subsequent literature builds on this framework, called the taxpayer-as-gambler (TAG) paradigm by Cowell (2004). Marelli (1984) is the first paper which extends the TAG framework to firm evasion. He studies simultaneous evasion and production decisions of a risk-averse monopolist facing an ad-valorem sales tax. Wang and Conant (1988) apply the same framework to corporate tax evasion, and Yaniv (1988) considers joint employment and evasion decisions of a perfectly competitive employer facing a payroll tax. All these papers show that firm activity is independent of the evasion decision (but not vice versa) as long as probability of audits are fixed, and that the separability breaks down when it becomes a function of reported activity. Yaniv (1995) later generalizes this result and show that irrespective of the form of the evaded (linear) tax, as long as the firm evades by underreporting its tax base or overreporting allowed deductions, separability result holds. The extension of this framework to the case of risk-neutral firms is first made by Virmani (1989), who considers evasion and output decisions of a risk-neutral firm under perfect competition in the presence of concealment costs, and an audit probability which is increasing in output. He shows that the previous separability result breaks down and both output and evasion decisions are jointly determined. Furthermore, firm's evasion response to increases in penalty rates become ambiguous. Cremer and Gahwari (1993) apply a similar framework to optimal commodity taxation. They show that if the audit probability is fixed, and concealment costs are separable in the evaded amount, then a one-sided separability result is recovered, with output depending on evasion but not vice versa. This allows obtaining clear comparative statics for firm's decision. This latter framework is later applied to different market and tax structures; for instance to profit taxation under monopoly by Eichhorn (2006), or oligopoly by Goerke and

Runkel (2006). It also constitutes the basis of firm decision in this chapter.

A somewhat related strand to the current chapter is the literature on endogenous enforcement. Kolm (1973) can be considered the initiator of this strand, where he assumes that fiscal authorities can spend resources to increase audit coverage and frequency. Subsequent papers gradually shift their focus to optimizing enforcement regimes by designing audit rules or penalty schedules making the best use of available information and behavioural patterns.[5] Sandmo (1981) considers welfare-maximizing proportional penalty and linear audit probability choices of a planner for the case of income tax with labor supplying risk averse agents. Reinganum and Wilde (1985), by limiting attention to cutoff audit rules as functions of liability declarations in a principle-agent framework, show that the latter weakly dominate fixed-probability audit rules if the principle aims to maximize expected tax collection. In another influential paper, using a mechanism design approach, Sánchez and Sobel (1993) show that when auditing is costly, the optimal audit rule for income tax is a function of tax rate and involves at most two cutoffs, dividing the distribution into at most three groups. In a more recent paper, Bayer and Cowell (2009) focus on the possible role of industry structure on audit rules. They show that in the context of corporate taxation and oligopolistic competition, a relative audit rule making use of prior information on the correlation between likelihood of evasion and size of industry-wide profit declarations reduces overall tax evasion relative to fixed-probability rules. There are also an array of papers aiming to endogenize other aspects of enforcement regimes. Several examples are Yitzhaki and Vakneen (1989), who study the optimal allocation of tax inspectors according to taxpayer complexity; Hashimzade, Huang and Myles (2010) who show that in case of VAT rebate fraud, the optimal penalty schedule should be convex in detected overstatement and Paramonova (2014) who studies the design of optimal information reporting schemes as part of an endogenous enforcement regime.

A more closely related strand is the literature on the political economy of redistribution in the presence of tax evasion. Borck (2004) is one of the first papers in this literature. Voters vote on a redistributive income tax and they each choose the fraction of their income to declare, taking audit probability and fine rate as given. He shows that under a weak enforcement regime, poorer workers can demand a higher taxation compared to no-evasion case as evasion can benefit rich voters disproportionately, which can hinder the progressivity of the tax

---

[5]A parallel literature that deals with optimal taxation when enforcement regimes are imperfect and evasion also exists.

system. Furthermore, he shows that an increase in enforcement frequency can actually increase evasion if this latter effect is too strong. Borck (2009) focuses on preference aggregation problems associated with tax evasion, showing how introduction of an evasion opportunity can cause cycling over tax policies, which would lead to non-existence of a majority winning tax level. Roine (2006) focuses on income tax avoidance instead.[6] Assuming a fixed cost of avoidance, he arrives at two conclusions. If the cost of avoidance is high, the standard conflict between rich and poor prevails. If the cost of avoidance is sufficiently low, then the very rich and the poor can coalesce to demand a high tax. Traxler (2009) deals with income tax evasion, but assumes that voters' marginal evasion costs depend on incomes. He shows that if marginal evasion costs are decreasing in income, then taxation preferences are characterized by a standard redistributive monotonicity. If they are increasing, taxation preferences can become non-monotonic, in the sense that redistribution occurs from middle class to rich and poor voters. Traxler (2012) studies welfare implications of avoidance opportunities in a standard voting model of linear income taxes.

Finally, there are two papers that deal with political determination of tax enforcement: Bárány (2009) and Besfamille, De Donder and Lozachmeur (2013). Bárány (2009) specifies a model of income tax evasion where voters are heterogenous along two dimensions: income and whether they have an evasion technology or not. Population is divided into three income groups and voters vote on tax and fine rates simultaneously, for a given audit probability. Using a probabilistic voting approach, she shows that a variety of enforcement-tax regimes can arise under different parameter configurations. These include "tax anarchies", where everyone who has the ability to do so evades, and equilibria in which nobody evades. One main difference from this chapter is that a demand for weak enforcement can coexist with voters having the power to influence tax policy. The reason for this is the fact that the opportunity to evade is a private good which not everybody possesses. As a result, there is not only redistribution between income groups, but also redistribution within groups, from non-evaders to evaders. Besfamille et al. (2013) present a structurally similar model to mine with an industrial organisation focus aiming to explain discrepancies in sales tax evasion across different industries. Voters are consumers heterogenous in their tastes for the good produced by firms and voting allows them to influence audit frequencies. Their results indicate that consumers who have stronger preferences for the good

---

[6]Avoidance can be defined roughly as legal evasion, without penalties involved. So it is not subject to enforcement.

in question demand a weaker enforcement regime, as stricter enforcement acts as an additional tax, which is subsequently reflected in firms' pricing decisions. One weakness of their paper is that the way they specify voter heterogeneity does not allow for multidimensional voting unless specific functional forms are imposed.[7] This clouds the connection between the given tax policy and enforcement demand, which is one of the main focus points of this chapter. Furthermore, voting decisions of a consumption-motivated electorate having no influence on tax rates driving differences in enforcement patterns across industries is a hard story to sell, unless transaction costs or other factors lead to a clear-cut market dispersion on a regional basis.

There are several contributions of this chapter. First, it provides a reasonably simple framework to think about the political channels which can influence cross-regional variations in tax enforcement and offers a possible explanation for frequently observed reductions and pardons of evasion penalties in developing countries, which can hinder the credibility, and thus the deterrence power of enforcement regimes. Second, it contributes to the tax enforcement literature by endogenizing the credibility of an enforcement regime within a political economy framework such that the strictness with which the penalties for payroll tax evasion are carried out is determined by political competition. It shows that in the context of redistributive taxation, a pure voting explanation for weak enforcement always requires some form of non-negotiability of the tax policy as long as evasion is a public good with externalities. In addition, despite the fact that possibility of binding constraints are usually avoided in models of redistributive voting, it is shown that the existence of Condorcet-winning policies is robust to them, under both single and multidimensional voting.

Next section presents the model. It starts by describing decision problems of economic actors taking enforcement and tax policies as given. This is followed by voting decisions of workers and the resulting enforcement policy outcome as well as a discussion on how this outcome would respond to changes in certain parameters. A subsection discussing how enforcement preferences of voters relate to the given tax policy follows this, which also includes the policy outcome that would prevail under multidimensional voting and the explicit relation between optimal tax and enforcement policies under two special cases. Next, a subsection considers normative implications of the policy outcome, which is then followed by a subsection dedicated to showing the robustness of previous results when it

---

[7]They find a similar full enforcement result under CES utilities and quadratic concealment costs.

is possible that a subset of workers exit the labor market under some policies. Final section concludes.

2 Model

The model consists of three stages. At the first stage, there is an election. For an exogenously given labor tax rate and a probability of tax audit, two identical local politicians with perfect commitment technologies compete in majoritarian elections by offering policy platforms on tax enforcement. These correspond to proportional evasion fines to be paid by the firm (to which the workers supply their labor) in case it gets caught evading. The interpretation is that while the electorate is too small to have sufficient political clout for influencing the centrally imposed tax rate, local politicians can influence *de facto* penalties by either obtaining tax pardons or fine reductions by leveraging their contacts with the fiscal authority, or influencing auditors so that they misreport the amount evaded in case the firm is audited.[8] Voters are workers having different productivities and the trade-off they face is between government transfers and labor revenues. In the second stage, given the election outcome, workers choose optimal amounts of leisure to consume and labor to supply. Given labor supply decisions of workers, there is a single risk-neutral firm with market power in the labor market (a monopsonist) who makes profit-maximizing hiring and evasion decisions.[9]

t=1	t=2	t=3
Elections. Winning politician implements enforcement policy.	Market outcomes. Workers supply labor. Firm hires labor and chooses evasion.	Revenue collection. Taxes and (if audited) fines. Redistribution.
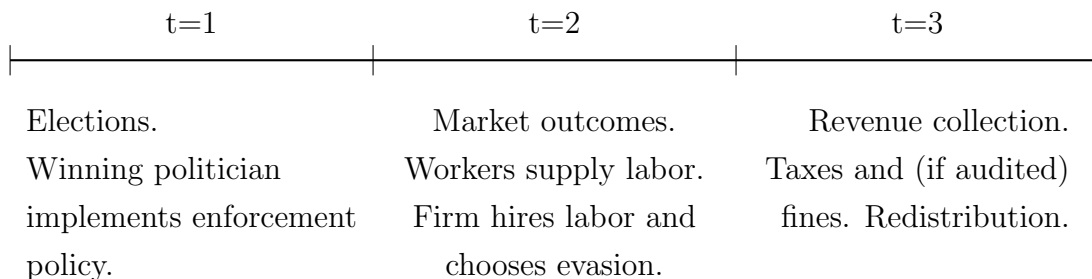
Figure 1: Timing and Structure.

---

[8]One may argue that politicians can achieve the same ends by offering policy platforms on *ex ante* tax breaks, but it is quite likely that there would be institutional obstacles for them to do so. For instance, while getting formal tax breaks might require a significant consensus within the fiscal authority, obtaining post-audit fine reductions on the evaded amount can possibly be achieved with a smaller base of influence. This is not only because the authority to grant formal fine reductions or pardons might have been left to the discretion of a smaller set of bureaucrats, but also because these can be obtained by informal means, e.g. making sure that the auditors misreport the amount evaded to their superiors.

[9]Model can easily be generalized to other forms of market structures with multiple firms without qualitatively altering any of the results.

This captures the big firm small town picture depicted in the introduction. In the final stage, an audit either occurs or does not occur according to an exogenously given probability distribution. Depending on the outcome, tax and fine revenues are collected and distributed to workers on an equal per capita basis. Figure 1 summarizes the timing and the general structure of the model. I start presenting the model with post election outcomes, as these will form the basis of how the voters will cast their votes.

2.1 Post Election

2.1.1 Government Revenues

At the final stage, tax and fine revenues are collected and redistribution occurs. I make two assumptions about auditing technology. First, it is assumed to be costless and reveals the evasion to auditors if it occurs. The former is not a substantial assumption, as even if there were costs associated with guaranteeing a certain frequency of audits and a non-negligible portion of these costs was reflected to the local electorate, there would likely be no additional costs for imposing higher penalties. So at the margin, these costs would be irrelevant. The latter is a simplification and if it were to be dropped, it would simply make voters *ceteris paribus* demand a stricter administration of penalties to achieve the same amount of evasion, which would not alter results ordinally. Second, I assume random auditing and proportional fines.[10]

There is a per-unit labor tax denoted by τ. Let α denote the fraction of employees not reported by the firm.[11] If no audit occurs, government revenues are given by

$$R_{na} = \tau(1 - \alpha)L. \tag{1}$$

If the firm is audited, government gets the following amount:

$$R_a = \tau(1 + \lambda\alpha)L, \tag{2}$$

where λ denotes the *de facto* proportional fine (so $\lambda = -1$ would imply either a full pardon for the evaded amount or that auditors engage in a complete cover-up). Let p_a denote the probability of an audit. Then *ex ante*, expected revenues

---

[10]Voting over more general non-linear penalty schemes to achieve the desired redistribution in the least distortion inducing way possible would certainly be interesting and possibly tractable under a probabilistic voting framework. I leave this possibility for future research.

[11]Strictly speaking, employee-hours not reported by the firm since as will be seen subsequently, L denotes the aggregate equilibrium labor hours supplied by the workers and hired by the firm. Henceforth, I will refer to L as employees and employee-hours interchangeably.

are given by:

$$\mathbb{E}[R] = p_a R_a + p_{na} R_{na}$$
$$= p_a \tau (1 + \lambda \alpha) L + (1 - p_a) \tau (1 - \alpha) L$$
$$= \tau (1 - (1 - e) \alpha) L = \tau_e L, \tag{3}$$

where $e = (1 + \lambda) p_a$ denotes the effective enforcement rate, and $\tau_e = \tau(1 - (1-e)\alpha)$ denotes the effective tax rate. Differentiating (3) with respect to τ and e by taking into account the fact that equilibrium tax evasion and employment will be functions of these allows to disaggregate the response of expected revenues to changes in tax and enforcement policies.

$$\frac{\partial \mathbb{E}[R(e, \tau)]}{\partial \tau} = \underbrace{(1 - (1-e)\alpha)L}_{+} + \underbrace{\tau(1 - (1-e)\alpha)\frac{\partial L}{\partial \tau}}_{-} - \underbrace{\tau(1-e)L\frac{\partial \alpha}{\partial \tau}}_{+} \tag{4}$$

$$\frac{\partial \mathbb{E}[R(e, \tau)]}{\partial e} = \underbrace{\tau L \left[\alpha - (1-e)\frac{\partial \alpha}{\partial e} \right]}_{+} + \underbrace{\tau(1 - (1-e)\alpha)\frac{\partial L}{\partial e}}_{-} \tag{5}$$

While an increase in tax level has a positive effect on expected revenues via increasing extraction from the already existing tax base, it also has a negative effect by reducing that base through two channels: decreased employment and increased evasion.[12] On the other hand, for a given tax level, a stricter enforcement policy will increase revenues not only because the firm will get caught more often and pay more fines and outstanding tax liabilities, but also because it will reduce the amount it evades due to having to pay higher fines in case of getting caught red-handed. Yet, it also has a decreasing effect via reducing employment as it leads to a higher effective taxation. For very low levels of tax ($\tau \to 0$), (4) should be positive and (5) should be zero. For very high levels of tax, the crowding-out effect will be too strong ($\lim_{\tau \to \infty} L(\tau) = 0$) and no revenues will be collected. This suggests a τ-Laffer curve for revenues, although it doesn't have to be single-peaked as this would ultimately depend on functional forms. On the other hand, as $e \to 1$, (5) will go to zero.[13] Furthermore, as $e \to 0$, evasion will be high (particularly for large values of τ), which suggests that in (5), the left term would dominate the right term. Yet, whether $\mathbb{E}[R]$ will be strictly increasing, or decreasing for some $e \in (0, 1)$ will ultimately depend on τ, as well as functional

---

[12] The signs of evasion and employment responses to policy changes follow from the discussion in the next subsection.

[13] This follows from the fact that as $e \to 1$, both $\alpha(e)$ and $\frac{\partial L}{\partial e}$ goes to zero. See the next subsection for the former and the appendix for the latter.

forms.

I now turn attention to the second stage, when the firm and workers make their hiring, evasion and labor supply decisions simultaneously.

2.1.2 The Firm

There is a single local firm with monopsony power in the labor market. It has no say in the policy process.[14] It produces a single good with its price normalized to 1 (numéraire good) to sell in a competitive national market using a single-input non-increasing returns technology $y(L)$. It takes an upward sloping inverse-labor supply function $w(L)$ and a per-employee (hour) tax τ as given and chooses the optimal number of employees to hire.[15] In addition, the firm can evade some of its tax liabilities by underreporting a fraction $\alpha \in [0, 1]$ of the labor it utilizes and get higher profits as long as it isn't audited. However, evasion is subject to a cost, which can be captured by a function $\tilde{c}(\alpha, L)$. Costly evasion might seem counterintuitive. For instance, if $\lambda = -1$, i.e. if the firm knows that it won't pay the evaded amount even if it gets caught, why would it engage in an effort to cover its tracks? It can be justified with the existence of a default cursory examination requiring the firm to undertake some concealment costs unless it is willing to face an audit with certainty without any scope for fine reductions. For instance, a large disparity between firm's output and its reported employment might raise a red flag with the central fiscal authority, grabbing the attention of a large set of people, which can make sure that the firm is audited and undermine the influence of local politicians on penalties.[16] For evasion costs, I adapt a common functional form from the indirect tax evasion literature popularized by Cramer & Gahwari (1993): $\tilde{c}(\alpha, L) = c(\alpha)\alpha L$, with $c(\cdot)$ being convex and increasing and $c(0) = c'(0) = 0$. Linearity of evasion costs in the evaded amount ensures a one-sided separability for the firm's problem, which leads to clear-cut comparative statics on hiring and evasion decisions.

Given evasion and hiring decisions of the firm, if no audit occurs, the firm's *ex post* profits can be written as follows.

$$\pi_{na} = y(L) - [w(L) + \tau(1 - \alpha) + c(\alpha)\alpha] L. \tag{6}$$

---

[14]So it cannot lobby and its profits are redistributed to a subset of voters with measure zero.

[15]It is possible to specify taxation as a fraction levied from wage payments but since wages won't differ across individuals (see next subsection), this wouldn't affect the results as long as the concealment cost function is modified slightly so that the one-sided separability result holds.

[16]So despite political interference, the firm might still need to do certain things "by the book". In any case, I provide an analysis with costless evasion later on.

If the firm is audited, its profits would reflect politically influenced fines it has to pay.

$$\pi_a = y(L) - [w(L) + \tau(1 + \lambda\alpha) + c(\alpha)\alpha] L. \tag{7}$$

The firm is risk-neutral, so *ex ante*, it will maximize its expected profits:

$$\max_{(L,\alpha)\in\mathbb{R}_+\times[0,1]} \mathbb{E}[\pi] = p_a\pi_a + p_{na}\pi_{na}$$

$$= y(L) - [w(L) + \tau] L - (c(\alpha) - (1 - e)\tau)\alpha L$$

$$= y(L) - [w(L) + \tau_e + c(\alpha)\alpha] L, \tag{8}$$

where τ_e and e are as defined previously.

Notice that if the firm is choosing a positive amount to evade, then it must be true that $c(\alpha)\alpha L < (1 - e)\tau\alpha L$. So even though evasion is costly, these costs will not dominate (expected) cost savings from paying less taxes. It follows that evasion opportunity, when combined with lax enforcement will translate into an increased employment via higher equilibrium wages. Lax enforcement is critical for this increase, as if $e \geq 1$, the firm would simply pay all of its tax liabilities.

Lemma 1 (Fully deterrent effective enforcement): If $e \geq 1$, then $\alpha = 0$.

Proof: For any $\alpha \in (0, 1]$ with $L > 0$, $c(\alpha)\alpha L - (1 - e)\tau\alpha L > c(\alpha)\alpha L > 0$ as long as $e \geq 1$. So the firm can increase its profits by setting $\alpha = 0$. ∎

Assuming continuity of $y(L)$, $w(L)$ and $c(\alpha)$, the problem defined in (8) has a solution, even though the feasible set of choice variables seems non-compact.

Proposition 1 (Existence of a profit maximizing choice): Firm's profit maximization problem has a solution under previous continuity assumptions.

Proof: With continuity assumptions, the objective function becomes continuous. Furthermore, there is an upper limit on L, given by the total time endowment in the economy. This limit can be expressed as $\overline{L} = \hat{\rho}$, where $\hat{\rho}$ is the average time-endowment (productivity) and the equality follows from the fact that population is of measure one (see next section). So the choice set is essentially compact, and the existence of a maximizing pair (L, α) follows from Weierstrass' theorem. ∎

Assuming that an interior solution exists, following equations describe first-order necessary conditions for it.

$$y'(L) = w'(L)L + w(L) + \tau + (c(\alpha) - (1 - e)\tau)\alpha, \tag{9}$$

$$c(\alpha) + c'(\alpha)\alpha = \tau(1 - e). \tag{10}$$

As long as $e < 1$, interior optima can be guaranteed by a sufficiently low τ, along with two border conditions, namely: $\lim_{L \to 0} [y'(L) - (w'(L)L + w(L))] > 0$ and $\lim_{L \to \overline{L}} [y'(L) - (w'(L)L + w(L))] < 0$. Henceforth, I also assume that the solution is unique.[17,18]

From (9) and the previous discussion, it can be immediately seen that firm's evasion response to a lax enforcement regime softens the impact of taxes on employment, resulting in an higher equilibrium wage then would prevail under no evasion opportunity at the same level of taxation. The following proposition gives the comparative statics for employment and evasion decisions.

Proposition 2 (Comparative statics for firm's choices): Assume that $w(\cdot), c(\cdot), y(\cdot)$ are twice continuously differentiable. Then at an interior maximum, firm choices are differentiable and the comparative statics are as follows.

$$\frac{\partial L}{\partial e} < 0 \quad ; \quad \frac{\partial L}{\partial \tau} < 0 \quad ; \quad \frac{\partial \alpha}{\partial e} < 0 \quad ; \quad \frac{\partial \alpha}{\partial \tau} > 0. \qquad (11)$$

Proof: Follows from an application of the implicit function theorem to (9) and (10). Details are provided in the appendix. ∎

These signs are intuitive. On the one hand, an increase in effective enforcement or tax level increases effective taxation, which increases marginal cost of the firm, leading it to hire less. On the other hand, while an increase in effective enforcement rate reduces evasion because it either reflects a higher probability of getting caught or a higher penalty in case of getting caught, an increase in tax level makes it more attractive by making evasion more cost-efficient.

2.1.3 Workers

Supply side of the labor market is populated by a continuum of workers with measure one, indexed according to their productivities $\rho \sim F$. F is a smooth distribution with support $[\rho, \overline{\rho}]$. Productivities are "time endowments", which the workers can allocate between labor and leisure.

---

[17]If the concavity and convexity conditions stated in the next footnote holds along with continuous differentiability of $y(\cdot), w(\cdot)$ and $c(\cdot)$, then (9) and (10) are also sufficient conditions for an interior global optima.

[18]Uniqueness is important but not crucial for the remainder of this chapter. It can be guaranteed by assuming strict concavity of $y(\cdot)$, strict convexity of $c(\cdot)$ and convexity of $w(\cdot)$. The latter requires imposing a restriction on the third-derivative of worker's utility. This would guarantee the strict concavity of the objective function and make sure of the existence of a unique maximizer. It is important because it ensures that at the election day, each policy is mapped to a unique outcome. It is not crucial because from Berge's Maximum Theorem, maximizers will be upper-hemicontinuous in policies so one can invoke a continuous ϵ-approximate selection theorem and base voting decisions on that selection.

Workers have quasilinear utilities, linear in the numéraire good with its price normalized to one. Quasilinearity is a convenience assumption. By eliminating income effects, it ensures an upward sloping labor supply and allows the incentives governing redistributive motives to be isolated down to individual productivities. As a corollary to the latter, it also helps with preference aggregation later on, making sure that individual preferences can be ranked according to gross incomes independently of the redistribution policy, a condition called Hierarchical Adherence by Roberts (1977). This is essentially Persson and Tabellini's (2000) take on the seminal Meltzer and Richard (1981) model of labor supply. More productive workers are not only better at earning labor income, but also have more leisure time to spend. Yet due to quasilinearity, everybody allocates the same amount of time to leisure, implying that more productive workers end up with higher labor revenues and thus more consumption. Uniformity of leisure across workers also helps with the political preference aggregation problem mentioned previously.

Following describes the choice problem faced by a worker with productivity ρ for a given wage w and expected (per capita) government transfers $\mathbb{E}[R]$.

$$\max_{(x,\ell,h)\in\mathbb{R}^3_+} u(x,h;\rho) = x + v(h)$$
$$\text{s.t. } x \leq w\ell + \mathbb{E}[R] \tag{12}$$
$$\ell + h \leq \rho,$$

where x denotes consumption of the numéraire good, h denotes leisure and ℓ denotes labor. The interpretation is that while labor supply and leisure enjoyment take place at the second period, consumption (or the consumption associated with public transfers) takes place at the third period. Due to risk-neutrality of the firm and quasilinearity of worker preferences, risk-attitude of workers and timing of numéraire consumption are irrelevant. Notice that the social transfer system functions on the principal of pure local financing. This is a simplification. One could assume that revenues are pooled and distributed according to some regional transfer function $g(\mathbb{E}[R])$. As long as $g(\cdot)$ is strictly increasing, qualitative conclusions of this chapter would remain intact.[19] Another simplification is the anonymity of workers, and thus irrelevance of who gets underreported. Within this framework, this is justified by the specification of productivity, which guar-

---

[19] Of course, if $g(0)$ is strictly positive and large enough, then the electorate might show political support for a very low effective taxation, i.e. very weak tax enforcement. This is likely to be an important source of cross-regional variation in the prevalence of tax evasion, especially in countries with high political interference to tax enforcement.

antees an equal hourly wage across all workers.[20]

Proposition 3 (Existence and uniqueness of a utility maximizing choice):
If $v(\cdot)$ is continuous, strictly increasing and strictly concave, then the problem in
(12) has a unique solution for each $(w, \mathbb{E}[R])$.

Proof: The objective is continuous and the constraint set is compact, so from
Weierstrass' theorem, a solution exists. Strict concavity of $v(\cdot)$ implies strict qua-
siconcavity of $u(\cdot)$, so the solution is unique. ∎

Since $v(\cdot)$ is strictly increasing and $u(\cdot)$ is strictly increasing in x, both constraints
in (12) will bind at the optimum. So I can rewrite worker's maximization problem
as follows.

$$\max_{h \in \mathbb{R}_+} \tilde{u}(h; \rho) = w\left[\rho - h\right] + \mathbb{E}[R] + v(h)$$

$$\text{s.t. } h \leq \rho. \tag{13}$$

There are two types of potential corner solutions associated with (13). First
one involves the worker spending all of her time working, second one involves her
not working at all.

$$v'(0) < w \implies h = 0 \text{ and } x = w\rho + \mathbb{E}[R]; \tag{14}$$

$$v'(\rho) > w \implies h = \rho \text{ and } x = \mathbb{E}[R]. \tag{15}$$

The first one of these is not very interesting and can easily be eliminated with the
sensible assumption $\lim_{h\to 0} v'(h) = \infty$. On the other hand, second one requires
more stringent assumptions to eliminate and is interesting because it implies that
some policy choices might lead to some low productivity workers being completely
dependent to the social transfer system.[21] For the remainder of this chapter up
until the end, I will ignore corner solutions as their existence do not alter the
qualitative conclusions. At the end of this section, I will return to them and
show that the political equilibrium is robust to possibility of "downward" corners

---

[20]Nevertheless, this would be an interesting extension, as it would not only open up a new
redistribution channel from formal workers to informal workers, but would also have the
potential to reverse the progressivity of the tax-transfer system. See the conclusion for a
discussion.

[21]One can assume $\lim_{h\to\underline{\rho}} v'(h) = 0$ to make sure no worker ever supplies zero labor, but
this would imply satiation; an assumption "hard to swallow". Alternatively, one can simply
assume that $\underline{\rho}$ and ρ^m (median's productivity) are high enough so nobody ends up at the
corner after elections. Answering the question "how high?" requires exact functional forms.
See Barnett et al. (2014) for an answer to the case with linear income taxes and isoelastic
utilities.

where some people do not supply any labor.

Assuming that the solution is interior, then it is characterized by the following first-order condition.

$$v'(h) = w. \tag{16}$$

Since $v(\cdot)$ is strictly concave, its derivative is strictly increasing and hence invertible. So the solution can be expressed as a function of the wage and transfers as follows.

$$
\begin{aligned}
h(w) &= v'^{-1}(w); \\
\ell_\rho(w) &= \rho - v'^{-1}(w); \\
x_\rho(w, \mathbb{E}[R]) &= w\left[\rho - v'^{-1}(w)\right] + \mathbb{E}[R].
\end{aligned}
\tag{17}
$$

Aggregate labor supply in the economy, as a function of w is in turn given by the following.

$$
\begin{aligned}
L(w) &= \int_{\underline{\rho}}^{\overline{\rho}} \ell_\rho(w)\, \mathrm{d}F(\rho) = \int_{\underline{\rho}}^{\overline{\rho}} (\rho - h(w))\, \mathrm{d}F(\rho) \\
&= \hat{\rho} - h(w),
\end{aligned}
\tag{18}
$$

where $\hat{\rho}$ is the average time endowment.[22] So aggregate labor supply is equal to aggregate labor endowment minus aggregate time spent on leisure. Next lemma shows that aggregate labor supply is upward sloping.

Lemma 2 (Upward monotonicity of the aggregate labor supply): $L(w)$ is continuous, differentiable and $L'(w) > 0$.

Proof: Continuity and differentiability follows from twice continuous differentiability of $v(\cdot)$, which carries over to $v^{-1}(\cdot)$. $L'(w)$ is given by the following.

$$
\begin{aligned}
L'(w) &= -h'(w) \\
&= -\frac{\mathrm{d}v'^{-1}(w)}{\mathrm{d}w} \\
&= -\frac{1}{v''(w)} > 0,
\end{aligned}
\tag{19}
$$

where the second equality follows from (17), third equality follows from the inverse function theorem and the inequality follows from strict concavity of $v(\cdot)$. ∎

Since $L(w)$ is strictly increasing, it is invertible and $w(L)$ is strictly increasing as

---

[22] And also the total time endowment, as the population is of measure one.

well. Given the inverse labor supply function, tax level and enforcement policy, the monopsonist determines the equilibrium employment as described previously. This will in turn determine social transfers.

This concludes the discussion of post-election outcomes, given the enforcement regime arising from electoral competition. In the next section, voters will declare policy preferences taking into account the outcomes described in this section, and politicians will offer policy platforms taking into account those preferences.

2.2 Election

2.2.1 Electoral Competition and Tax Enforcement

Since τ is set by the central government and outside the political reach of local politicians, voters' (workers') aim is to influence effective taxation via their local politicians' influence on tax enforcement. Given the random audit rule of the tax authority (p_a), enforcement is in turn influenced by obtaining credible concessions to the legally imposed penalty rate $(\overline{\lambda})$ either via authorized fine reductions or pardons, or through influencing auditors to collude and misreport the detected evasion. Without loss of generality, I assume that $\overline{e} = (1 + \overline{\lambda})p_a = 1$, i.e. enforcement policy is fully deterrent on the paper. From the discussion in the previous section, it should be clear that voting on λ is equivalent to voting on e.

Lemma 3 (Outcome equivalence of λ and e): Holding p_a fixed, choosing $\lambda \in \left[-1, \overline{\lambda}\right]$ is equivalent to choosing $e \in [0, 1]$.

Proof: Follows directly from the definition $e = (1 + \lambda)p_a$ and the fact that e is the relevant variable affecting firm's decision and revenues from (3) and (8). ∎

For any λ resulting from the political process, $\overline{\lambda} - \lambda$ should be interpreted as the "credibility gap" in the enforcement regime. Whereas $\lambda = -1$ corresponds to full pardon for evasion, $\lambda = \overline{\lambda}$ corresponds to no political interference, hence no credibility gap.

Political competition manifests in the form of two-candidate representative democracy. Politicians only care about winning the election and commitment is costless. It can be described in three stages.

1. Local politicians announce and commit to policies λ_A, λ_B (with corresponding e_A, e_B) respectively.

2. Voters pick the policy that gives them the highest utility.

3. Winning politician assumes the office and implements the announced policy.

This is a standard Downsian electoral competition framework. Whether this election game has a pure strategy Nash equilibrium will in turn depend on whether there is a Condorcet winning policy.[23]

Proposition 4 (Equilibrium existence in the Downsian policy competition game): If the set of Condorcet-winning policies $E^c \subset [0,1]$ is non-empty, and in case of a tie both politicians win with equal probability, then there exists a pure strategy Nash equilibrium where the candidates announce $e_A, e_B \in E^c$ and win with equal probability.

Proof: Suppose E^c is singleton and let $e^c \in E^c$. If $e_A \neq e^c$ and $e_B \neq e^c$, then one of the candidates announce e^c and win the election. If $e_A \neq e^c$ and $e_B = e^c$, then candidate A can announce e^c to increase its winning chance from 0 to $\frac{1}{2}$ and vice versa. If E^c is non-singleton, then with a similar reasoning, one concludes that any profile where politicians announce policies from E^c is a Nash equilibrium which results in them winning with equal probability.[24] ∎

Voters' policy calculus will be based on their indirect utilities, given the post-election outcomes discussed previously. That is, they will take into account that they will be optimizing their labor supplies after the election. Recognizing that these outcomes will depend on policies, the indirect utility of a voter with productivity ρ can be written as follows.

$$V(e, \tau; \rho) = \max_{h \in \mathbb{R}_+} \{ w(e, \tau) [\rho - h] + \mathbb{E}[R(e, \tau)] + v(h) \}$$
$$= w(e, \tau) [\rho - h(w(e, \tau))] + \mathbb{E}[R(e, \tau)] + v(h(w(e, \tau))), \quad (20)$$

where $w(e, \tau) \equiv w(L(e, \tau))$.[25]

If (20) is single-peaked in e, then workers' optimal policies (bliss points) under different productivities can be monotonically ordered, and any deviation to any direction from these optimal policies would decrease voter utilities monotonically. This would allow one to establish the existence of a unique Condorcet winning policy by a separation argument, which would coincide with the bliss point of the median voter. However, at this level of generality, (20) can easily fail to be single-

---

[23]A Condorcet winning policy is a stable policy in the sense that it beats every other alternative in pairwise voting with majority support, assuming that the voters are voting sincerely.

[24]If E^c is non-singleton, then their elements are sometimes called weak Condorcet-winners. By definition, when two weak Condorcet-winning policies are put against each other, they produce a tie. Equal chance tie-breaking can be achieved through independent randomization by voters with equal probabilities in case of indifference.

[25]Since I am assuming slack downward labor supply constraints at any policy choice, the constraint $h \leq \rho$ can be ignored.

peaked and instead, say, could be characterized by multiple local maxima.[26] In that case, deviations from bliss points would no longer decrease voter utilities monotonically, so it becomes a possibility that the bliss point of the median voter is beaten by another policy because that policy is not only closer to the global maxima of the voters at one side of the median, but is also closer to the local maxima of the voters at the other side of the median.[27] So instead, I will show that preferences satisfy single-crossing condition over the policy space $[0, 1]$. This is a property not of the individual preferences as in single-peakedness, but rather of the set of preferences, concerning how they are ordered with respect to each other and in particular, whether that ordering is monotonic.

Lemma 4 (Single-crossing and the existence of a Condorcet winning policy): Suppose that enforcement preferences satisfy the single-crossing condition for any given tax policy:

$$\text{If } e' > e \text{ and } \rho' > \rho \text{ or if } e' < e \text{ and } \rho' < \rho, \text{ then}$$
$$V(e', \tau; \rho') \geq V(e, \tau; \rho') \implies V(e', \tau; \rho) \geq V(e, \tau; \rho).$$

Then a Condorcet winning policy exists and coincides with an effective enforcement rate that maximizes indirect utility of the median voter.

Proof: If the above condition is satisfied, then letting $e(\rho^m)$ denote an optimal policy of the median voter, against any alternative $e > e(\rho^m)$, $e(\rho^m)$ is weakly preferred by all workers with $\rho \geq \rho^m$ and against any alternative $e < e(\rho^m)$, $e(\rho^m)$ is weakly preferred by all workers with $\rho \leq \rho^m$.[28] ∎

In other words, single-crossing requires that if a worker prefers a weaker (stricter) alternative when comparing two enforcement regimes, then all workers with higher (lower) productivities should prefer the weaker (stricter) alternative as well. This ensures that for any alternative against her optimal policy, median's type divides the productivity line into supporters and non-supporters.

In a seminal paper, Roberts (1977) shows that in the case of linear income taxation, single-crossing boils down to two conditions: increasing progressivity

---

[26]Why can that be the case? A worker might dislike a small increase in the effective tax level resulting from a small reduction in penalty rates due to concavity of the production function and the curvature of worker's utility combined with the convexity of evasion costs leading to a pronounced downward response in wages. Yet further reductions might be more palatable if they don't lead to large shifts in equilibrium employment.

[27]A median voter theorem, since it relies on a separation argument, always requires the unanimous support of either all the voters to the left or to the right of the median against any alternative.

[28]Because median is median, its right and left each constitute half of the population.

and hierarchical adherence. The first one refers to the set of feasible tax schedules being completely ordered with respect to Lorenz dominance and as shown by Hemming & Keen (1983), amounts to the requirement that after-tax incomes should be single-crossing in tax and gross incomes.[29] The second one refers to a property of voter heterogeneity, and requires that to be separable in a fashion ensuring that voters can be ranked according to their gross incomes independently of the redistributive policy. Here, workers are essentially voting over non-linear income tax schedules with the penalty rate playing the role of the taxation parameter, the form of non-linearity is induced by the shapes of leisure utility, production function and evasion cost, and gross incomes are nothing but labor incomes that would prevail under no enforcement (full pardon).

Gans and Smart (1996) show that insights of Roberts (1977) are true for any voting decision over sets of non-linear tax schedules and that they are equivalent to preferences satisfying the single-crossing condition in Lemma 4. So in essence, Lemma 4 is equivalent to enforcement regimes being rankable according to Lorenz dominance, and workers' gross labor incomes being rankable independently of the enforcement policy. Using Milgrom and Shannon's (1994) equivalence result between the single-crossing condition and the Spence-Mirrlees condition for cross-partial derivatives, Gans and Smart (1996) also provide a sufficient condition for Lemma 4 to hold. In the current model, the requirement is that marginal rates of substitution between effective enforcement rate and public transfers are monotonic in time endowments, i.e. indifference curves in $(e, \mathbb{E}[R])$ plane are single-crossing in ρ.
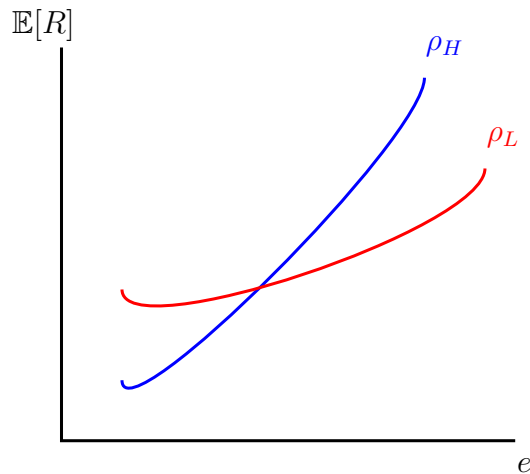
Figure 2: Increasing Marginal Rates of Substitution.

---

[29]This makes sure that for any two tax schedules, Lorenz curves do not cross.

Lemma 5 (Increasing marginal rates of substitution): Indirect utilities described in (20) satisfy Spence-Mirrlees condition of type-monotonic marginal rates of substitution in $(e, \mathbb{E}[R])$ plane for any τ.

Proof: Holding $\mathbb{E}[R]$ constant, $V(e, \tau; \rho)$ is differentiable in e at any τ and ρ, due to differentiability of $w(\cdot)$, $h(\cdot)$ and $v(\cdot)$. It is also differentiable in $\mathbb{E}[R]$. So using the envelope theorem, marginal rate of substitution between e and $\mathbb{E}[R]$ can be expressed as follows.

$$\text{MRS}_{e,\mathbb{E}[R]}(e, \tau; \rho) = -\frac{\partial V(e,\tau;\rho)/\partial e}{\partial V(e,\tau;\rho)/\partial \mathbb{E}[R]} = -\frac{\partial w(e,\tau)}{\partial e}\left[\rho - h(w(e,\tau))\right], \tag{21}$$

for $e < 1$. Differentiating (21) with respect to ρ, one gets the following.

$$\frac{\partial}{\partial \rho}\text{MRS}_{e,\mathbb{E}[R]}(e, \tau; \rho) = -\frac{\partial w(e,\tau)}{\partial e} > 0, \tag{22}$$

where the inequality follows from the fact that $\partial w(e,\tau)/\partial e = \frac{\partial w(L(e,\tau))}{\partial L}\frac{\partial L(e,\tau)}{\partial e} < 0$ from proposition 2 and lemma 2. If $e = 1$; $\alpha = 0$ so $\frac{\partial L(e,\tau)}{\partial e} = 0$ and thus the marginal rate of substitution between e and $\mathbb{E}[R]$ is zero for any ρ. ∎

All the discussions so far culminate in the following lemma.

Proposition 5 (Electoral outcome): Electoral competition has a pure-strategy Nash equilibrium, where both politicians announce and commit to the optimal penalty rate $\lambda^m = \frac{e^m}{p_a} - 1$ of the median voter and win with 50% chance.

The optimal penalty rate of the median voter will be pinned down by the enforcement policy that maximizes her indirect utility. Due to continuity of indirect utilities in e and compactness of the policy space, such an optimal policy exists due to Weierstrass' theorem. In case its not unique (multiple global maxima), any one of these policies can emerge from the electoral competition.

$$e^m \equiv e(\rho^m) = \underset{e \in [0,1]}{\arg\max}\, V(e, \tau; \rho^m). \tag{23}$$

Using the envelope theorem, median's optimal policy will be characterized by the following first-order condition as long as $e^m < 1$.

$$\frac{\partial w(e^m, \tau)}{\partial e}\left[\rho^m - h(w(e^m, \tau))\right] + \frac{\partial \mathbb{E}[R(e^m, \tau)]}{\partial e} \leq 0 \quad (= 0 \text{ if } e^m > 0). \tag{24}$$

Notice that any interior maximum requires the enforcement policy to be chosen at the increasing portion of the expected social transfer curve, because $\partial w(e^m, \rho)/\partial e <$

0. If such a portion did not exist than the solution would necessarily be at $e^m = 0$. A too high τ also makes $e^m = 0$ and therefore $\lambda^m = -1$ possible. In addition, $e^m = 1$, and thus $\lambda^m = \overline{\lambda}$ would be a possibility, if no $e \in [0, 1)$ were to satisfy (24) due to, say, τ being set at a too low level with respect to median voter's redistributive tastes.

So far, it has been established that any "credibility gap" in the enforcement regime arises due to political competition among opportunistic local politicians and that it can be tracked down to median worker's policy preferences. What determines those policy preferences? A first fact arises directly as a consequence of Lemma 3.

Lemma 6 (Increasing credibility gap in audit frequencies): An increase in the frequency of audits ($p_a \uparrow$) increases the credibility gap ($\overline{\lambda} - \lambda^m \uparrow$).

The more effort the central government puts into detecting evasion, the more aggressive voters become in pursuing *ex post* penalty reductions in order to achieve their preferred effective enforcement rates.

A second fact is ensued by lemma 5. Because marginal rates of substitution are increasing, a worker with a higher time endowment requires greater social transfers for a given amount of increase in effective taxation due to stricter enforcement. Put differently, as a worker becomes more productive, her marginal disutility from stricter enforcement (due to lower wages) increases. This should imply that the credibility gap is increasing in median worker's productivity.

Proposition 6 (Increasing credibility gap in median productivity): $\overline{\lambda} - \lambda^m$ is increasing in ρ^m.

Proof: Because $\partial V(e, \tau, \rho)/\partial \mathbb{E}[R] = 1$, (21) implies the following.

$$\frac{\partial^2 V(e, \tau; \rho)}{\partial \rho \partial e} = \frac{\partial w(e, \tau)}{\partial e} < 0, \tag{25}$$

for $e < 1$. From theorem 6 in Milgrom and Shannon (1994), this implies that $V(e, \tau; \rho)$ is supermodular in $(-\rho, e)$, which, from theorem 4 in Milgrom and Shannon (1994), implies that the optimal effective enforcement $e(\rho)$ is increasing in $-\rho$, i.e. decreasing in ρ. It follows that $\partial(\overline{\lambda} - \lambda^m)/\partial\rho^m > 0$.[30] ∎

In addition, if $e(\rho^m) = 0$ or $e(\rho^m) = 1$, then infinitesimal changes in the produc-

---

[30]If the problem in (23) have multiple global optima, then supermodularity implies that set of optimizers are decreasing in strong set order. That is, for $\rho > \rho'$ let $E(\rho)$, $E(\rho')$ denote the set of optimal enforcement policies for ρ and ρ'. Then $E(\rho) < E(\rho')$ in the sense that the smallest (largest) element in $E(\rho)$ is smaller than the smallest (largest) element in $E(\rho')$.

tivity of the median worker has no effect on the credibility gap. In the former case, all workers with $\rho > \rho^m$ prefers full pardon, whereas in the latter case, all workers with $\rho < \rho^m$ prefers a fully deterrent enforcement regime with no credibility gap.

At this point, a natural question arises: How does enforcement preferences respond to changes in the given tax level? Unfortunately, at this level of generality, it is impossible to answer this question with precision, neither using the implicit function theorem on (24), nor using a monotone comparative statics approach as in the above proposition because they both require signs on cross-partials. However, there is an inherent connection between the given tax policy and voters' enforcement demands due to the fact that they both serve the same redistributive purpose. Next subsection explores this idea further.

2.2.2 Tax policy and enforcement preferences

Consider a hypothetical situation in which local politicians can commit to setting tax levels and the firm has no evasion technology. Assuming that the increasing portion of the Laffer curve is active, a marginal increase in the tax level increases social transfers. This increase is financed in part by the firm, in part by the workers themselves, the extent of which will depend on the tax incidence.[31] Whatever the incidence is, "richer" workers (those having a higher time endowment) end up losing more relative to poorer workers, which makes redistribution less palatable for them. In the end, all workers have optimal tax levels reflecting their motivations to appropriate from the firm and the workers richer then themselves, and strength of that motivation depends on their marginal losses during this process, which in turn depends on their productivities.

Now consider the actual situation. For a given tax policy, voters can achieve more or less redistribution by calling for stricter or looser enforcement, and marginal changes in the penalty rate will have a similar effect to marginal changes in the tax level. However, two reasons suggest that having an influence on the enforcement regime is not a perfect substitute for having a direct influence on the tax policy. First, the highest effective taxation attainable by influencing enforcement is bounded above by the given tax level. This would be relevant if the labor tax was set at a level lower than median worker's preferred level. Second,

---

[31]More accurately, compared to the perfect competition, monopsonist bears the full tax burden (net of deadweight losses) because it optimizes w.r.t. wage. However, there is an excess burden due to increased deadweight loss. The extent to which this is borne by workers is captured by the marginal loss in labor revenues due to the reduction in wages. This loss reflects not only the relative elasticities as in perfect competition (a more inelastic labor supply implies a higher burden for workers), but also the curvature of the labor supply function.

for lax enforcement to reduce effective taxation, the firm has to engage in tax evasion by underreporting the number of its employees. Since this involves concealment costs, it creates an additional source for deadweight losses, which must be partially borne by the workers. This would be especially relevant if the given tax policy implied an excessive redistribution according to the median worker.

This reasoning hints that if voters were given a chance to influence both tax and enforcement policies, they would make sure to not bother themselves with extra costs associated with evasion and instead they would implement their preferred redistribution scheme by influencing the tax policy alone. This is indeed the case. Proving it requires two results. First one involves the nature of the policy trade-off faced by workers and shows the hegemony of wages and transfers when the worker is comparing two policies. It is proven by taking potential binding labor constraints into account as the result will also be useful in the last section. Second one is about multidimensional voting and makes sure that voting on both tax and enforcement policies lead to a cycling-free preference aggregation, which ensures equilibrium existence for the multidimensional Downsian policy competition game. Even though each individual voter might prefer complete elimination of costs associated with lax enforcement, median voter's preferred tax policy would differ from the rest. This difference would imply that some voters might prefer a policy pair characterized with a weaker enforcement regime in combination with the median voter's tax policy, despite the costs associated with it. So one cannot readily limit attention to unidimensional policies just by looking at individual preferences in isolation.

Proposition 7 (Unambiguously Preferred Policies): Consider a worker with productivity ρ and two policies β and β' at which the worker supplies non-zero labor ($\beta \equiv (e, \tau)$). If $w(\beta') > w(\beta)$ and $\mathbb{E}[R(\beta')] > \mathbb{E}[R(\beta)]$, then $V(\beta'; \rho) > V(\beta; \rho)$.

Proof: Appendix. ∎

While this result seems obvious, it should be kept in mind that a policy that yields both lower wage and transfers also results in higher leisure utility. Yet in the margin, this extra utility is nullified by the reduction in labor revenues due to lower labor supply, which is the essence of the (multidimensional) envelope theorem.

Next, I show that multidimensional voting does not lead to cycling and that the set of optimal policies for the median worker coincide with bidimensional Condorcet-winning policies.

Lemma 7 (Condorcet-winning bidimensional policies): Let β^m denote an optimal tax and enforcement pair for the median worker and β be an arbitrary suboptimal one. Then β^m gets majority support against β.

Proof: From the indirect utility of the median worker, if β^m is an optimal choice for the median, then the following inequality must hold.

$$v(h(\beta^m)) + \mathbb{E}[R(\beta^m)] + w(\beta^m)\left[\rho^m - h(\beta^m)\right]$$
$$> v(h(\beta)) + \mathbb{E}[R(\beta)] + w(\beta)\left[\rho^m - h(\beta)\right].$$

Rearranging this, one gets the following inequality.

$$\rho^m \gtrless \frac{v(h(\beta)) - v(h(\beta^m)) + \mathbb{E}[R(\beta)] - \mathbb{E}[R(\beta^m)] + w(\beta^m)h(\beta^m) - w(\beta)h(\beta)}{w(\beta^m) - w(\beta)}.$$

It follows that if $w(\beta^m) > w(\beta)$, then β^m wins with the unanimous support of the workers in $[\rho^m, \overline{\rho}]$ and if $w(\beta^m) < w(\beta)$, β^m wins with the unanimous support of the workers in $\left[\underline{\rho}, \rho^m\right]$. ∎

Lemma 7, along with proposition 4 implies that a two-policy election would result in commitment to median worker's optimal tax and enforcement pair. This is due to preferences being affine with respect to productivity, a condition called intermediate preferences by Grandmont (1978). In essence, this allows projecting the bidimensional policy conflict into a unidimensional space by summarizing it within a single measure: $\rho w(\beta)$. This measure is monotonically indexed by productivities, which allows for the separation argument necessary for establishing the existence of Condorcet-winners.

I now show that in any Condorcet-winning policy pair, the enforcement regime should be fully deterrent, i.e. there should be no credibility gap.

Proposition 8 (Fully deterrent enforcement regime under tax setting): Allowing politicians to commit on both the tax level and penalty rates results in a fully deterrent enforcement regime, along with a tax level optimal for the median voter.

Proof: From lemma 7 and proposition 4, the outcome of the election will be determined by the following problem.

$$\max_{(e,\tau) \in [0,1] \times \mathbb{R}_+} V(e, \tau; \rho^m).$$

This problem is equivalent to the following problem.

$$\max_{(c_e, \tau_e) \in \Gamma} \tilde{V}(c_e, \tau_e; \rho^m);$$

$$\Gamma = \Big\{ (c_e, \tau_e) : c_e = c(\alpha(e, \tau))\alpha(e, \tau) \ \&$$
$$\tau_e = \tau(1 - (1 - e)\alpha(e, \tau)) \ \& \tag{26}$$
$$(e, \tau) \in [0, 1] \times \mathbb{R}_+ \Big\}.$$

To see that this is indeed the case, just notice that any effect that a pair (e, τ) has on the indirect utility of a worker operates via its effect on $w(\cdot)$ and $\mathbb{E}[R(\cdot)]$. These are in turn determined by the following.

$$y'(L) = w(L) + w'(L)L + \tau_e + c_e, \tag{27}$$
$$\mathbb{E}[R] = \tau_e L. \tag{28}$$

Any pair (c_e, τ_e) uniquely pins down the labor demand (which determines $w(L)$), which in turn (along with τ_e) determines expected transfers. Yet worker is not unconstrained in her choice of (c_e, τ_e). First, there is a functional dependence between them which arises due to concealment costs and firm's optimal choice of evasion taking that cost into account. Second, given that functional relation, choice of (c_e, τ_e) is limited by the allowable range for (e, τ). These are all embedded into the constraint set Γ. Now for a given τ_e, any $c_e > 0$ would make the voter worse off by giving him both a lower wage and a lower expected revenue by reducing employment. This follows from proposition 7. It follows that the only case where the voter might potentially choose a $c_e > 0$ is if such choice was required for an otherwise infeasible τ_e. But there is no such τ_e, as any τ_e that is feasible with $c_e > 0$ can also be chosen by setting $c_e = 0$. This choice corresponds to setting $e = 1$, i.e. $\lambda = \bar{\lambda}$ and picking $\tau \in \mathbb{R}_+$ in an unrestricted fashion.

Once the choice of a fully deterrent enforcement regime is established, median worker's problem reduces to the following.

$$\max_{\tau \in \mathbb{R}_+} \tilde{V}(1, \tau; \rho^m). \tag{29}$$

The choice of τ will influence $w(\cdot)$ and $\mathbb{E}[R]$ via two equations.

$$y'(L) = w(L) + w'(L)L + \tau, \tag{30}$$
$$\mathbb{E}[R] = \tau L. \tag{31}$$

So the environment reduces to a standard political economy setting of redistributive taxation under no evasion opportunity and full tax collection. Although the

choice set in (29) is non-compact, it is possible to compactify it. Previously, it was assumed that $\lim_{L\to 0} [y'(L) - (w'(L)L + w(L))] > 0$. So let $\bar{\tau}$ be the tax level satisfying $\lim_{L\to 0} [y'(L) - (w'(L)L + w(L))] = \bar{\tau}$. Then the worker would never choose a $\tau > \bar{\tau}$, hence the choice set is effectively $[0, \bar{\tau}]$. So from Weierstrass' theorem, (29) has a solution. ∎

Proof of proposition 8 is in essence a revealed preference argument. For any effective taxation τ_e, $\tilde{V}(0, \tau_e) > \tilde{V}(c_e, \tau_e)$ should hold true for any $c_e > 0$. So if a policy pair (c_e, τ_e) with lax enforcement (and thus with positive concealment costs, corresponding to (c_e, τ_e)) was picked, $(0, \tau_e)$ should have been infeasible. But there is no such $(0, \tau_e)$, so full enforcement always prevails. From this result follows the next corollary.

Corollary 1 (Lax enforcement follows unpopular taxation): If tax level coincides with median voter's preferred level under a setting with no evasion, then fully deterrent enforcement prevails.

Therefore, a political demand for a weak enforcement regime is always associated with a divergence of tax policy from its no-evasion optimal level for the electorate. An argument analogous to the one made in proposition 6 establishes that this optimal level is decreasing in the productivity of the worker. It follows that there will almost always be some political demand for lax enforcement and the fate of the local credibility gap will hinge on whether such divergence occurs for the median voter.

Proposition 8 and its corollary also underlines penalty rate as an imperfect control for tax policy. Consider some worker with productivity ρ and let τ^* denote her optimal tax policy under no evasion. Suppose $\tau > \tau^*$. Relaxing the enforcement regime from full deterrence (reducing the penalty rate from $\bar{\lambda}$), allows the voter to close the gap between τ^* and τ_e.

$$\frac{\partial \tau_e}{\partial e} = \tau \left[\alpha - (1 - e) \frac{\partial \alpha}{\partial e} \right] > 0. \tag{32}$$

But this comes at a cost of partially internalizing the distortion associated with concealment costs. Although the effect of these costs never dominate the reduction in effective taxation, they still hinder it to an extent preventing the voter from reaching her optimal pair $(w(\tau^*), \mathbb{E}[R(\tau^*)])$.[32] So the voter's choice of *ex post* penalty rate for a given tax level can be thought as a constrained optimum

---

[32] If the worker sets the penalty rate so that $\tau_e = \tau^*$, then she would still obtain $\mathbb{E}[R(\tau^*)]$ but a lower wage due to $c_e > 0$.

for the underlying redistribution problem.

I now turn attention to the effect of a change in tax level to enforcement demand. Although the general case with increasing per-unit concealment costs remains elusive, it is possible to get functional-form free comparative statics under two special cases. This is done by exploiting the limited isomorphism between e and τ_e, which offers a tractable framework to map the changes in tax level to changes in the enforcement demand when the per-employee cost of underreporting $(c_e = c(\alpha)\alpha)$ can be held constant or is linear in α.

2.2.2.1 Costless Evasion

I start by rewriting the optimality condition governing firm's evasion decision.

$$c(\alpha) + c'(\alpha)\alpha = \tau(1 - e). \tag{33}$$

Suppose that $c(\alpha) = 0$ for all α. If $\tau > 0$ and $e < 1$ ($\lambda < \overline{\lambda}$), then $\alpha = 1$, i.e the firm evades all of its tax liabilities, and revenues are collected only if an audit occurs. Let $\tau^{m*} > 0$ denote median worker's preferred tax level under no evasion, i.e. her solution to the problem in (29). First, suppose that the given tax level is higher: $\tau > \tau^{m*}$. Then the worker would demand a weak enough enforcement regime to ensure that $\tau_e = \tau(1 - (1 - e)) = \tau^{m*}$, i.e. $e = \frac{\tau^{m*}}{\tau} < 1$. Next, suppose that the given tax level is lower: $\tau < \tau^{m*}$. First thing to notice is that the range of effective tax levels implementable by influencing the enforcement regime is $[0, \tau]$, as e is relaxed from 1 to 0. There are two possibilities. First, $\tilde{V}(\tau; \rho^m) > \tilde{V}(\tau_e; \rho^m)$ for any $\tau_e \in [0, \tau)$. In this case, $e = 1$. This is possible if $\tilde{V}(\cdot; \rho^m)$ is single peaked in τ or there are local maxima in $[0, \tau)$ but they all yield a lower indirect utility compared to τ. Second, it is possible that there is a local maximum $\tau^{m**} \in [0, \tau)$ which yields a higher indirect utility compared to τ. In this case, worker would relax the enforcement regime until she obtains $\tau_e = \tau^{m**}$, i.e. $e = \frac{\tau^{m**}}{\tau} < 1$. Now consider the effect of a small increase in τ to the enforcement demand e. If $e = 1$, then it has no effect. If $e < 1$, then either to reach the globally optimal tax policy, or to reach the second-best locally optimal tax policy, worker would further relax the enforcement regime. This discussion is summarized in the following lemma.

Lemma 8 (Costless evasion comparative statics for τ): If there are no concealment costs, then the credibility gap $\overline{\lambda} - \lambda^m$ is non-decreasing in τ (strictly increasing if $\lambda^m < \overline{\lambda}$).

So under costless evasion, local enforcement regime weakens in response to in-

creased taxation.

Next special case considers constant per-unit concealment costs and it still yields tractable comparative statics. The argument is similar, albeit a little more complicated.

2.2.2.2 Constant Average Cost of Evasion

Suppose that $c(\alpha) = k$ for all α. Then the firm will engage in full evasion if $\tau(1 - e) \geq k$ and no evasion otherwise. Again let $\tau^{m*} > 0$ denote the median worker's preferred tax level under no evasion and suppose $\tau \neq \tau^{m*}$. If $k > \tau$, then no matter how weak the worker sets the enforcement regime, she can't get the firm to evade. In that case, she will be indifferent towards any enforcement regime. Suppose instead $k \leq \tau$. First thing to notice is that $\tilde{V}(0, \tau^{m*}; \rho^m)$ is no longer obtainable because if the voter gets the firm to evade, she has to face the lower wage consequences of costly evasion. Second, the strength of enforcement regime that the worker can set is bounded above by $1 - \frac{k}{\tau}$. Any e above that, and the firm ceases to evade. So the range of effective taxation policy under reach by influencing the enforcement regime and nudging the firm towards evasion is $[0, \tau - k]$. Also notice that the voter would never set $\tau_e = \tau - k$ by letting the firm evade because that would be equivalent to taxing the firm at a rate $\tau_e + c_e = \tau$ and getting $\mathbb{E}[R] = (\tau - k)L$ in return. Instead, she could set a strong enforcement regime with $e > 1 - \frac{k}{\tau}$, making sure that evasion costs are zero, and hence the firm is still taxed at a rate τ but she is getting $\mathbb{E}[R] = \tau L$. Now the question is: is there a $\tau_e \in (0, \tau - k)$ such that the following holds?

$$\tilde{V}(k, \tau_e; \rho^m) > \tilde{V}(0, \tau; \rho^m). \tag{34}$$

If there is not, then the median worker demands an enforcement regime strong enough to prevent evasion, and $\partial e/\partial \tau = 0$. If there is, then denoting it by $\tau_k^{m**} \in (0, \tau - k)$, she will set $e = \tau_k^{m**}/\tau$, in which case $\partial e/\partial \tau < 0$. This discussion is summarized in the following lemma.

Lemma 9 (Constant cost evasion comparative statics for τ): If per-unit concealment costs are constant, then the credibility gap $\overline{\lambda} - \lambda^m$ is non-decreasing in τ (strictly increasing if $e^m < 1 - \frac{k}{\tau}$).

The intuition behind these two results is clear. As long as the median voter is picking a weak enforcement regime, she is declaring a preference towards less redistribution. In the second case, she is even willing to absorb deadweight losses associated with concealment costs for it. Since an increase in the given tax level

increases redistribution for the given enforcement policy, she reduces redistribution by decreasing the penalty rates. The second case is also useful to show the limited effectiveness of influencing the enforcement regime in compensating the value loss from being unable to decide on tax policy. The best deal the swing voter can get by doing so is:

$$\max \left\{ \tilde{V}(0, \tau; \rho^m), \max_{\tau_e \in [0, \tau - k]} \tilde{V}(k, \tau_e; \rho^m) \right\} \leq \tilde{V}(0, \tau^{m*}; \rho^m). \tag{35}$$

Unfortunately, the usefulness of this approach runs out of steam once $c(\alpha)$ start varying continuously with the fraction underreported. Intuitively, lemmas 8 and 9 should still be true in the general case as long as an increase in tax level translates into more redistribution. If the worker is making an interior choice for the enforcement regime, she is declaring a preference towards a lower redistribution compared to the one implied by the imposed tax policy under no evasion. For this, she is willing to internalize some of the costs associated with deadweight losses stemming from $c_e > 0$. Her choice also implies that locally, instead of a slightly higher redistribution and lower costs (a greater e), she prefers a slightly lower redistribution.[33] This can be true only if at that slightly higher redistribution, marginal utility from lower redistribution dominates the marginal disutility from higher deadweight losses. Now consider a slight increase in τ holding the enforcement choice constant. If this implies a slightly higher redistribution, then the voter should find herself in an analogous situation, willing to reduce redistribution by relaxing the enforcement regime slightly at the cost of incurring slightly higher deadweight losses. However, it is no longer clear whether an increase in τ implies a higher redistribution. Because voters are not choosing τ, one might as well have $\frac{\partial \mathbb{E}[R]}{\partial \tau} < 0$.

2.3 Normative Considerations

A typical exercise in spatial voting models of policy determination is to show how the electoral outcome diverges from the outcome that would be picked by a social planner. Take a standard social welfare function assigning equal weights to each worker and assume that instead of having a voting outcome, a planner sets the enforcement regime at the first stage by maximizing that welfare function.[34]

---

[33] $\frac{\partial \mathbb{E}[R]}{\partial e} > 0$ at the optimal enforcement choice and $\frac{\partial c_e}{\partial e} < 0$ because $\frac{\partial \alpha}{\partial e} < 0$.

[34] Note that producer surplus is not included in the welfare objective. This can be justified either by non-local ownership, or by assuming that profits are redistributed to a measure-zero set of residents.

$$\max_{e\in[0,1]} \left\{ v(h(e,\tau)) + \mathbb{E}[R(e,\tau)] + \int_{\underline{\rho}}^{\overline{\rho}} (\rho - h(e,\tau))w(e,\tau)\mathrm{d}F(\rho) \right\}. \qquad (36)$$

Assuming an interior solution, the planner's optimal enforcement policy should satisfy the following.

$$[\hat{\rho} - h(e,\tau))] \frac{\partial w(e,\tau)}{\partial e} + \frac{\partial \mathbb{E}[R(e,\tau)]}{\partial e} = 0, \qquad (37)$$

where $\hat{\rho}$ corresponds to the average worker's productivity. So relative to an egalitarian planner outcome, the voting outcome differs to the extent of the productivity difference between median and average workers. This is a standard result. From proposition 6, this implies that voting results in a weaker enforcement and a higher credibility gap under a left skewed distribution ($\rho^m > \hat{\rho}$) and a stricter enforcement and lower credibility gap under a right skewed distribution ($\rho^m < \hat{\rho}$) compared to (36).

However, both of these outcomes are constrained-Pareto efficient, as the median voter outcome coincides to a planner outcome where the planner simply maximizes a Benthamite social welfare function putting all the weight to the median worker. Instead, if the planner had more power, he could obtain a welfare improvement over both outcomes by implementing an unconstrained Pareto efficient outcome. To see this, just note that under any social weighting function $G(\cdot)$, and any given tax level $\tilde{\tau}$, planner's problem for optimal enforcement can be written as follows.

$$\max_{\substack{(e,\tau)\in[0,1]\times\mathbb{R}_+ \\ \text{s.t. } \tau=\tilde{\tau}}} \left\{ v(h(e,\tau)) + \mathbb{E}[R(e,\tau)] + \int_{\underline{\rho}}^{\overline{\rho}} (\rho - h(e,\tau))w(e,\tau)\mathrm{d}G(\rho) \right\}. \qquad (38)$$

Giving the power of tax-setting to the planner would be equivalent to dropping the constraint $\tau = \tilde{\tau}$, which would unambiguously increase (38), hence the name unconstrained Pareto efficient.[35] From proposition 8, we know that this increase would stem from the elimination of deadweight losses associated with concealment costs, as the solution would involve a fully deterrent enforcement regime. In fact, if the planner had even more power, he could have gone further by adopting a less distortionary form of taxation. For instance, one possibility would be to impose a neutral profit tax combined with a fully deterrent enforcement, and

---

[35]This improvement in social welfare due to a reallocation from a constrained Pareto efficient outcome to an unconstrained Pareto efficient outcome does not necessarily correspond to a Pareto improvement because some workers might end up worse off during the process. Yet, it is an efficiency gain in the sense that the outcome is no longer at the interior of the utility-possibilities set (when the set is defined by allowing both e and τ to vary), but on its frontier.

to redistribute the receipts according to the optimal trade-off between wages and expected transfers.[36]

I now turn attention to potentially binding non-negative labor supply constraints (equivalently, upward leisure constraints) and their implications for the preceding analysis.

2.4 Binding Constraints

For a given wage w, suppose that there exists some $\tilde{\rho} \in \left[\rho, \overline{\rho}\right]$ such that $v'(\tilde{\rho}) = w$. This is the indifferent worker whose leisure constraint $h(w) \leq \rho$ just binds and she supplies no labor at w, i.e. $\tilde{\rho} = h(w)$. Since $v(\cdot)$ is strictly concave, $v'(\cdot)$ is strictly decreasing. So any worker with productivity $\rho \in \left[\rho, \tilde{\rho}\right)$ have a binding constraint as well and therefore supplies no labor. This is because $\rho \in \left[\rho, \tilde{\rho}\right)$ implies $v'(\rho) > w$, which implies that the marginal benefit of leisure exceeds its marginal cost even though all the available time is spent on it. Taking this into account, the aggregate labor supply function is given below.

$$L(w) = \int_{h(w)}^{\overline{\rho}} \ell_{\rho}(w) \, \mathrm{d}F(\rho) = \int_{h(w)}^{\overline{\rho}} (\rho - h(w)) \, \mathrm{d}F(\rho)$$
$$= (1 - F(h(w))) \left\{ \mathbb{E}[\rho | h(w) < \rho \leq \overline{\rho}] - h(w) \right\}. \tag{39}$$

From lemma 2, we know that the strict concavity of $v(\cdot)$ implies $h'(w) < 0$. So an increase in workers' wages will increase their labor force participation. Furthermore, it is straightforward to show that the aggregate labor supply is strictly increasing.

Lemma 10 (Upward monotonicity of total labor supply with binding constraints): $L(w)$ is continuous, differentiable and $L'(w) > 0$ with binding constraints.

Proof: Continuity and differentiability follows from twice continuous differentiability of $v(\cdot)$, which carries over to $v^{-1}(\cdot)$ and smoothness of F. Using Leibniz rule, $L'(w)$ is given by the following.

$$L'(w) = -\int_{h(w)}^{\overline{\rho}} h'(w) \, \mathrm{d}F(\rho) - h'(w)(h(w) - h(w))f(h(w))$$
$$= -h'(w)(1 - F(h(w))) > 0, \tag{40}$$

where the sign in (40) follows from $h'(w) < 0$. ∎

Since $L(w)$ is strictly increasing, it is invertible and from the inverse function theorem, $w'(L) > 0$. The rest of the analysis covering post-election outcomes remains the same.[37]

At the election day, voters choose their optimal policies by maximizing indirect utilities as usual. But their maximization problems should now take into account the possibility that some policy choices will lead them to supply no labor after the elections.

$$\max_{e \in [0,1]} V(e, \tau; \rho) = \max_{e \in [0,1]} \left\{ \max_{0 < h \leq \rho} \{ w(e, \tau) [\rho - h] + \mathbb{E}[R(e, \tau)] + v(h) \} \right\}. \quad (41)$$

Proposition 9 (Existence of a maximum): Optimal policy choice problem described in (41) has a solution.

Proof: Utility function is continuous in (h, e, τ, ρ). Moreover, the correspondence $\rho \mapsto [0, \rho]$ is continuous and compact-valued, so from Berge's maximum theorem, $V(e, \tau; \rho)$ is continuous in (e, τ, ρ). Since the policy space is compact ($[0,1]$), the result follows from Weierstrass' theorem. ∎

For a given tax level τ, consider the worker with productivity ρ having the indirect utility described in (41). If $v'(\rho) < w(1, \tau)$, then this worker will supply a positive amount of labor for any $e \in [0,1]$. If, on the other hand, there exists some $e_\rho^c(\tau) \in [0,1]$ such that $v'(\rho) = w(e_\rho^c(\tau), \tau)$; then for any $e > e_\rho^c(\tau)$, worker will exit from the labor market, because proposition 2 is still valid and hence $w(\cdot)$ is decreasing in e.[38]

Lemma 11 (Comparative statics for the cutoff policy): Suppose that at the given tax level τ, there exists a cutoff policy $e_\rho^c(\tau) \in [0,1]$, above which the worker with productivity ρ exits the labor market. Then $e_\rho^c(\tau)$ is increasing in ρ and decreasing in τ.

Proof: If such cutoff policy exists, then the following equation should hold at that policy.

$$v'(\rho) - w(e, \tau) = 0. \quad (42)$$

---

[37]Guaranteeing a unique interior solution for the firm's problem becomes harder, as $h''(w) > 0$ (which implies $w''(L) > 0$ assuming no binding constraints), which is obtainable via a sign restriction to the third derivative of $v(\cdot)$ in the previous case is no longer enough. One also requires $H(h(w)) < \frac{h''(w)}{(h'(w))^2}$, where $H(\cdot)$ is the hazard rate of F. Yet, as before, this is not a necessary condition but a rather strong sufficient condition. Moreover, as long as the solution is interior, previous comparative statics results remain unchanged.

[38]If there exists such $e_\rho^c(\tau) \in [0,1]$, then there exists a corresponding $\lambda_\rho(\tau) \in [-1, \overline{\lambda}]$. Signs of the comparative statics in lemma 11 carry over to $\lambda_\rho(\tau)$.

From the implicit function theorem, (42) defines e as a function of ρ and τ on a small neighbourhood around (ρ, e, τ) and the comparative statics are given by:

$$\frac{\partial e_\rho^c(\tau)}{\partial \tau} = -\frac{-\partial w / \partial \tau}{-\partial w / \partial e} < 0,$$

$$\frac{\partial e_\rho^c(\tau)}{\partial \rho} = -\frac{v''(\rho)}{-\partial w / \partial e} > 0,$$

where the signs follow from strict concavity of v and proposition 2. ∎

A worker who is more productive requires a lower wage (and therefore a stricter enforcement) to exit the labor force. This results from the way the productivity is specified, which implicitly defines leisure as the "outside option", the marginal utility of which falls with productivity at the boundary. Furthermore, an increase in tax level reduces the prevailing wage. For any worker, this reduces the required wage reduction for exiting the labor force, which implies that workers would cease supplying labor at a weaker enforcement regime.

Assuming the existence of a cutoff policy $e_\rho^c(\tau)$, (41) can be rewritten to get a cleaner view of the incentives that might lead a voter to pick a policy that would lead to her exit from the labor market.

$$\max \left\{ \max_{e \in \left[0, e_\rho^c(\tau)\right]} \left\{ w(e, \tau) \left[\rho - h(w(e, \tau))\right] + \mathbb{E}[R(e, \tau)] + v(h(w(e, \tau))) \right\} , \right.$$

$$\left. \max_{e \in \left[e_\rho^c(\tau), 1\right]} \left\{ \mathbb{E}[R(e, \tau)] + v(\rho) \right\} \right\}. \tag{43}$$

First, notice that if the worker decides to pick a policy leaving her supplying zero labor, than that policy should correspond to the enforcement regime that is maximizing the expected transfers alone. This policy can be at $e = 1$, but if it is not in $[e_\rho(\tau), 1]$, then the worker should necessarily pick a slack policy.[39] Second, rewrite the first-order condition associated with the first component of (46), again using the envelope theorem.

$$\frac{\partial w(e, \tau)}{\partial e} \left[\rho - h(w(e, \tau))\right] + \frac{\partial \mathbb{E}[R(e, \tau)]}{\partial e} \leq 0 \quad (= 0 \text{ if } e^m > 0). \tag{44}$$

If (44) is strictly positive at every $e \in \left[0, e_\rho^c(\tau)\right)$, then the solution to (43) necessarily occurs at a binding policy in $\left[e_\rho^c(\tau), 1\right]$. If there exists an interior solution in $\left(0, e_\rho^c(\tau)\right)$, than that can indeed be the optimal policy, but it should be contrasted

---

[39]Henceforth, I refer to policies leading to positive (zero) labor supply as slack (binding) policies.

with the revenue maximizing policy in $\left[e_\rho^c(\tau), 1\right]$, if one exists in this set. In any case, the lower ρ is, the more likely that the voter would pick a binding policy. This is due to Lemma 11, which implies a smaller range for e before worker's constraint starts binding, and the fact that $\partial w/\partial e < 0$ getting multiplied by a smaller number leaves more scope for (44) being positive. This can be seen as the logical continuation of the previous result, with higher productivity workers being more sensitive towards their wages, and lower productivity workers caring more about transfers.

Under potentially binding constraints, the existence of a Condorcet-winning policy becomes slightly more involved. To see why this is the case, notice that the indirect utility of any worker might become non-differentiable with respect to e at the policy where her leisure constraint starts binding ($e = e_\rho^c(\tau)$). This is because that policy would likely correspond to a kink in her value function. This implies that the argument of Gans and Smart (1996), which relies on ordering marginal rates of substitution at each policy to establish the single-crossing property, runs into technical problems. At any policy, if the set of voters supplying zero labor is of non-zero measure, then not only the marginal rate of substitution will be undefined for the indifferent worker, but marginal rates of substitution will jump as one moves to a slightly lower productivity. Now even if the median worker doesn't pick a binding policy, the policy she picks can leave other voters outside the labor market. Even if this were not the case, a Condorcet-victory requires median's optimal policy to beat every other policy, including policies possibly binding for others. So at the heart of the complication lies the potential need to compare binding policies against slack policies and vice versa. The following proposition shows that despite this complication, median worker's optimal policy is a Condorcet-winning policy and thus constitute an electoral outcome.

Proposition 10 (Unidimensional MVT with binding constraints): Let e^m denote an optimal effective enforcement rate for the median worker. Then e^m is a Condorcet winning policy.

Proof: Appendix. ∎

Although the proof is long, its intuition is simple. Any policy that leads to some workers exiting the labor market should divide the productivity interval into two connected parts, with the left part composed of workers supplying no labor and vice versa. This threshold is increasing in enforcement rate, because a stricter enforcement regime leads to a lower wage, which causes more workers to exit the labor market.
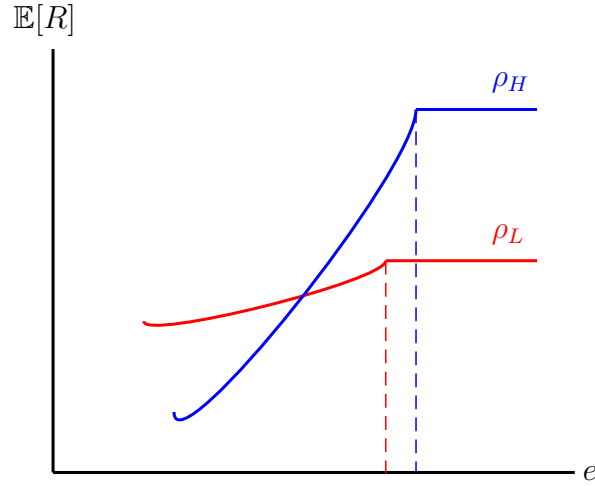
Figure 3: Single-Crossing Indifference Curves with Binding Constraints.

Thus for any two workers, not only the more productive worker will have a higher marginal rate of substitution at any policy where they both supply a positive amount of labor, but will also require a stricter enforcement before exiting the labor market. This makes sure that kinks are also ordered in a monotonically increasing fashion, which ensures that the single-crossing condition is not violated. Figure 3 provides a demonstration. In fact, a similar argument also holds for the "upward" constraints where workers spend no time in leisure and only work.[40]
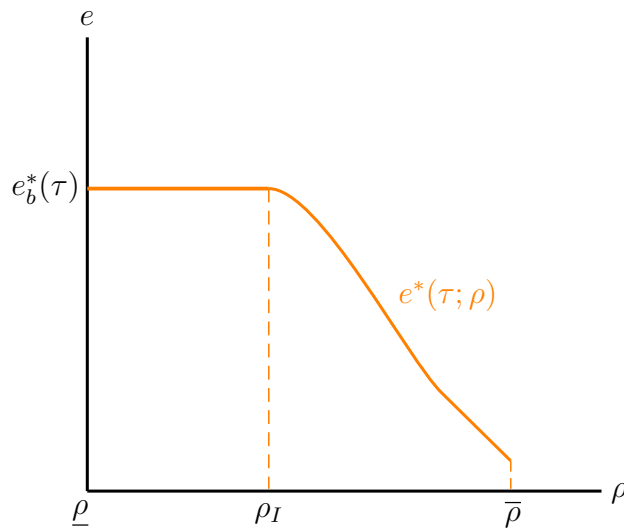
Figure 4: Optimal enforcement policies with binding constraints.

Proof of proposition 10 reveals some additional facts. First, if a worker with productivity $\tilde{\rho}$ decides to pick a binding policy, then all workers with $\rho < \tilde{\rho}$ pick

---

[40]In that case, zero-leisure kinks would occur at the same policy for all workers.

binding policies. On the other hand, if $\tilde{\rho}$ picks a slack policy, then all workers with $\rho > \tilde{\rho}$ pick slack policies. Furthermore, if a worker picks a binding policy, then she is picking the policy that maximizes expected transfers. On the contrary, optimal slack policy is type-dependent and strictly decreasing in productivity as long as it is interior. The missing link in this picture is the ordering between binding and slack policies. Proposition 11 shows that the optimal binding policy should be no less than optimal slack policies and figure 4 depicts preferred enforcement rate as a function of labor productivity.

Proposition 11 (Anatomy of binding and slack policies): Assume a unique solution for the policy problem and let $e_b^*(\tau)$ denote the policy that maximizes expected transfers. A worker with ρ finds $e_b^*(\tau)$ optimal only if it is a binding policy for her and if she finds a binding policy optimal, it should be $e_b^*(\tau)$. Let ρ_I denote the highest productivity worker who finds it optimal to pick a binding policy. Then ρ_I separates $[\underline{\rho}, \overline{\rho}]$ into two connected parts, with $\rho \leq \rho_I$ picking $e_b^*(\tau)$ and $\rho > \rho_I$ picking slack policies. Furthermore, letting $e(\tau; \rho)$ denote the optimal policy of a slack policy picker, $e_b^*(\tau) \geq \sup_{\rho \in (\rho_I, \overline{\rho}]} e(\tau; \rho)$. Moreover, $\partial e(\tau; \rho) / \partial \rho < 0$ for any $\rho \in (\rho_I, \overline{\rho}]$. Finally, $e_b^*(\tau) = e_{\rho_I}^c(\tau)$.

Proof: Appendix. ∎

It follows that the marginal effect of a productivity increase on the credibility gap will critically depend on the productivity of the median worker, specifically on whether she prefers a binding or a slack policy. Notice that $e_b^*(\tau) = 1$ is possible (and will be the case if expected transfers are single peaked in e), in which case if the median had sufficiently low productivity, enforcement regime would be completely free of political interference.

Corollary 2 (Upward type-monotonicity of the credibility gap with binding constraints): $\overline{\lambda} - \lambda^m$ is non-decreasing in ρ^m. More specifically, let $\rho_I(\tau)$ be the highest type that would pick a binding policy. If $\rho^m < \rho_I(\tau)$, then $\frac{\partial(\overline{\lambda} - \lambda^m)}{\partial \rho^m} = 0$ and if $\rho^m > \rho_I(\tau)$, then $\frac{\partial(\overline{\lambda} - \lambda^m)}{\partial \rho^m} > 0$.

Corollary 2 is just an extension of proposition 6. In addition, proposition 8 also holds under potentially binding constraints. To see this, first note that neither the proof of proposition 7, nor the parts of the proof of proposition 8 establishing full-deterrence assumed all-slack policies. The only part which assumed labor force participation by all workers under any policy was lemma 7, which established the existence of a Condorcet-winning policy pair under multidimensional voting.

Incorporating the constraint into the indirect utility of the worker, a first look might suggest that multidimensional preference aggregation breaks down under potentially binding policies.

$$V(e, \tau; \rho) = v(\min\{\rho, h(e, \tau)\}) + \mathbb{E}[R(e, \tau)] + w(e, \tau)(\rho - \min\{\rho, h(e, \tau)\}).$$
(45)

As can be seen from (45), indirect utility is no longer affine in ρ. Yet, as the following proposition shows, a multidimensional Condorcet-winner exists and corresponds to the optimal choice of the median worker.

Proposition 12 (Bidimensional MVT with binding constraints): Let (e^m, τ^m) denote an optimal effective policy pair for the median worker. Then it is a Condorcet winning policy and $e^m = 1$.

Proof: Proposition 8 and appendix. ∎

The proof is similar to the proof of proposition 10, with few differences arising from the fact that a marginal rates of substitution ordering argument in the differentiable portions of value functions can no longer be used.[41] The intuition behind the result is slightly more convoluted, but not much more. If one limits attention to the set of slack policies, then (45) has the intermediacy property, which still allows projecting the bidimensional conflict into a unidimensional space by summarizing it within a single measure: $\rho w(e, \tau)$. Holding w constant, this measure is monotonically indexed in ρ. On the other hand, for each given w, decisions to pick a binding policy are summarized within a single measure that also varies monotonically with ρ: $\rho - h(w)$. Finally, holding ρ constant and varying (e, τ) (the effect of which propagates only through the equilibrium wage) there is a monotonic relationship between these two measures. This latter monotonicity makes sure that these two measures act as a single monotonically ordered (in ρ) measure, which allows projecting the entire policy conflict into a single dimensional space and invoking the necessary separation argument.

It follows that all of the main results are valid under potentially binding constraints, and that when the voters are given the power to vote over tax policy, they still choose to eliminate deadweight losses associated with concealment costs. They then choose their optimal tax level, which can be high enough to leave some people (including themselves) outside the labor force, living purely off social transfers.

---

[41]In fact proposition 12 implies proposition 10. Yet the proof of proposition 10 itself has the additional benefit of implying unidimensional type-monotonicity of enforcement choices.

3 Conclusion

This chapter developed a pre-election politics model of payroll tax enforcement where the strength of the enforcement regime is undermined by local politicians responding to their electorate's redistributive demands by interfering with the post audit penalty administration process in return for electoral support. The motivation was to construct a political economy framework to explain the prevalence of weak tax enforcement regimes, especially in developing countries, and the accompanying patterns of widespread evasion and informality. The proposed mechanism that weakens the enforcement regime is the inability of the fiscal authority to commit to not granting fine reductions or even tax pardons, which nudges the firm towards more evasion by shortening the stick that it would face in the event it gets caught. This is a common phenomenon in developing countries such as Turkey for which the model provides a potential explanation if one infers that the pre-election pervasiveness of such fine reductions and pardons is associated with a political bargaining process akin to the one described in this chapter. Although the model is built around payroll taxes, its essence would remain intact if one were to consider an alternative form of firm-level taxation, as long as voters' livelihoods were at stake and they were facing a similar redistributive trade-off.[42]

Several conclusions have arisen from the model. First, voter demand for lax enforcement, and therefore equilibrium tax evasion is increasing in the productivity of the median voter. This is the familiar "richer voters demand less distribution" story in disguise. A strengthening of the enforcement regime corresponds to a higher effective taxation on workers' labor revenues, therefore those who supply more effective labor time desires less of it. The implication is that assuming a common tax level, enforcement regimes faced by firms in polities where gross labor earnings are higher should be subject to higher credibility gaps, as measured by the difference between *de jure* and *de facto* fine rates. Second, electoral demand for a weak enforcement regime always arises as a political response to not being able to influence tax policy. If local politicians were in fact able to commit on both tax and enforcement policies, then voters would always pick a fully deterrent enforcement regime and no credibility gap would occur in political equilibrium. Third, from the perspective of the voters, enforcement policy is an imperfect control for implementing an optimal redistribution scheme. Not only

---

[42]The first version of this chapter was built around a non-neutral profit tax and the results were similar.

it can't be used to increase the effective taxation above the imposed level and therefore might constrain some voters to a lower redistribution then they would prefer, it also comes with deadweight losses in case the firm is undertaking a costly effort to conceal its evasion. Fourth, even though the generality of the model prevents an inference regarding the effect of an increase in the given tax level on the enforcement demand, an argument is made that enforcement demand should be usually non-decreasing in tax level, and it is shown that this is indeed the case under two special cases. While this might seem counterintuitive at first, particularly if the tax was set at a level implying a lower-than-optimal redistribution for the worker, it should be kept in mind that as long as the voter prefers a weak enforcement regime to a fully deterrent enforcement regime, she is declaring a preference towards less redistribution than implied by the given tax level. It follows that a tax increase, particularly if it implies more redistribution, should imply an even weaker enforcement preference. Finally, it is shown that the results of the model are robust to allowing for the possibility of workers preferring policies that would lead some of them to exit the labor market. Corner solutions are usually avoided in models of redistributive voting.

Policy inferences from models at this level of generality should always be taken with a grain of salt. Taking the story of the model literally would lead one to conclude that the best way to solve the problem of politically-weakened enforcement would be to take the path of fiscal disintegration, allowing each polity to set its own tax policy. However, to come up with such policy recommendation, one needs to reconsider the problem in a broader framework, integrating wider welfare implications of interregional transfers, spillovers etc. A case can also be made for stronger centralization of enforcement-related services, as it can be easier for local politicians to influence the enforcement policy when its administration is undertaken by a local agency. One can even consider the establishment of an autonomous tax collection agency subject to some performance metric.[43] Again, these claims should be scrutinized in the context of more comprehensive models (incorporating agency elements in this case), besides being investigated empirically.

There are several promising ways to extend the analysis in this chapter. First, payroll tax evasion usually involves a collusive agreement between an employee and an employer. The reason for that is, especially in the case of social security contributions, the identity of the employee whose employment was underreported

---

[43]This is one of the standard inclusions to IMF's blueprint of recommendations to developing countries. See Crandall (2010).

matters. This chapter got around that difficulty by assuming anonymous workers and equal per capita social transfers. While the second assumption is not very substantial as long as the tax-transfer system is progressive, the first one can be important because dropping it would imply the emergence of a redistribution channel from formal workers to informal ones. One way of thinking about this involves assigning different hourly wages to different workers by specifying a different form of productivity, as well as allocating differing outside opportunities. Next step can be to consider incentive-compatible (for workers) underreporting strategies for the firm, e.g. underreporting below a certain productivity level. An interesting possibility to consider would be the informality of workers with relatively high productivities due to, say, some form of an incentive compatible rent-sharing agreement between the firm and workers. This has the potential of reversing the progressivity of the tax system, and breaking the limited isomorphism between enforcement and tax level. Second, the formal burden of payroll taxes are usually distributed among workers and firms according to legally set rates and the idea that the legal apportionment of tax burden having an effect on the enforcement preferences of voters has intuitive appeal. Yet, the model at its current form would not leave any room for such an effect to operate because the incidence is fully determined by the shapes of utilities and the production function. One way of assessing the potential impact of this would be to introduce bargaining elements in the process of wage determination.[44] Third, lobbying can be an important source of political interference and its effect can act in conjunction with that of the voting. Thinking in very simple terms, a firm such as the one in this chapter would always want the weakest possible enforcement regime. So if it were allowed to effectively lobby, this would introduce a further upward pressure on the credibility gap. Finally, the model in this chapter studied a single electorate in isolation and it was assumed that local politicians were able to exercise a perfect influence on the fiscal authority. It is usually the case that different electoral districts carry different clout levels with the central government. This would be an important factor in limiting local politicians' ability to interfere with the enforcement regime, if such interference required a certain degree of collusion with the central fiscal authority. Taking this into account would not change much if the focus remained on a single local electorate, but it would be an additional explanation for cross-regional enforcement variations. A more interesting approach would be to consider the problem of an incumbent national party who cares about maximizing the number of local electoral victories by either its own candidates

---

[44]See for instance, Chae (2002)

or candidates who are somehow more aligned with it. The incumbent would be limited in its ability to arrange such collusion to the extent of its hold in the bureaucracy and by the strength of checks and balances so it would have to behave selectively. This can, for instance, imply that swing polities, *ceteris paribus*, are characterized by weaker enforcement regimes. These possible extensions are left for future research.

Appendix

Proof of Proposition 2 (Comparative statics for firm's choices): Due to continuous differentiability assumptions; the implicit function theorem implies that (9) and (10) define two continuously differentiable functions $L(e, \tau)$ and $e(e, \tau)$ on an open neighbourhood around the given (e, τ). Totally differentiating the system in (9) and (10) and using the first-order conditions, one obtains the following.

$$
\begin{bmatrix} y''(L) - (2w'(L) + w''(L)L) & 0 \\ 0 & 2c'(\alpha) + c''(\alpha)\alpha \end{bmatrix} \begin{bmatrix} \mathrm{d}L \\ \mathrm{d}e \end{bmatrix}
$$
$$
= \begin{bmatrix} (1 - (1 - e)\alpha)\,\mathrm{d}\tau + \tau\alpha\,\mathrm{d}e \\ (1 - e)\,\mathrm{d}\tau - \tau\,\mathrm{d}e \end{bmatrix}. \tag{46}
$$

A locally unique interior maximum implies $y''(L) - (2w'(L) + w''(L)L) < 0$ and $2c'(\alpha) + c''(\alpha)\alpha > 0$. Solving the above system using Cramer's rule yields the following:

$$
\frac{\partial L}{\partial \tau} = \frac{(1 - (1 - e)\alpha)}{y''(L) - (2w'(L) + w''(L)L)} < 0 \;; \quad \frac{\partial L}{\partial e} = \frac{\tau\alpha}{y''(L) - (2w'(L) + w''(L)L)} < 0 \;;
$$
$$
\frac{\partial \alpha}{\partial \tau} = \frac{(1 - e)}{2c'(\alpha) + c''(\alpha)\alpha} > 0 \;; \quad \frac{\partial \alpha}{\partial e} = \frac{-\tau}{2c'(\alpha) + c''(\alpha)\alpha} < 0, \tag{47}
$$

where the signs follow straightforwardly. ∎

Proof of Proposition 7 (Unambiguously Preferred Policies): Think of β as a pair of enforcement policy and tax level: (e, τ). The total change in the indirect utility of the worker with productivity ρ resulting from a policy change of β to β' can be expressed as follows.

$$
V(\beta'; \rho) - V(\beta; \rho) = \int_\chi \nabla V(\tilde{\beta}) \cdot \mathrm{d}\tilde{\beta}
$$

$$= \int_\chi \ell_\rho(\tilde{\beta}) \nabla w(\tilde{\beta}) \cdot \mathrm{d}\tilde{\beta} + \int_\chi \nabla \mathbb{E}[R(\tilde{\beta})] \cdot \mathrm{d}\tilde{\beta}, \qquad (48)$$

where the first line follows from the fundamental theorem of line integrals and the second line follows from the envelope theorem. χ represents a smooth oriented curve starting at β and ending at β'. Now let $\kappa : [a, b] \to \chi$ denote an arbitrary bijective parametrization of χ with $\kappa(a) = \beta$ and $\kappa(b) = \beta'$. (48) can be rewritten as follows.

$$\int_a^b \ell_\rho(\kappa(t)) \nabla w(\kappa(t)) \cdot \kappa'(t) \mathrm{d}t + \int_\chi \nabla \mathbb{E}[R(\tilde{\beta})] \cdot \mathrm{d}\tilde{\beta}. \qquad (49)$$

Using the mean value theorem for integrals and fundamental theorems of calculus and line integrals, this can be rewritten as follows.

$$\ell(\beta'') \left[w(\beta') - w(\beta) \right] + \left[\mathbb{E}[R(\beta')] - \mathbb{E}[R(\beta)] \right] > 0, \qquad (50)$$

where $0 < \ell_\rho(\beta'') < \rho$ is the individual labor supply $(\rho - h(\beta''))$ at some policy on the path χ. It follows that $V(\beta'; \rho) > V(\beta; \rho)$.

Implicit in this argument is the existence of a smooth curve χ joining β and β', on which the value function is continuously differentiable everywhere. This is true for the single dimensional case, because there is a cutoff wage below which a worker with endowment ρ does not supply any labor. Given that the wage decreases monotonically with e, this implies a cutoff e above which the worker ρ doesn't work. This means that any enforcement policy in-between two enforcement policies where the endowment constraint is slack is a slack policy itself, and differentiability is an issue only at the policies at which the labor constraint just starts binding. It is also true for the multidimensional case. To see this, just notice that the wage decreases monotonically both in τ and e. This implies that on the (e, τ) space $([0, 1] \times \mathbb{R}_+)$, for each ρ, there exists a decreasing locus of policies, with slack policies lying below it and binding policies lying above it. This implies that the space of slack policies for each worker is path-connected. ∎

Proof of Proposition 10 (Unidimensional MVT with binding constraints): Proof is by exhaustion. Let e^m denote the optimal effective enforcement rate of the median voter.

1. Suppose that e^m is a binding policy for the median. This implies $\rho^m \leq h(e^m) \equiv h(w(e^m))$ and therefore everybody with $\rho \leq h(e^m)$ (including the median) spends all their available times in leisure, i.e. $h_\rho(e^m) = \rho$ for all $\rho \leq h(e^m)$. Hence, they constitute a majority. This also implies that expected transfers are

maximized at e^m because once the leisure constraint becomes binding, the only consideration is expected transfers.

First, pick any $e' > e^m$. Then $w(e') < w(e^m)$ and thus $h(e') > h(e^m)$, which implies more people would end up with binding constraints. All voters whose constraints were previously binding would still have binding constraints and therefore pick e^m; the revenue maximizing policy. Since they constitute a majority, e' gets beaten.

Next, pick any $e' < e^m$. There are two cases two consider. First, if $\rho^m \leq h(e') < h(e^m)$, then the above argument still holds and e' gets beaten by voters whose constraints are binding under both policies. Second, consider the case where $h(e') < \rho^m$. Voters in $[\underline{\rho}, h(e'))$ have binding constraints under both policies so they pick e^m. On the other hand, voters who are in $[h(e'), \rho^m)$ have binding constraints under e^m but supply a positive amount of labor under e'. These would support e^m only if the following inequality held.

$$\mathbb{E}[R(e^m)] + v(\rho) \geq \mathbb{E}[R(e')] + v(h(e')) + w(e')(\rho - h(e')). \tag{51}$$

We know from median voter's optimal choice that the following holds.

$$\mathbb{E}[R(e^m)] + v(\rho^m) \geq \mathbb{E}[R(e')] + v(h(e')) + w(e')(\rho^m - h(e')). \tag{52}$$

(51) would imply (52) for all $\rho \in [h(e'), \rho^m)$ if the following condition was satisfied.

$$v(\rho^m) - v(\rho) \leq w(e')(\rho_m - \rho). \tag{53}$$

From concavity of v, one gets the following chain of inequalities.

$$\frac{v(\rho_m) - v(\rho)}{\rho^m - \rho} \leq \frac{v(\rho_m) - v(h(e'))}{\rho^m - h(e')} \leq v'(h(e')) = w(e'), \tag{54}$$

where the final equality follows from the fact that $h(e')$ is the worker just at the corner. So all $\rho \in [h(e'), \rho^m)$ prefers e^m over e' and e^m wins.

2. Suppose that e^m is a slack policy for the median. This implies $h(e^m) < \rho^m$. First, pick any $e' > e^m$. If $h(e') \leq h(e^m)$; then e' gets beaten by the votes of every worker to the right of the median. This is because the value functions of all to the right are differentiable for any $e \in [h(e'), h(e^m)]$ as these are workers who make interior choices in that range. This implies that their marginal rates of substitution exist and are ordered in an increasing fashion, which we know implies that if a worker with a lower type (ρ^m) picks a lower policy (e^m), then all workers

to the right of that worker picks that lower policy as well. Now suppose instead $h(e') > \rho_m$. Consider first the workers in $(\rho^m, h(e')]$. From median's choice, we know that the following inequality holds.

$$v(h(e^m)) + \mathbb{E}[R(e^m)] + w(e^m)(\rho^m - h(e^m)) \geq v(\rho^m) + \mathbb{E}[R(e')]. \qquad (55)$$

For a unanimous political support at this range, we need the following inequality to hold for any $\rho \in (\rho^m, h(e')]$:

$$v(h(e^m)) + \mathbb{E}[R(e^m)] + w(e^m)(\rho - h(e^m)) \geq v(\rho) + \mathbb{E}[R(e')]. \qquad (56)$$

This inequality holds for the relevant range as the following sufficient condition (which makes sure that (55) implies (56)) is implied by concavity of $v(\cdot)$ and optimality of leisure choices.

$$\frac{v(\rho) - v(\rho^m)}{\rho - \rho^m} \leq v'(\rho^m) = w(e^m); \quad \forall \rho \in (\rho^m, h(e')]. \qquad (57)$$

Next, consider workers in $(h(e'), \bar{\rho}]$. Since they are making interior choices under both policies, they would pick e' if the following inequality held.

$$v(h(e^m)) + \mathbb{E}[R(e^m)] + w(e^m)(\rho - h(e^m))$$
$$\geq v(h(e')) + \mathbb{E}[R(e')] + w(e')(\rho - h(e')). \qquad (58)$$

Let $\epsilon > 0$ be an arbitrarily small number. Because $w(e^m) > w(e')$, evaluating (58) at the indifferent type $h(e')$ and adding $w(e^m)\epsilon$ and $w(e')\epsilon$ to left and right-hand sides respectively, we get the following.

$$v(h(e^m)) + \mathbb{E}[R(e^m)] + w(e^m)(h(e') + \epsilon - h(e^m))$$
$$\geq v(h(e')) + \mathbb{E}[R(e')] + w(e')(h(e') + \epsilon - h(e')). \qquad (59)$$

Since the voter within an epsilonic distance to the right picks e^m over e' and since all the value functions are differentiable at any $\rho \in [h(e') + \epsilon, \bar{\rho}]$, one can invoke the marginal rates of substitution ranking argument and the fact that $\epsilon > 0$ was arbitrary to conclude that all workers in $(h(e'), \bar{\rho}]$ would support e^m over e'. This proves that e^m beats e' with the support of all voters to the right of the median. Second, pick any $e' < e^m$. Then we have $w(e') > w(e^m)$. Start by considering workers in $[\underline{\rho}, h(e')]$. These have binding constraints under both policies and therefore pick e^m because $\mathbb{E}[R(e^m)] > \mathbb{E}[R(e')]$ (see proposition 7). Next, consider

workers in $(h(e^m), \rho^m)$. These workers have slack constraints under both policies and differentiable value functions at any $e \in [e', e^m]$. Hence, increasing marginal rates of substitution makes sure that they pick e^m over e'. Finally, consider the workers on $(h(e'), h(e^m)]$. These are workers with binding constraints under e^m and slack constraints under e'. From the above argument, we know that for any $\epsilon > 0$, the following holds.

$$v(h(e^m)) + \mathbb{E}[R(e^m)] + w(e^m)(h(e^m) + \epsilon - h(e^m))$$
$$> v(h(e')) + \mathbb{E}[R(e')] + w(e')(h(e^m) + \epsilon - h(e')). \quad (60)$$

Since $\epsilon > 0$ was arbitrary and since $w(e') > w(e^m)$, it follows that the following inequality holds as well.

$$v(h(e^m)) + \mathbb{E}[R(e^m)] + w(e^m)(h(e^m) - h(e^m)) = v(h(e^m)) + \mathbb{E}[R(e^m)]$$
$$\geq v(h(e')) + \mathbb{E}[R(e')] + w(e')(h(e^m) - h(e')). \quad (61)$$

Rearranging this, I get:

$$\frac{v(h(e^m)) - v(h(e'))}{h(e^m) - h(e')} + \frac{\mathbb{E}[R(e^m)] - \mathbb{E}[R(e')]}{h(e^m) - h(e')} \geq w(e'). \quad (62)$$

Now, using concavity of $v(\cdot)$ and the fact that $\mathbb{E}[R(e^m)] > \mathbb{E}[R(e')]$, the following holds for any $\rho \in (h(e'), h(e^m)]$.

$$\frac{v(\rho) - v(h(e'))}{\rho - h(e')} + \frac{\mathbb{E}[R(e^m)] - \mathbb{E}[R(e')]}{\rho - h(e')} \geq w(e'). \quad (63)$$

Upon rearranging, (63) becomes:

$$v(\rho) + \mathbb{E}[R(e^m)] \geq v(h(e')) + \mathbb{E}[R(e')] + w(e')(\rho - h(e')). \quad (64)$$

It follows that e^m gets the support of all $\rho \in (h(e'), h(e^m)]$ and beats e' by the unanimous support of all voters to the left of the median. Since all possibilities are exhausted, the proof is complete. ∎

Proof of Proposition 11 (Anatomy of binding and slack policies): If a worker with ρ finds a binding policy optimal, then she is not receiving any wages, so the policy she picks must maximize expected revenues, which is what $e_b^*(\tau)$ does. If she finds $e_b^*(\tau)$ optimal but this is a slack policy for her, then $\partial \mathbb{E}[R(e_b^*(\tau), \tau)] / \partial e = 0$, so (44) is strictly negative, contradiction. $\rho > \rho_I$ picking slack policies follows from the definition of ρ_I and $\rho \leq \rho_I$ picking $e_b^*(\tau)$

46

follows from the proof of proposition 10. Suppose $e_b^*(\tau) < \sup_{\rho \in (\rho_I, \bar{\rho}]} e(\tau; \rho)$. Then from proposition 7, any slack policy picker in $(\rho_I, \bar{\rho}]$ would be better off picking $e_b^*(\tau)$, because this is a slack policy for her (by definition of ρ_I), a contradiction. Pick any $\rho \in (\rho_I, \bar{\rho}]$, then $V(e, \rho; \tau)$ is differentiable in e so $\partial e(\tau; \rho)/\partial \rho < 0$ follows from proposition 6. Finally, suppose $e_{\rho_I}^c(\tau) > e_b^*(\tau)$. Then $e_b^*(\tau)$ is a slack policy for ρ_I, contradiction. Suppose instead $e_{\rho_I}^c(\tau) < e_b^*(\tau)$. Let $e^*(\tau; \rho)$ denote the optimal policy for any worker. Then $e^*(\tau; \rho_I) = e_b^*(\tau)$, so $e^*(\tau; \rho_I) - e_{\rho_I}^c(\tau) > 0$. $e_{\rho_I}^c(\tau)$ is continuous from lemma 11 and $e^*(\tau; \rho_I)$ is continuous in ρ from Berge's maximum theorem, so one can find a sufficiently small $\epsilon > 0$ such that $e^*(\tau; \rho_I + \epsilon) - e_{\rho_I + \epsilon}^c(\tau) > 0$.[45] But this is a contradiction, as $e^*(\tau; \rho_I + \epsilon)$ should be a slack policy for $\rho_I + \epsilon$. ∎

Proof of Proposition 12 (Bidimensional MVT with binding constraints):
Proof is again by exhaustion and similar to the unidimensional case. Few differences arise because marginal rates of substitution ordering argument is no longer valid. Let $\beta^m \equiv (e^m, \tau^m)$ be the optimal policy pair for the median voter.

1. Suppose β^m is a policy under which median voter's constraint binds. This implies $\rho^m \leq h(\beta^m)$. Pick any alternative β'.

First, suppose that β' produces a corner solution for the median. Since expected transfers are maximized under β^m, all workers to the left of the median prefer β^m because they all have binding constraints under both policies.

Second, suppose that β' produces an interior solution for the median. Then we must have $h(\beta') < \rho^m$. Consider workers with $\rho \in [\underline{\rho}, h(\beta')]$. These are the workers with binding constraints under both policies and therefore pick β^m. Now consider workers with $\rho \in (h(\beta'), \rho^m)$. They would support β^m if the following inequality held for all $\rho \in (h(\beta'), \rho^m)$.

$$v(\rho) + \mathbb{E}[R(\beta^m)] \geq v(\rho) + \mathbb{E}[R(\beta')] + w(\beta')(\rho - h(\beta')). \tag{65}$$

Median voter's revealed preference would imply (65) if the following held for all ρ in the relevant range.

$$v(\rho^m) - v(\rho) \leq w(\beta')(\rho^m - \rho) \tag{66}$$

But this holds for all $\rho \in (h(\beta'), \rho^m)$ due to concavity of $v(\cdot)$ and the solution to indifferent type's $(h(\beta'))$ optimal decision:

---

[45] Because $\rho \mapsto [0, 1]$ is a continuous and compact-valued correspondence and $V(e, \tau; \rho)$ is continuous, the solution correspondence is upper hemicontinuous. Any single-valued upper hemicontinuous correspondence is continuous.

$$\frac{v(\rho^m) - v(\rho)}{\rho^m - \rho} \leq \frac{v(\rho^m) - v(h(\beta'))}{\rho^m - h(\beta')} \leq v'(h(\beta')) = w(\beta'). \tag{67}$$

It follows that β^m beats β' with the unanimous support of all voters to the left of the median.

2. Suppose that median's constraint is slack at β^m. This implies $h(\beta^m) < \rho^m$. Pick an alternative β'.

First, suppose that $h(\beta^m) \leq h(\beta') < \rho^m$. Since the median and every worker to the right of the median have slack constraints under both policies, intermediacy of preferences within that range implies that β^m beats β' with the support of the right.

Next, suppose that $\rho^m < h(\beta')$. Then $w(\beta') < w(\beta^m)$. First consider workers in $(\rho_m, h(\beta')]$. These workers would choose β^m if the following condition held.

$$v(h(\beta^m)) + \mathbb{E}[R(\beta^m)] + w(\beta^m)(\rho - h(\beta^m)) \geq v(\rho) + \mathbb{E}[R(\beta')]. \tag{68}$$

Given median's revealed preference, this condition is implied by the following.

$$v(\rho) - v(\rho^m) \leq w(\beta^m)(\rho - \rho^m). \tag{69}$$

(69) holds because concavity of $v(\cdot)$ and the optimization problem of the median voter imply:

$$\frac{v(\rho) - v(\rho^m)}{\rho - \rho^m} \leq v'(\rho^m) = w(\beta^m). \tag{70}$$

Next, consider workers in $(h(\beta'), \bar{\rho}]$. Even though median voter's constraint binds under β', the following must still be true:

$$v(h(\beta^m)) + \mathbb{E}[R(\beta^m)] + w(\beta^m)(\rho^m - h(\beta^m))$$
$$\geq v(h(\beta')) + \mathbb{E}[R(\beta')] + w(\beta')(\rho^m - h(\beta')). \tag{71}$$

To see this, just note that the following describes median voter's optimization problem given that she has chosen a policy where she supplies non-zero labor.

$$\max_{\beta \in \mathbb{R}_+ \times [0,1]} V(\beta; \rho^m) = v(h(\beta)) + \mathbb{E}[R(\beta)] + w(\beta)(\rho^m - h(\beta))$$
$$\text{s.t.} \quad h(\beta) \leq \rho^m. \tag{72}$$

If (71) didn't hold, then the constraint in (72) should have been binding and

48

β^m should have been yielding a binding policy for the median. Moreover, since $w(\beta^m) > w(\beta')$; I can add $w(\beta^m)(\rho - \rho^m)$ to the left-hand side and $w(\beta')(\rho - \rho^m)$ to the right-hand side of (71) to obtain:

$$v(h(\beta^m)) + \mathbb{E}[R(\beta^m)] + w(\beta^m)(\rho - h(\beta^m))$$
$$\geq v(h(\beta')) + \mathbb{E}[R(\beta')] + w(\beta')(\rho - h(\beta')), \qquad (73)$$

for any $\rho \in \left(h(\beta'), \overline{\beta}\right]$. It follows that β^m beats β' with the unanimous support of all voters to the right of median.

Finally, suppose $h(\beta') < h(\beta^m) < \rho^m$. Since this implies $w(\beta') > w(\beta^m)$, from proposition 7, we must have $\mathbb{E}[R(\beta')] \leq \mathbb{E}[R(\beta^m)]$. It follows that workers in $\left[\underline{\rho}, h(\beta')\right]$, whose constraints are binding under both policies, pick β^m over β'. Moreover, since workers in $(h(\beta^m), \rho^m]$ have slack constraints under both policies, from the intermediacy of preferences under such situation and the fact that $w(\beta^m) < w(\beta')$, these prefer β^m as well. Finally, a concavity argument analogous to the one made in the proof of the unidimensional case shows that β^m is supported by voters in $\left[h(\beta'), h(\beta^m)\right]$ and thus beats β' by the unanimous support of all workers to the left of the median. Since all possibilities are exhausted, the proof is complete. ∎

References

[1] **Allingham, M.J. and A. Sandmo (1972)**, "Income tax evasion: a theoretical analysis", *Journal of Public Economics*, 1(3-4): 323-338.

[2] **Bailey, C. and J. Turner (2001)**, "Strategies to reduce contribution evasion in social security financing", *World Development*, 29(2): 373-386.

[3] **Bárány, Z.L. (2009)**, "Some tax evasion - more redistribution: a political economy model of tax evasion", mimeo, London School of Economics and Social Sciences, London (UK).

[4] **Barnett, R.C., J. Bhattacharya and H. Bunzel (2014)**, "Voting for income-immiserizing redistribution in the Meltzer-Richard model", *Economic Inquiry*, 52(2): 682-695.

[5] **Bayer, R. and F.A. Cowell (2009)**, "Tax compliance and firms' strategic interdependence", *Journal of Public Economics*, 93(11-12): 1131-1143.

[6] **Becker, G. (1968)**, "Crime and punishment: an economic approach", *Journal of Political Economy*, 76(2): 169-217.

[7] **Besfamille, M., P. De Donder and J.M. Lozachmeur (2013)**, "The political economy of the (weak) enforcement of indirect taxes", *Journal of Public Economic Theory*, 15(6): 856-883.

[8] **Borck, R. (2004)**, "Stricter enforcement may increase tax evasion", *European Journal of Political Economy*, 20(3): 725-737.

[9] **Borck, R. (2009)**, "Voting on redistribution with tax evasion", *Social Choice and Welfare*, 32(3): 439-454.

[10] **Cahuc, P. and G. Laroque (2014)**, "Optimal taxation and monopsonistic labor market: does monopsony justify the minimum wage?", *Journal of Public Economic Theory*, 16(2):259-273.

[11] **Casaburi, L. and U. Troiano (2013)**, "Ghost-house busters: the electoral response to a large anti tax evasion program", MPRA Paper, 52242, University Library of Munich, Munich, Germany.

[12] **Chae, S. (2002)**, "Tax incidence with bargaining", *Economics Letters*, 77(2): 199-204.

[13] **Cowell, F.A. (2004)**, "Carrots and sticks in enforcement", in: Aaron, H. J. and J. Slemrod (eds.), *The Crisis in Tax Administration*, Brookings Institution Press, Washington DC (US), pp. 230-275.

[14] **Cremer, H. and F. Gahvari (1993)**, "Tax evasion and optimal commodity taxation", *Journal of Public Economics*, 50(2): 261-275.

[15] **Crandall, W. (2010)**, "Revenue Administration: Autonomy in Tax Administration and the Revenue Authority Model", IMF Technical Notes and Manuals, 10/12, Washington DC (US).

[16] **Eichhorn, C. (2006)**, "Optimal policies in the presence of tax evasion", Ph.D. thesis, Ludwig Maximilian University of Munich, Munich (Germany).

[17] **Gans, J. and M. Smart (1996)**, "Majority voting with single-crossing preferences", *Journal of Public Economics*, 59(2): 219-237.

[18] **Gillian, C., J. Turner, C. Bailey and D. Latulippe (eds.) (2000)**, *Social Security Pensions: Development and Reform*, ILO Press, Geneva (Switzerland).

[19] **Goerke, L. and M. Runkel (2006)**, "Profit tax evasion under oligopoly with endogenous market structure", *National Tax Journal*, 59(4): 851-857.

[20] **Grandmont, J.M. (1978)**, "Intermediate preferences and the majority rule", *Econometrica*, 46(2): 317-330.

[21] **Hashimzade, N., Z. Huang and G.D. Myles (2010)**, "Tax fraud by firms and optimal auditing", *International Review of Law and Economics*, 30(1): 10-17.

[22] **Hemming, R. and M.J. Keen (1983)**, "Single-crossing conditions in comparisons of tax progressivity", *Journal of Public Economics*, 20(3): 373-380.

[23] **Karadeniz, O. (2013)**, "Pensions, health and long-term care: Turkey", Asisp country document, Cologne (Germany).

[24] **Kolm, S.C. (1973)**, "A note on optimum tax evasion", *Journal of Public Economics*, 2(3): 265-270.

[25] **Marelli, M. (1984)**, "On indirect tax evasion", *Journal of Public Economics*, 25(1-2): 181-196.

[26] **Meltzer, A.H. and S.F. Richard (1981)**, "A rational theory of the size of government", *Journal of Political Economy*, 89(5): 914-927.

[27] **Mesa-Lago, C. (1998)**, "Comparative features and performance of structural pension reforms in Latin America, *Brooklyn Law Review*, 64(3): 771-793.

[28] **Milgrom, P. and C. Shannon (1994)**, "Monotone comparative statics", *Econometrica*, 62(1): 157-180.

[29] **Paramonova, Y. (2014)**, "The optimal information reporting and tax audit policy", working paper, University of Michigan, Michigan (US).

[30] **Persson, T. and G. Tabellini (2000)**, *Political Economics: Explaining Economic Policy*, MIT Press, Cambridge (US).

[31] **Reinganum, J.F. and L.L. Wilde (1985)**, "Income tax compliance in a principal agent framework", *Journal of Public Economics*, 26(1): 1-18.

[32] **Roberts, K.W.S. (1977)**, "Voting over income tax schedules", *Journal of Public Economics*, 8(3): 329-340.

[33] **Roine, J. (2006)**, "The political economics of not paying taxes", *Public Choice*, 126(1-2): 107-134.

[34] **Sanchez, I. and J. Sobel (1993)**, "Hierarchical design and enforcement of income tax policies", *Journal of Public Economics*, 50(3): 345-369.

[35] **Sandmo, A. (1981)**, "Income tax evasion, labor supply and the equity-efficiency tradeoff", *Journal of Public Economics*, 16(3): 265-288.

[36] **Sandmo, A. (2005)**, "The theory of tax evasion: a retrospective view", *National Tax Journal*, 58(4): 643-663.

[37] **Saunders, P. and X. Shang (2001)**, "Social security reform in China's transition to a market economy", *Social Policy and Administration*, 35(3): 274-289.

[38] **Traxler, C. (2009)**, "Voting over taxes: the case of tax evasion", *Public Choice*, 140(1): 43-58.

[39] **Traxler, C. (2012)**, "Majority voting and the welfare implications of tax avoidance", *Journal of Public Economics*, 96(1): 1-9.

[40] **Virmani, A. (1989)**, "Indirect tax evasion and production efficiency", *Journal of Public Economics*, 39(2): 223-237.

[41] **Wang, L.F.S. and L. Conant (1988)**, "Corporate tax evasion and output decisions of the uncertain monopolist", *National Tax Journal*, 41(4): 579-581.

[42] **Yaniv, G. (1988)**, "Withholding and non-withheld tax evasion", *Journal of Public Economics*, 35(2): 183-204.

[43] **Yaniv, G. (1995)**, "A note on the tax-evading firm", *National Tax Journal*, 68(1): 113-120.

[44] **Yitzhaki, S. and V. Yitzhak (1989)**, "On the shadow price of a tax inspector", *Public Finance*, 44(3): 492-505.

Chapter 2: A Sleeping Beauty - Media Strength and Politician Selection

Abstract

This chapter develops a signalling game of pure adverse selection with costly political participation, rational voters and an altruistically benevolent media having a costly and imperfect technology of investigation for a dual purpose: to understand political selection effects of a stronger media and to assess the impact of improvements in media's technology on the intensity of its journalistic activities. Two results stand out. First, improvements in investigation costs and journalistic strength reduce adverse selection by increasing media's threat of exposing low quality challengers. Second, while reductions in investigation costs always lead the media to be more active in equilibrium, strength gains can lead to increasing media passivity, especially if its investigative technology is already strong relative to the extent of potential adverse selection. This last result underlines a possible "paradox" of good journalism: if the media is already posing a strong threat of exposure, then it does not need to use it as often to achieve the desired outcome.

1 Introduction

1.1 Motivation and Summary

The well functioning of a democracy hinges on its ability to establish two types of institutions: those that provide proper incentives to make sure that those who get in power govern properly once they are in power, and those that allow its citizens to choose the most willing and able. The former refers to the problem of moral hazard and the importance of accountability.[1] The latter refers to the importance of selection. Notice the distinction between "willing" and "able". As noted by Besley (2006), this distinction goes (at least) back to the ideas of Madison in *The*

---

[1] Accountability can be supplied via many channels. Two obvious examples are (informed) voting and checks and balances.

Federalist Papers ("virtue" and "wisdom"), and embodies the implicit assumption that agents possess or lack intrinsic moral characteristics which make them more or less suitable for public service. There is also a tradition going back to Brennan and Buchanan (1980) arguing that morality issues are in essence incentive issues and selection concerns should focus exclusively on the ability of agents. This chapter deals with selection in this ability/competence sense. Although the job definition of a politician is usually not as straightforward as a job definition of a firm employee, I assume that there is a summary statistic that translates to better public service (in my case, higher public good), reflecting the quality of a politician.[2]

Politician selection may depend on many factors: organization of political parties, size of government, degree of political competition, public remuneration, etc. For example, Besley (2004) and Caselli and Morelli (2004) show that a combination of fixed public pay and performance-contingent outside options can lead to a particularly toxic form of adverse selection by excluding high quality people and only attracting low quality candidates to political life. This sort of selection issues can be mitigated by interventions such as allowing for moonlighting, or they might simply not arise in a threatening form due to factors such as imperfect substitutability of private/public sector skill sets, or already-availability of a politician class accommodating a wide range of skills with a sufficiently strong stake in the quality of public services. A second sort of selection issue can appear in the form of non-excludability of low quality politicians, particularly if non-public service related office benefits/rents are high compared to costs of political participation. This is the setting considered in this chapter and it is when the media can shine as the fourth estate. By providing valuable screening services too costly to be undertaken at an individual level, media can help with elections to operate as "filters" as termed by Cooter (2003).

In case the media assumes the role of screening for low quality politicians, it can use a variety of tools to acquire and disseminate office-relevant information about a candidate. For instance, it can use its contacts or reporters to acquire background information or track record for a candidate, or obtain informed opinions about candidates using its network of political experts.[3] Regarding dissem-

---

[2]There is some evidence supporting this. For instance, Congleton and Zhang (2013) estimate a statistically significant and positive impact of US presidents' human capital on economic growth.

[3]For example, a candidate in a mayoral election might have previously served as a public official in some state institution or as a provincial governor in a relatively small polity. Even if the candidate is previously untried in public service, media can retrieve performance relevant information such as education or assess the candidate's competence via phone interviews and/or

ination, it can produce news shows contrasting the candidate(s) and possibly a relatively better known incumbent or even arrange an interview with the goal of "exposing" the incompetence of a candidate.[4,5] When combined with costly electoral participation, this sort of media screening can improve selection by discouraging low quality politicians from running for office, as an increased likelihood of being exposed would reduce their chance of an electoral victory. On the other hand, the media would usually be limited by the strength of its journalistic resources in discovering such office-relevant information. Furthermore, when the idea of a media outlet providing this sort of service is taken to its logical extremity, say, an altruistic agent who exclusively cares about the welfare of its fellow citizens and thus acts as the "extended arm" of voters, a hypothetical possibility appears. If the media becomes sufficiently strong in its ability to sort out the good apples from the bad apples, then the mere threat of exposure can be enough to deter incompetent candidates from political entry without requiring media to engage (as) actively in political journalism. This argument can also be extended to situations involving more than selection issues and can even be considered as a complementary explanation for the commonly perceived decline in political journalism in the West.[6]

To bring these ideas together, this chapter builds a signalling model combining costly electoral participation, endogenous media investigation and voting. It is a pure adverse selection model, so there are no pre-election ideological or post-election policy-wise differences. The outcome of the election is a public good, the quantity of which is assumed to be increasing in the competence of the office holder. As such, the model should be interpreted as reflecting elections where competence is the (almost) only concern (e.g. mayoral elections), or as approximating political environments where competence is one of the primary concerns.[7] The setting features a single candidate whose quality (competence)

---

experts judging the coherence of the candidate's project proposals and arguments.

[4] Of course, being able to disseminate such information requires an electorate that has a relatively unhindered access to mediums of dissemination (TV, newspapers, etc.) and a certain level of public trust to the media outlet in question. The former can in fact be an alternative interpretation for media strength (which is mainly used as capturing media's investigation powers in this chapter) as discussed in footnote 18. The latter problem is swept aside by the assumption of an altruistic media and the common knowledge of this fact, but should no doubt constitute one of the next steps in extending the analysis.

[5] For particularly humorous examples of incompetence exposure via interviews, see the TV series *Newsroom*, especially episodes concerning Tea Party.

[6] See for instance Lloyd and Toogood (2014).

[7] Of course ideological differences can play a large role in undermining selection and create all sorts of problems leading to media bias. This dimension is ignored for the sake of tractability and sharpening the main point of the chapter. See the models of Ashworth and Bueno de Mesquita (2008) and Bernhardt, Krasa and Polborn (2008), which are discussed in the next

is private information, who decides whether to challenge an incumbent (whose quality is commonly known) by paying participation costs and who cares about the public good as well as office rents; a "watchdog" media which can undertake a costly investigation that may or may not reveal the challenger's quality before the elections; and a set of voters rational in their beliefs and interested in electing the most competent politician. Note that the incumbent is not actually a player but simply a payoff parameter. The model is presented in more detail at the beginning of section 2 but some of the modelling assumptions are worth going over. The existence of an incumbent as described above serves three purposes. First, it captures the idea that in elections, it is usually the case that there is a well-known incumbent who faces competition from a relatively less known challenger for whom electoral participation at least involves the initial PR costs. Second, it makes sure that the office never remains empty.[8] Third, it provides a reference point which allows an intuitive characterization of the consequences of adverse selection. Allowing only for a single candidate challenging against the incumbent simplifies the presentation and the model can easily be extended to incorporate multiple candidates as explained subsequently.[9] The choice of an altruistic media with limited investigation powers provides a simple framework for studying selection effects of a stronger media, as well as emphasizing the hypothetical possibility outlined in the previous paragraph. Finally, allowing voters to be sophisticated in their beliefs (in a Bayesian sense) serves the dual purpose of underlining the relevance of media in selection even when the voters do a relatively good job in it by modifying their voting strategies, and preempting potential criticisms concerning irrational voters. As discussed before concluding, this assumption can indeed imply too much voter rationality, but dropping it would only increase the relevance of media selection as well as leaving the conclusions qualitatively intact.

The chapter has two main results. First, whenever adverse selection is present, an improvement in media's information technology, whether it involves cost reductions or gains in investigation strength, always (weakly) improves selection by increasing the minimum quality of an equilibrium challenger. Basically, when voters are rational in their beliefs, media's task becomes sparing voters from a retrospective mistake of electing an inferior challenger, and as the threat of expo-

---

subsection.

[8]Since the sender is signalling by challenging or not, this is a possibility if there is no default incumbent. An election with no contestants should be rare but uncontested incumbents can occur from time to time, e.g. Wrighton and Squire (1997) document that it is especially common in US House elections.

[9]For a more thorough discussion of the assumptions including how they could be relaxed and how would their modification affect the results, see pages 117-119 before the conclusion.

sure becomes strong, the more reluctant inferior candidates become to challenge. Second, while cost reductions always make the media more likely to engage in political journalism, a gain in investigative strength has a non-monotonic effect on journalistic activity in equilibrium, reducing it if the media is already relatively strong. Yet, this decrease never becomes strong enough to undermine media's deterrence. This is in line with the previous discussion and underlines the possibility that even if the media appears "asleep", it can still be doing its job.

1.2 Previous Literature and Contribution

This chapter stands at the intersection of two distinct but somehow interwoven strands of literature: political economy of mass media and political selection. I start with the former, which can be roughly divided into three substrands: capture, coverage and slant.[10]

Capture literature deals with consequences of manipulation attempts to media's information stream by other actors using incentive tools such as bribes or threats. The most famous paper in this literature is by Besley and Prat (2006). Their model is a pure adverse selection one with a politician being either good or bad in the productivity sense. There are multiple symmetrically informed media outlets which (all at once) can receive a perfectly informative signal about the incumbent with some probability, only if the incumbent is bad. This probability is analogous to media's investigation strength in this chapter, albeit media investigation occurs automatically and costlessly in their model. If media outlets observe the signal, they receive a bribe offer from the incumbent. They than either suppress the signal in return for bribes, or transmit it to voters in return for journalistic benefits which are decreasing in the number of media outlets. This is followed by voters either retaining the incumbent, or replacing him with a random politician drawn from a readily available pool. Their results indicate that while selection is improved with media strength, possibility of capture is independent of it and rather depends on the relative magnitudes of office rents and journalistic benefits. More precisely, an improvement in media strength improves selection only if the media is not captured and media won't be captured if office rents don't cover the necessary amount for bribing the entire industry. This allows them to conclude that a more pluralistic media industry is harder to capture, and thus it is beneficial for selection. Other than obvious differences in media's motives and possibility of bribes, there are two important differences between their model

---

[10]Prat and Strömberg (2011) provide a thorough survey for this literature.

and the model in this chapter. First, information acquisition is not contingent on media's decision and there is no signalling by politicians to guide this decision if it were. Second, selection occurs at the incumbency level once the politician is in the office, i.e. media determines *ex post* selection.[11] Corneo (2006) devises a model where capture can originate from an heterogenous electorate which consists of agents differing in their ownership shares for some efficiency-enhancing prospective public project, the initiation of which depends on a voting outcome. Media can either be opportunistic or idealistic, and is privately informed about the social costs (externalities) associated with this project. If it turns out to be opportunistic, it can then collude with a subset of voters in a rent-sharing deal to manipulate public beliefs. It can do so despite the rationality of voter beliefs due to uncertainty regarding its type. His results indicate that a higher wealth (ownership share) concentration can facilitate media capture, but this can be welfare-enhancing depending on the initial wealth distribution. While quite different in its focus, his depiction of the media is quite similar to the one in this chapter, in that it shares a common concern (payoff) with voters. In fact, when their media is an idealist, its incentives are analogous. Petrova (2008) studies a very similar problem (inequality and capture) in the context of taxation. Instead of a matching based collusion initiated by an opportunistic media, she takes a lobbying approach where rich voters can coalesce to buy the media off. One interesting result she derives is that an electorate with an easier access to media can decrease media freedom by increasing the likelihood of capture. In a more recent paper, Gehlbach and Sonin (2014) study government capture for a mobilization purpose in the context of an autocratic regime with a monopolistic media as in Besley and Prat (2006), i.e. caring about journalistic benefits (advertisement revenues) and government "contributions". A captured media manipulates (mobilizes) an electorate by misreporting the state of nature to nudge it towards a certain action. They show that while a stronger mobilization motive can encourage the government towards following a more aggressive capture strategy, higher journalistic benefits can incentivize it to nationalize the media altogether.

There are also papers that study how voters' ability to monitor elected officials and therefore political outcomes respond to media coverage, assuming that the latter is governed by market dynamics. Prat and Strömberg (2005) combine

---

[11]Political selection may originate from two mechanisms: *ex post* selection, i.e. the process by which politicians are evaluated once they are in the office and retained or replaced as seems fit, or *ex ante* selection, referring to the elimination of subpar candidates at the election stage, or even inducement of self-selection upon them so that they don't run for office in the first place. So far, the literature on the political economy of mass media have exclusively focused on selection in the former sense and this chapter is about the latter.

a probabilistic voting model featuring an informed/uninformed voter dichotomy with an industrial organization approach assuming a monopolistic media endowed with an increasing returns technology. Voters are segregated into multiple groups and they can be targeted with public goods using a combination of group transfers and group-specific politician abilities. Media (a newspaper) decides on how much coverage to allocate across different issues (group transfers) basing on certain group characteristics. Voters can purchase the newspaper to find out (stochastically) about the amount of transfers they received, and the probability of this is increasing in media's coverage. This allows them to update their beliefs on incumbent's competence regarding their own group, which determines their decision to retain or replace the incumbent with a random untried candidate. This setting allows them to deduce a wide array of predictions, among which the most important are: larger groups and groups consisting of voters more valuable to advertisers receive more coverage to their issues, which makes the incumbent more accountable towards these groups.

Finally, several studies deal with political consequences of ideological bias in media, also called media slant to underline the distinction from settings where the bias in reporting arises due to some form of capture and favours a specific group. Bernhardt, Krasa and Polborn (2008) study a model where such bias arises due to polarization in audience preferences: they assume that leftist voters like to receive positive news about left-wing politicians and negative news about right-wing ones, and vice versa. They assume that there exists two media outlets operating in left and right markets respectively and show that profit maximizing behaviour lead them to suppress negative facts about politicians associated with their respective markets, where the politicians differ along an intrinsic corruption dimension. Voters, despite being ideologically biased towards the kind of news they receive, care about corruption regardless of ideology and can deduce the nature of media's suppression. This leads them to rationally disregard media suppression and expect the politician to be as corrupt as the average politician. Their results indicate that this sort of demand-driven slant can negatively affect selection: a candidate more corrupt than average can be elected and a candidate less corrupt than average may fail to get elected due to their ideological labels. Duggan and Martinelli (2011) take a supply-side approach to media slant. There is an incumbent with a known policy position and a challenger whose policy stance is unknown. Policy positions are pairs of public good levels and tax rates, and they are taken as fixed. Their focus is on media's problem, which provides information to its audience on challenger's position by projecting it onto a single

dimensional space. This captures the idea of cognitive limitations on readers' front which necessitates the media to transmit the news in an easily digestible form. The way it does the projection (specifically, its slope) represents media slant, which is optimally determined (from media's perspective) depending on whether the media is pro-incumbent or pro-challenger. While media's strategy does not have any selection or accountability effects due to politician behaviour being exogenous, it still have welfare consequences because it affects voter decisions. In particular, they show that a partisan media may be better for voters compared to an objective one.

Regarding political selection, earlier models focus on *ex post* selection and assume that candidates are passive players, fulfilling the role of an exogenous outside option for voters.[12] Banks and Sundaram (1990) consider an infinitely repeated game of elections with quality-heterogenous politicians where the politician in the office produces an output at each period.[13] Output is stochastic and only depends on politician's type which takes values from a finite set, i.e. incumbents do not act. At each period, a representative voter decides whether to retain the incumbent or replace him with a randomly chosen challenger basing on the history of incumbent's output realization. They show that voter's optimal policy is belief-stationary, and that it is characterized by a cutoff rule representing the output level at which the voter is indifferent between retaining and replacing the incumbent. Cutoff rules based on voter indifference are a common theme in selection literature and holds in this chapter as well unless the media is active in equilibrium, in which case voters strictly prefer replacing the incumbent. This emphasizes the welfare benefits (from voters' perspective) of having a media caring about selection. Coate and Morris (1995) consider a two-period model combining elements of moral hazard and adverse selection to study the problem of inefficient transfers to special interest groups. Their politicians can either be good or bad, where the former only cares about voters and the latter cares about the special interest group more than they care about voters. There is a single project financed by tax receipts which generate a net public benefit in the good state of the world and a net loss in the bad state. The project generates benefits for the special interest in both states so it represents an inefficient transfer from voters to the special interest group in the bad state. Politicians not only have private information about their own types, but also about the probabilities of these states and this latter introduces a moral hazard aspect to their model. In

---

[12]See Besley (2005) for a decent survey of this literature.

[13]They also build a two-period model combining selection and moral hazard in Banks and Sundaram (1998) exhibiting similar results.

addition, voters possess initial reputations (priors) about politicians and they can update these for the incumbent but not for the challenger. Although their main focus is accountability, i.e. the disciplining effect induced by voters' reelection rules to the behaviour of the bad incumbent, their results have selection implications. In particular, they show that if challenger's reputation is high relative to incumbent's initial reputation, voters replace bad incumbents more often despite the disciplining effect improving bad incumbents' behaviour to a certain extent. However, taking reputation priors as given, their approach still treats challengers as exogenous outside options.

The first steps towards endogenizing the outside option posed by challengers came from Osborne and Slivinski (1996) and Besley and Coate (1997). Dubbed citizen-candidate models, they endogenize political entry by studying the problem of citizens who can either stay out of politics and vote, or run for office and implement their favourite policies as well as receive office benefits in case they win. Their main concern is to endogenize the number of candidates competing in equilibrium. So in both papers, all citizens are allowed to run for office and the heterogeneity is along the policy-preference dimension. A common simplification made when applying the citizen-candidate framework to selection problems is to limit potential challengers to a subset of the electorate (the politician class). One of the first such applications is made by Besley (2004), who considers a model of political entry intending to capture the first sort of selection issue mentioned in the previous subsection (i.e. exclusion of good candidates). In his setting, politicians differ in two dimensions: morality (corrupt/honest, or in Besley's terms "dissonant" and "congruent"), and opportunity costs (foregone private sector wages). There is a continuum of both corrupt and honest potential challengers and electoral participation is costless, besides the foregone private sector wage which is a concern only if the challenger wins. Winner is assumed to be determined by a random selection among the pool of participants in case voters decide to replace the incumbent. Since challenging is costless, politicians challenge if their expected office benefits exceed their opportunity costs. Former consist of public wages for both types and rents on top of those for corrupt types. As a result, the opportunity cost above which politicians do not challenge is higher for corrupt types. This implies that if morality does not influence private sector productivities, then a larger fraction of corrupt politicians will challenge. Consequently, an increase in politician wages improves average honesty of the challenger pool, with the improvement being stronger if the population distribution initially favours honesty. Besley (2004) also considers an extension by adding a political compe-

tence dimension positively correlated with outside options. He finds that average challenger competence can go both ways in response to an increase in public remuneration because it provides a relatively stronger entry incentive to honest but weak politicians. Caselli and Morelli (2004) consider a citizen-candidate model of large elections (continuum of public positions) focusing on the competence of elected body. Their results concerning selection effects of public remuneration are similar to Besley (2004). They also study the effects of an informative pre-elections signal when combined with costly electoral participation: each type of challenger (low and high) sends a binary signal about his type after challenging but before voting takes place. The signal is correct with a certain probability and knowledge of this probability allows voters to update their priors. This signal is interpreted as campaign quality. They show that an increase in signal accuracy increases average challenger quality by discouraging bad politicians from running for office. This result is akin to the impact of media strength in my model, where a comparable signal is produced at the media's discretion. Although Caselli and Morelli (2004) use a citizen-candidate framework, potential challengers do not take their own competence into account when deciding to run for office. The reason for this is the existence of a continuum of political positions and the dependence of the public good on the average quality of the elected body. Messner and Polburn (2004) consider the problem of *ex ante* selection for competence in the context of a single-office election where two candidates decide on whether to run for office simultaneously. Candidates care about net office benefits, as well as the quality of the politician who gets elected. These considerations enter candidates' payoffs in an additively separable fashion which is very close to the specification used in this chapter except for the fact that their costs represent opportunity costs rather than actual participation costs. They assume that politician qualities are common knowledge so voters' problem is trivial. However, they assume that opportunity costs are private information and that these are stochastically greater for the good candidate, with the average difference exceeding the gain in public good due to his competence advantage. This ensures that bad candidates are more likely to challenge in equilibrium, which is a result similar to the papers discussed above. But when they consider the effects of an increase in public wages, they find that it can worsen selection. In essence, the only benefit of political life for the good type is a higher public good. But due to opportunity costs, this alone does not justify running most of the time. Since higher public wages increase the likelihood of a low type challenging, it also increases high type's incentives to free ride. Poutvaara and Takalo (2007) study a

selection model which distinguishes itself from previous ones by one feature: a continuum of competence levels (distributed uniformly). They also assume that challenging is costly and challengers send an informative but inaccurate signal about their types as in Caselli and Morelli (2004), but modified to suit an infinite type space. Their results imply that candidates use a cutoff rule and lower participation costs, as well as higher office benefits lead to a deterioration in *ex ante* selection by attracting worse challengers. These results are similar to ones obtained in my model and they are a byproduct of an infinite type space, as well as the presence of candidates with partial public good motivations. Finally, a recent paper by Estache and Foucart (2013) combine challenger entry with *ex post* selection, focusing on the role of two types of institutions: auditory and judiciary. For a given rule of punishment, two politicians differing in competence decide on whether to run for office to implement a preset policy in case they get elected. If they both run, one of them is randomly picked as the winner. Once the politician is in the office, he implements the policy and decides on whether to engage in theft. In particular, a high quality politician can mimic a low quality one (which has a higher cost of implementing the policy) and pocket the difference. This is followed by the auditor receiving an imperfect signal regarding the cost of implementation. If the signal indicates high cost, the judiciary then receives a similar signal on the source of inefficiency and punishes the politician if the signal indicates theft. Their results indicate that auditors and judiciary are complements and if both of them are sufficiently accurate, there exists an optimal punishment rule (a constitution) that solves both the selection and moral hazard problems, i.e. only high quality candidates challenge and they never steal.

A common element of the models in the previous paragraph is the lack of a strategic feedback mechanism between challengers and agents who have a stake in selection. This is an important dimension that underlies the non-trivial media response to gains in journalistic strength in my model. There is a literature dealing with this aspect of the political selection process, which can be dubbed the strategic challengers literature.[14] One of the earliest papers in this literature is by Banks and Kiewiert (1989), who focus on the problem of two potential challengers from the same party with differing qualities simultaneously deciding on whether to face an incumbent of the opposing party in the primaries, or to wait one term and run in open elections after the incumbent retires. They assume that the probability of an electoral victory is increasing in quality, as

---

[14]In fact, models in this literature can be seen as variants of the citizen-candidate model but this literature appeared chronologically earlier and followed an epistemologically distinct development.

well as a previous victory (and decreasing in a previous loss). The risk for the high quality candidate in waiting for incumbent's retirement is that if the weak candidate were to challenge and win, he would never be able to assume office, as being from the same party precludes him from competing against his weak colleague. While many different entry patterns are possible depending on politician qualities, they show that a larger quality gap between potential challengers lead to weak candidate challenging in primaries and strong candidate waiting for open elections. This implies (holding incumbent quality constant) that a larger quality difference between candidates can increase the likelihood of a weak candidate assuming the office, undermining political selection. In Epstein and Zemsky (1995), ex ante selection is steered by an incumbent wishing to face the weakest (quality-wise) challenger possible. Incumbent has a limited amount of resources which he can allocate between two activities: fund raising and governing. The combination of these determines incumbent's electoral strength. They assume that the probability of the incumbent getting reelected increases in his electoral strength and decreases in the quality of the challenger. A strong incumbent is better in raising funds, he has greater electoral strength for a given amount of funds raised. Potential challengers can only observe the amount of funds raised by the incumbent, and thus they have to deduce electoral strength before deciding on whether to challenge. They show that if costs of challenging relative to office benefits are neither too high nor too low, two types of equilibria might prevail: a separating one where no deterrence occurs, and a pooling one where the low quality incumbent raises a lot of funds to mimic the strong incumbent, deterring high quality candidates from challenging. Furthermore, the prevalence of the pooling equilibrium becomes more likely if the incumbents have more resources. From this, they deduce that a larger "war chest" can undermine political selection. Goodliffe (2005) builds a repeated elections model where these war chests are endogenously determined and arrives at similar conclusions. More specifically, he assumes that war chests build up over time, carrying over from funds raised in previous elections. One implication of that is, *ex ante* selection can deteriorate over time if low quality politicians manage to win several elections consecutively. All these models assume a reduced form specification for voting and look for the origins of selection elsewhere. Gordon, Huber and Landa (2007) consider a model with a single rational voter and quality-wise heterogenous candidates. Candidates have private information regarding their own competence as well as the incumbent's competence, where the latter is unobservable to the voter except for a noisy signal. Challenging is costly and allows the voter to

not only update his prior for candidates but also revise his beliefs concerning the incumbent. This is argued to be capturing the idea of politicians having inside information on other politicians. Their results indicate that a challenger entry always results in a downward revision in voter beliefs for incumbent quality, which they interpret as a force countering incumbency advantage stemming from incumbent's office performance giving him an informational edge. Notwithstanding, they show that a higher incumbent signal not only improves selection but also decreases turnover, i.e. increases incumbency advantage. Their dually asymmetric information structure lets them endow their candidates with the ability to mount a strong challenge reversing the incumbency advantage, and allows them to analyze selection effects of incumbent performance. At the same time, it requires them to impose distributional monotonicity conditions for equilibrium existence and prevents them from studying equilibrium transition patterns due to the difficulty it induces on establishing continuity properties. In addition, they consider the case where the voter can engage in costly private learning to discover both politicians' types with certainty. Their reason for limiting the number of voters to one is that with a larger set of voters, a strong free-riding incentive arises, preventing learning to take place as the probability of being pivotal becomes negligible. Their set of equilibria under that scenario is akin to the special case of full media strength in my model, except for differences in cutoffs due to different information structures. Most notably, the voter indifference result also seen in Banks and Sundaram (1990) always holds in Gordon et al. (2007), and it fails in my model precisely when the media is active and constrained by its journalistic strength. As mentioned in the previous subsection, ideological differences can play a role in undermining selection. Ashworth and Bueno de Mesquita (2008) construct a model where politicians instead differ in two (fixed) dimensions: quality and policy positions. Voters care about both issues in a quasi-linear manner. Politicians transmit a (normally distributed) noisy signal at the pre-election stage, and the incumbent transmits an additional signal before the pre-election stage. This gives the incumbent an informational advantage if he is competent. Observing the incumbent's signal, a party decides on whether to recruit a challenger among its ranks to face the incumbent, but without knowing the quality of its choice. When both signals realize, voters update their beliefs and elections take place. Besides showing that an improvement in signal quality (measured by reductions in the variance of noise) leads to an improvement in selection (i.e. a decrease in the likelihood that a high quality incumbent will be replaced by a low quality challenger), they also show that party polarization

(in terms of policy positions), electoral polarization and stronger partisanship all contribute negatively to political selection. Finally, a more recent paper by Dewan and Hortala-Vallve (2013) considers the impact of campaign asymmetry between an incumbent and a challenger. Their model differs from the previous ones by assuming that there is a challenger who is already participating to the election, but that this challenger can either choose a safe campaign or a high profile campaign, where the latter is more likely to signal his competence (succeed) only if he is competent to begin with. The incumbent faces the same choice in the policy dimension, but with the assumption that when he is competent, he has a greater advantage when choosing a high risk high reward policy (so policy choice is incumbent's campaign technology). They show that compared to the case where challenger campaigning is uninformative, giving the challenger the technology of an informative campaign can actually worsen selection by causing the incumbent to engage too aggressively in implementing the high risk policy, leading to him being replaced by an inferior challenger more frequently.

Contributions of this chapter are twofold. First, it attempts to bridge the gap between the literatures of political economy of mass media and strategic challenger entry by studying the selection consequences of costly electoral participation under imperfect media monitoring. From the standpoint of the former, it uses media screening to endogenize the outside option that voters face when deciding on whether to retain the incumbent. It emphasizes media's *ex ante* selection role, rather than *ex post* selection as is usually assumed. From the standpoint of the latter, it outsources the task of selection to a strategic media and studies the selection effects of a stronger or weaker media in the context of media-challenger interaction. Second, by characterizing the journalistic incentives of an idealized media with limited screening powers in a setting of political selection, it shows that the media can appear silent but still be doing its job in the background if its journalistic expertise and resources are strong enough so that it can detect low quality politicians with relative ease. This might constitute a complementary explanation for the contemporary decline in political journalism, one that doesn't need to invoke explanations involving capture or hegemony of commercial interests.

Next section presents the model. It starts by introducing players, timing, payoffs, equilibrium concept and so on. After deriving some general results, it proceeds to presenting equilibria under different parameter configurations. Presentation of equilibria starts with less interesting configurations (no adverse selection, etc.) and gradually progresses to more interesting ones where media is

actively involved in selection, with an intermission to refine some trivial equilibria where the incumbent always remains uncontested. Next, concentrating on parametric regions where selection issues are present, continuity and transition properties of the equilibria are studied, with a particular focus on selection and media incentives. Final section concludes.

2 Model

The model is a multi-stage game with observed actions and incomplete information. More accurately, it is a signalling game of pure adverse selection intended to capture the effects of strengthening the media on *ex ante* politician selection and media incentives when the quality of the latter is all that matters and the task of the former is limited to finding out about it. There is an incumbent whose type is common knowledge, and a candidate (potential challenger) whose type is private information decides whether to challenge the incumbent in a partly citizen-candidate, partly opportunistic spirit. The former means that candidates care about the public good, which solely depends on the quality of the politician in the office. The latter implies that they also care about office rents. Challenging is costly so it acts as a signal which is received by two receivers who act consecutively. First one is an intermediary receiver: a benevolent media monopolist who decides whether to undertake a costly investigation that may or may not reveal challenger's type and the result of which can credibly and fully disclosed. Second one is a group of decisive receivers: a set of voters who decide on whom to vote for, either based on the initial signal or in a mechanical manner due to the uncertainty resolution provided by the media.

2.1 Players, Timing and Payoffs

There is no moral hazard and no policy conflict. Politicians differ in the number of units of public good they produce if elected, which is assumed to be perfectly observable and is interpreted as politician's quality.[15] At the beginning of the game, incumbent's quality $\theta_i \in [\underline{\theta}, \overline{\theta}]$ is given and is common knowledge. There is a single candidate (potential challenger), whose privately known quality θ_c is drawn from $[\underline{\theta}, \overline{\theta}]$ according to a commonly known distribution F with a con-

---

[15]Assuming the potential policy conflict away serves to emphasize the adverse selection dimension and dropping it would require significant alterations to the model, starting with media's position. One should also keep in mind that there are important elections in democracies in which competence and organizational ability of candidates are at the center stage, e.g. mayoral elections.

tinuous and full-support density f.[16] Candidate cares about the public good, as well as an office utility rent r she would receive if elected. Challenging requires undertaking a fixed and non-recoverable utility cost c_E that can be interpreted as some fixed campaign cost. Both the office rent and the candidacy cost affects candidate's utility in an additively separable fashion for simplicity. If the candidate decides not to challenge, then the game ends with the incumbent staying in the office. If she decides to challenge, then both the media and voters observe this and the game continues. First, an intermediary receiver; a "benevolent" media acts.[17] Caring only about the quality of the politician, the media decides whether to undertake an investigation by incurring an additive cost c_I. Investigation is a costly lottery that has a fixed probability Ψ of being successful. If it is successful, its result is costlessly transmitted to voters fully eliminating the uncertainty (e.g. it is bundled with TV content and everybody watches TV). If it fails, it generates no informative signal and nothing is transmitted to voters (e.g. media provides a neutral coverage of the elections). Ψ can be interpreted as media's journalistic strength, reflecting the extensiveness of its sources, quality of its political analysts etc.[18] It captures the quality detection (screening) power of the media and has the leading role in this chapter. Following the media stage, a continuum of voters (decisive receivers) who cannot abstain and who exclusively care about electing the best politician vote either in an informed, or in an uninformed fashion. Because the probability of a single voter affecting the outcome of election is nil, each individual voter will be indifferent in any equilibrium. I focus on equilibria characterized by sincere voting, which is an assumption commonly made in voting literature.[19] Since there is no policy conflict, this implies that all voters vote identically. All players are assumed to be risk-neutral. The timing

---

[16]No restriction is imposed on $\underline{\theta}, \overline{\theta}$, i.e. $\underline{\theta}, \overline{\theta} \in \mathbb{R}$ with $\underline{\theta} < \overline{\theta}$. Restricting them to be non-negative would be more suitable for a public good interpretation, whereas allowing them to take negative values permits for a potential public *bad* interpretation.

[17]Media can be interpreted either as an altruistic media outlet, or a stable coalition of altruistic media outlets who came up with an incentive-compatible cost sharing agreement to deal with potential free riding issues.

[18]Alternatively, media investigation can be successful with certainty and Ψ can be interpreted as the probability of the truth being transmitted to a decisive measure of voters. For instance, a very simple model of nonstrategic TV watching assuming no across-voter information transmission can be specified along the following lines. Assuming a continuum of measure 1 sincere voters who cannot abstain; each voter watches TV before the elections if $\epsilon > \delta$ (both stochastic and realizations of which are unknown to the media), where ϵ is some idiosyncratic entertainment benefit with distribution G_ϵ and δ is some common cost with distribution G_δ. Then the probability of more than half being aware of the true quality ordering at the time of voting is given by $G_\delta(G_\epsilon^{-1}(\frac{1}{2})) = \Psi$.

[19]Sincere voting implies that voters vote as if they had an infinitesimal chance of being pivotal. This is also called conditional sincerity. See Alesina and Rosenthal (2000).

structure of the game is summarized below.

1. **(t=0)** Nature picks candidate's quality θ_c from $[\underline{\theta}, \overline{\theta}]$ according to a distribution F.

2. **(t=1)** Candidate decides whether to challenge ($\tilde{C} = 1$) by incurring a fixed cost $c_E > 0$ or to stay out ($\tilde{C} = 0$). If she stays out, then the game ends and all players get the payoff θ_i.

3. **(t=2)** At the beginning of the period, media decides whether to investigate ($I = 1$) or not to investigate ($I = 0$) challenger's quality by incurring a fixed cost $c_I > 0$. If undertaken, investigation concludes at the end of the period; successfully revealing θ_c with probability Ψ, or failing to generate an informative signal with probability $1 - \Psi$.

4. **(t=3)** At the beginning of the period, elections occur and voters either retain the incumbent ($R_j = 1$) or replace her with the challenger ($R_j = 0$), where the subscript $j \in \{0, 1\}$ refers to voters' informational status depending on media's previous decision and outcome of the investigation if undertaken ($j = 1$ if uncertainty is resolved). At the end of the period, winner is determined and the public good, and thus the payoffs realize.

The following table gives players' *ex post* payoffs depending on the outcome.

	Candidate		Media		Voter
	Challenge	Don't	Investigate	Don't	
Incumbent Remains	$\theta_i - c_E$	θ_i	$\theta_i - c_I$	θ_i	θ_i
Challenger Wins	$\theta_c + r - c_E$		$\theta_c - c_I$	θ_c	θ_c

Table 1: Ex Post Payoffs

Let $\tilde{C}, I, R_0, R_1 \in [0, 1]$ be the (possibly mixed) actions taken by the candidate, media and voters respectively, with notations in line with the timing scheme described above.[20],[21] Below, I give the *ex ante* payoffs of players, which are the

---

[20]If mixed, they correspond to probabilities of respective actions, e.g. $I = 0.3$ correspond to a thirty percent probability of media investigation.

[21]Notice that there are two ways for the voter to remain uncertain about challenger's type: either the media remains passive, or it conducts an investigation but it fails. These occurrences

decision-relevant payoffs representing their expected prospects at the moment of choosing their actions.

Candidate: Consider a candidate with type θ_c. If she decides not to challenge, then her payoff is the commonly known incumbent quality, i.e.

$$V_c(\tilde{C} = 0; I, R_0, R_1 \mid \theta_c) = \theta_i. \tag{1}$$

If she challenges, then her *ex post* payoff will depend on the outcome of the election. So *ex ante*, given risk neutrality, candidate's (expected) payoff from challenging will be the average of the payoffs given in the first column of table 1, weighted by probabilities of loss and victory respectively. Victory is contingent on the realization of three mutually exclusive events: either the media does not investigate and voters blind-vote for the challenger, or the media does investigate but it fails and voters blind-vote for the challenger, or the media does investigate and the investigation successfully reveals challenger's quality and voters support the challenger knowingly. The probability of one of these occurring is given as follows.

$$(1 - I)(1 - R_0) + I(1 - \Psi)(1 - R_0) + \Psi I(1 - R_1)$$
$$= 1 - R_0 + \Psi I(R_0 - R_1). \tag{2}$$

So the expected payoff of the candidate from challenging is given by the following expression.

$$V_c(\tilde{C} = 1; I, R_0, R_1 \mid \theta_c) = [1 - R_0 + \Psi I(R_0 - R_1)](\theta_c + r)$$
$$+ [R_0 + \Psi I(R_1 - R_0)]\theta_i - c_E. \tag{3}$$

Allowing for mixed actions, the expressions in (1) and (3) can be summarized into the following expected payoff function, which will constitute the basis in the candidate's electoral participation decision.

$$V_c(\tilde{C}; I, R_0, R_1 \mid \theta_c) = \tilde{C}V_c(\tilde{C} = 1; I, R_0, R_1 \mid \theta_c)$$
$$+ (1 - \tilde{C})V_c(\tilde{C} = 0; I, R_0, R_1 \mid \theta_c). \tag{4}$$

---

correspond to different continuations and hence lead to different parts of the game tree. Since voters can observe whether the media has attempted an investigation or not, one might argue that these distinct continuations should be taken into account with an additional subscript. However, as the media has no additional information unavailable to the voters, the equilibrium concept I employ makes sure that they lead to identical beliefs and hence they are "decision-wise identical" occurrences. This point is further clarified in the next subsection.

Media: Given that the candidate has decided to challenge (which I indicate with "$C = 1$", to distinguish from the possibly mixed candidate action $\tilde{C}$) and taking into account the follow-up actions of the voters, media's (expected) payoff in case it decides not to investigate is given by the following.

$$V_M(I = 0; R_0 \mid C = 1) = R_0 \theta_i + (1 - R_0)\mathbb{E}_M[\theta_c \mid C = 1], \tag{5}$$

where $\mathbb{E}_M[\cdot]$ denotes media's expectation of candidate's type given the fact that the candidate decided to challenge.[22] If it decides to undertake the costly investigation, its (expected) payoff will not only depend on the subsequent voter behaviour, but also on the outcome of the investigation.

$$\begin{aligned} V_M(I = 1; R_0, R_1 \mid C = 1) = \Psi\mathbb{E}_M[R_1\theta_i + (1 - R_1)\theta_c \mid C = 1] \\ + (1 - \Psi)V_M(I = 0; R_0 \mid C = 1) - c_I. \end{aligned} \tag{6}$$

As before, allowing for mixing, media's payoff function can be written as follows.

$$\begin{aligned} V_M(I; R_0, R_1 \mid C = 1) = V_M(I = 1; R_0, R_1 \mid C = 1)I \\ + (1 - I)V_M(I = 0; R_0 \mid C = 1). \end{aligned} \tag{7}$$

Voters: If a voter faces no uncertainty at the time of voting, then retaining or replacing the incumbent yields θ_i or θ_c with certainty. So voter payoff given a successful media investigation is as follows.

$$V_v^1(R_1 \mid \theta_c, C = 1) = R_1\theta_i + (1 - R_1)\theta_c. \tag{8}$$

If on the other hand the media did not investigate or its investigation failed, then the following describes voters' expected payoffs.

$$V_v^0(R_0 \mid C = 1) = R_0\theta_i + (1 - R_0)\mathbb{E}_v[\theta_c \mid C = 1], \tag{9}$$

where the last term denotes voters' expectations regarding challenger's competence level.

Given these expected payoff functions which depend on actions, next subsection introduces the equilibrium concept, (optimal) strategies (history-contingent actions) and the resulting belief system which pins down the expectations in (5) and (9).

---

[22]This will depend on media's beliefs regarding challenger's type, which will be explained in detail in the next subsection.

2.2 Equilibrium Concept, Strategies and Beliefs

The equilibrium concept employed is perfect Bayesian equilibrium (henceforth PBE), as defined by Fudenberg and Tirole (1991).[23] PBE requires that given beliefs, each players' strategies should comply with sequential rationality. That is, given beliefs, actions of each player at each information set should maximize its expected payoff taking strategies of other players as given. Beliefs should satisfy several conditions.[24] First, in a signalling model such as this, they are defined over the type space, they are defined at each information set for each player, and they are common across symmetrically informed players.[25] Second, beliefs should be derived from prior beliefs and equilibrium strategies using Bayes' Rule "whenever possible". Formally, a PBE of the current model is a strategy profile $(\tilde{C}(\theta_c), I^*, R_0^*, R_1(\theta_c))$ such that:

$$\tilde{C}(\theta_c) \in \arg\max_{\tilde{C} \in [0,1]} V_c(\tilde{C}; I^*, R_0^*, R_1(\theta_c) \mid \theta_c), \tag{10}$$

$$I^* \in \arg\max_{I \in [0,1]} V_M(I; R_0^*, R_1(\theta_c) \mid C = 1), \tag{11}$$

$$R_0^* \in \arg\max_{R_0 \in [0,1]} V_v^0(R_0 \mid C = 1), \tag{12}$$

$$R_1(\theta_c) \in \arg\max_{R_1 \in [0,1]} V_v^1(R_1 \mid \theta_c, C = 1), \tag{13}$$

and a measurable belief system with a density satisfying:

$$\begin{aligned}
\beta(\theta_c \mid C = 1) &= \frac{\mathbb{P}(C = 1 \mid \theta_c, \tilde{C}(\theta_c)) f(\theta_c)}{\int_{\underline{\theta}}^{\overline{\theta}} \mathbb{P}(C = 1 \mid \tilde{\theta}_c, \tilde{C}(\tilde{\theta}_c)) f(\tilde{\theta}_c) \mathrm{d}\tilde{\theta}_c} \\
&= \frac{\tilde{C}(\theta_c) f(\theta_c)}{\int_{\underline{\theta}}^{\overline{\theta}} \tilde{C}(\tilde{\theta}_c) f(\tilde{\theta}_c) \mathrm{d}\tilde{\theta}_c},
\end{aligned} \tag{14}$$

---

[23]Strictly speaking, their definition of PBE covers finite games but with a continuous type space, the only practical difference is the application of a continuous Bayes' rule whenever possible. Finiteness is important for existence of equilibria, which can fail in some infinite games as shown by van Damme (1987). Since I show and study the equilibria, this is of no concern for this chapter. In addition, Manelli (1996) has shown that signalling games having a compact and infinite type space but finite message and action spaces such as the one presented here always have a PBE.

[24]These can be found in a more concise form in Fudenberg and Tirole (1991), page 332.

[25]Beliefs regarding types stand in contrast with beliefs regarding nodes as in sequential equilibrium. Beliefs being defined for all players at every information set implies that even if a player does not move at a given information set, he/she still has beliefs at that information set. The way Fudenberg and Tirole (1991) formalize this idea consists of assuming that all players "act" and hold beliefs after any history but those that do not really act at a given information set have empty action sets for any history leading to that particular information set. The final one implies in the current context that media and voters should have common beliefs after any history.

at histories on the equilibrium path.[26] Notice the lack of a subscript indicating whether the belief belongs to the media or the voter. This is a direct consequence of common beliefs requirement and holds true off the equilibrium path as well. An immediate implication of this is that in any equilibrium, expectations in (5) and (9) are equal. Regarding beliefs at histories off the equilibrium path, "whenever possible" implies several explicit restrictions, but two of them are especially relevant for the current model. First, they should be updates of beliefs at previous histories, and second is what Fudenberg and Tirole (1991) call the "no signalling what you don't know" condition, which stresses that the updating process should not be influenced by actions of other players (nor should it be influenced by any other randomization uncorrelated with types about which beliefs are formed) as long as these other players possess the same information with the player that is doing the updating.
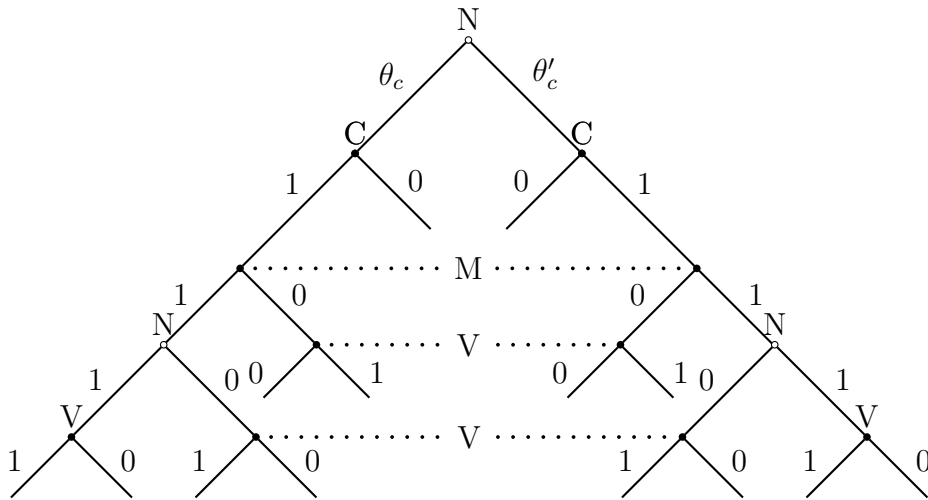
Figure 1: Simplified Extensive Form Representation.

To better understand the implications of these restrictions, consider the extensive form representation depicted in figure 1 with two types for simplicity. With a slight notational abuse, C refers to the candidate, N refers to nature, M and V refer to media and voters respectively. For sake of brevity, actions are denoted with 1 (challenge, investigate, retain the incumbent) and 0 (don't challenge, don't investigate, replace the incumbent), and payoffs are excluded. In addition, a successful investigation is denoted by 1 and a failed investigation is denoted by 0.

---

[26] Of course this formula is valid only if the set of challengers is of non-zero measure. In my model, there are possible equilibria with only $\overline{\theta}$ challenging with positive probability. In that case, $\beta(\theta_c \mid C = 1)$ would be degenerate at $\overline{\theta}$. It is possible to capture both degenerate and non-degenerate cases with a more general measure-theoretic notation using Lebesgue integration but I skip this.

If $\tilde{C}(\theta_c) = 0$ for all $\theta_c \in [\underline{\theta}, \overline{\theta}]$, then notice that (14) is undefined. Under such scenario, PBE imposes no restrictions on the beliefs at the first non-singleton information set (the one at which the media moves). Once they are determined however, PBE requires these beliefs to be carried over to subsequent non-singleton information sets (ones at which uninformed voters move) without change, because neither the pre-investigation media knows anything unknown to voters, nor the outcome of the investigation depends on challenger's quality.[27] Similarly, if in a PBE there exists some subset of types who challenge with positive probability, i.e. if the set of candidates who challenge with a positive probability:

$$\Omega = \left\{ \theta_c \in [\underline{\theta}, \overline{\theta}] : \tilde{C}(\theta_c) \in (0, 1] \right\}.$$

is non-empty, then all the beliefs at the three non-singleton information sets depicted in figure 1 are identical and are uniquely pinned down by (14), even though one of the information sets at which voters act will be off the equilibrium path when the media is playing a pure strategy.[28] A corollary of this belief identicalness in non-singleton information sets is that when constructing a PBE and considering voter strategies, one does not need to distinguish between different histories leading to voters remaining uninformed. The decision-relevant contingency from voter's perspective is whether the uncertainty persists or is resolved, as the differences in the manner by which the uncertainty persists do not lead to different beliefs and hence different expected payoffs. This is captured by subscripts 0 and 1 in (12) and (13).

Before moving on to presenting various equilibria, it is possible to cover more ground regarding strategies using the requirement of sequential rationality, which is valid under any PBE. Starting from the last stage, I do so below.

Voters: If media investigation is successful in revealing challenger's quality, voters would simply pick the best politician.

$$R_1(\theta_c) \in \begin{cases} \{0\}, & (\theta_c > \theta_i) \\ [0, 1], & (\theta_c = \theta_i) \\ \{1\}, & (\theta_c < \theta_i) \end{cases} . \tag{15}$$

---

[27]This would no longer be true, say, if investigation success depended on the quality gap between the incumbent and the challenger, capturing the idea of easier sorting when quality difference is large.

[28]Note that with continuum of types, only the number of singleton information sets in the extensive form representation are affected, i.e. the number of non-singleton information sets is still three.

If uncertainty persists for whatever reason, a voter strategy in any PBE should satisfy the following.

$$R_0^* \in \begin{cases} \{0\}, & (\mathbb{E}\left[\theta_c \mid C = 1\right] > \theta_i) \\ [0, 1], & (\mathbb{E}\left[\theta_c \mid C = 1\right] = \theta_i), \\ \{1\}, & (\mathbb{E}\left[\theta_c \mid C = 1\right] < \theta_i) \end{cases} \tag{16}$$

where the expectation will depend on beliefs that are pinned down uniquely from (14) if possible, and specified appropriately to support the equilibrium under consideration if not. Note that the subscript v in $\mathbb{E}_v$ is dropped due to commonality of beliefs.

Media: Inducting backwards, media's expected payoffs can be rewritten as follows after plugging in for the sequentially rational informed voting strategy.

$$V_M(I = 0; R_0^* \mid C = 1) = R_0^* \theta_i + (1 - R_0^*)\mathbb{E}\left[\theta_c \mid C = 1\right], \tag{17}$$

$$V_M(I = 1; R_0^*, R_1(\theta_c) \mid C = 1) = \Psi\mathbb{E}\left[\max\{\theta_i, \theta_c\} \mid C = 1\right]$$
$$+ (1 - \Psi)\left\{R_0^* \theta_i + (1 - R_0^*)\mathbb{E}\left[\theta_c \mid C = 1\right]\right\} - c_I. \tag{18}$$

Using (7), (17) and (18), and simplifying; I^* should maximize the following objective function.

$$V_M(I; R_0^*, R_1(\theta_c) \mid C = 1) = I\left[\Psi\mathbb{E}\left[\max\{\theta_i, \theta_c\} \mid C = 1\right] - c_I\right]$$
$$+ (1 - \Psi I)V_M(I = 0; R_0^* \mid C = 1). \tag{19}$$

After plugging in for the sequentially rational blind-voting strategy using (16) and rearranging, one can easily deduce that media's sequentially rational investigation policy under any PBE must satisfy the following.

$$I^* \in \begin{cases} \{0\}, & \left(\mathbb{E}\left[\theta_c \mid C = 1\right] < \theta_i \text{ and } \mathbb{E}\left[\max\{0, \theta_c - \theta_i\} \mid C = 1\right] < \frac{c_I}{\Psi}\right) \\ \{0\}, & \left(\mathbb{E}\left[\theta_c \mid C = 1\right] \geq \theta_i \text{ and } \mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] < \frac{c_I}{\Psi}\right) \\ [0, 1], & \left(\mathbb{E}\left[\theta_c \mid C = 1\right] < \theta_i \text{ and } \mathbb{E}\left[\max\{0, \theta_c - \theta_i\} \mid C = 1\right] = \frac{c_I}{\Psi}\right) \\ [0, 1], & \left(\mathbb{E}\left[\theta_c \mid C = 1\right] \geq \theta_i \text{ and } \mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] = \frac{c_I}{\Psi}\right) \\ \{1\}, & \left(\mathbb{E}\left[\theta_c \mid C = 1\right] < \theta_i \text{ and } \mathbb{E}\left[\max\{0, \theta_c - \theta_i\} \mid C = 1\right] > \frac{c_I}{\Psi}\right) \\ \{1\}, & \left(\mathbb{E}\left[\theta_c \mid C = 1\right] \geq \theta_i \text{ and } \mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] > \frac{c_I}{\Psi}\right) \end{cases}, \tag{20}$$

where expectations are again pinned down using (14) if possible. The form of the

investigation policy is quite intuitive. The marginal benefit of conducting a successful investigation is the expected quality gain from exposing inferior politicians who would be elected if voters were allowed to vote in an uninformed manner. For instance, if $\theta_i > \mathbb{E}[\theta_c \mid C = 1]$, then voters blind-vote for the incumbent, so incumbent is the politician who would be elected, and who would be the inferior one only if $\theta_c > \theta_i$. So the marginal benefit of a successful investigation under that scenario is $\mathbb{E}[\max\{0, \theta_c - \theta_i\} \mid C = 1]$. Taking into account the fact that the investigation might fail, its expected marginal benefit is therefore this expected quality gain multiplied by the probability of a successful investigation. If this exceeds the marginal cost of investigation (c_I), then the media investigates. Otherwise, it remains passive.

Candidate: Given the mutually sequentially rational strategies of the media and voters, start by considering the indifference condition of an arbitrary candidate with quality θ_c.

$$V_c(\tilde{C} = 1; I^*, R_0^*, R_1(\theta_c) \mid \theta_c) = \theta_i. \tag{21}$$

Using (3) and rearranging yields the following.

$$\theta_c = \theta_i - r + (1 - R_0^* + \Psi I^*(R_0^* - R_1(\theta_c)))^{-1} c_E = \hat{\theta}_c(\theta_c). \tag{22}$$

If $\theta_c > \hat{\theta}_c(\theta_c)$, then the candidate will strictly prefer challenging and vice versa. Notice that the value of $\hat{\theta}_c(\theta_c)$ will depend on the follow up strategies $I^*, R_0^*, R_1(\theta_c)$, as well as on payoff parameters θ_i, r, c_E. In particular, depending on whether the candidate is superior, or inferior (or equal) to the incumbent, $R_1(\theta_c)$ will take different values and (22) will define three distinct thresholds. This is because aside a victory by blind-voting, a superior challenger will win the election if the media performs a successful investigation, whereas an inferior challenger will lose in such situation. Using (15) to substitute for $R_1(\theta_c)$ in (22) yields three thresholds defining the decision rule used by the candidate in a PBE.

$$\hat{\theta}_c(\theta_c) = \begin{cases} \hat{\theta}_{c,w} = \theta_i - r + (1 - R_0^* + \Psi I^*(R_0^* - 1))^{-1} c_E, & (\theta_c < \theta_i) \\ \hat{\theta}_{c,e} = \theta_i - r + (1 - R_0^* + \Psi I^*(R_0^* - R_1(\theta_c)))^{-1} c_E, & (\theta_c = \theta_i), \\ \hat{\theta}_{c,b} = \theta_i - r + (1 - R_0^* + \Psi I^*(R_0^* - 0))^{-1} c_E, & (\theta_c > \theta_i) \end{cases} \tag{23}$$

where $R_1(\theta_c = \theta_i) \in [0, 1]$, as indicated in (15).[29] Using (23), it is possible to

---

[29]The subscripts w, e, b stands for "worse", "equal" and "better" respectively, referring to candidate's quality relative to the incumbent's.

express candidate's sequentially rational strategy in a concise form.

$$\tilde{C}(\theta_c) \in \begin{cases} \{0\}, & \left(\theta_c < \hat{\theta}_c(\theta_c)\right) \\ [0,1], & \left(\theta_c = \hat{\theta}_c(\theta_c)\right) \\ \{1\}, & \left(\theta_c > \hat{\theta}_c(\theta_c)\right) \end{cases}. \tag{24}$$

(24) is quite given candidate's risk neutrality. Consider for instance some candidate who is of better quality than the incumbent. If she challenges, she would win only if the media successfully uncovers her type, or if the voters blind-vote for her. Given the subsequent strategies, the probability of one of these occurring is exactly $(1 - (1 - \Psi I^*)R_0^*)$. Given this probability and the fact that she cares about the quality-dependent public good as well as the office rents, she would find it worthwhile to challenge only if her type exceeded $\hat{\theta}_{c,b}$. Although (24) is intuitive, it is in a somehow more complicated form then necessary, as it features several different thresholds.

Two observations make it possible to get a much simpler representation for candidate's PBE strategy. First, indifference thresholds in (23) themselves can be ordered, as for instance the probability of victory for a superior candidate is always greater then the probability of victory for an inferior candidate. Second, depending on equilibrium values of these thresholds, some of the cases implied by (24) will be infeasible, e.g. if $\hat{\theta}_{c,w} \geq \theta_i$, then one cannot have $\theta_c < \theta_i$ and $\theta_c > \hat{\theta}_{c,w}$ at the same time.

Lemma 1 (Connectedness of the set of strict challengers and uniqueness of the indifferent candidate): Consider an arbitrary PBE described by a strategy profile $(\tilde{C}(\theta_c), I^*, R_0^*, R_1(\theta_c))$. Let Ω_p denote the set of pure-strategy challengers, i.e.

$$\Omega_p = \left\{ \theta_c \in \left[\underline{\theta}, \overline{\theta}\right] : \theta_c > \hat{\theta}_c(\theta_c) \right\}.$$

Then Ω_p is connected. If it is non-empty, then its right end-point is $\overline{\theta}$. Furthermore, if, at this PBE, there exists a θ^* at which the candidate is indifferent, then θ^* is unique and $\Omega_p = \left(\theta^*, \overline{\theta}\right]$. Conversely, if Ω_p is non-empty and a strict subset of $\left[\underline{\theta}, \overline{\theta}\right]$, then it is a left-open interval with the unique indifferent challenger at its left boundary and $\overline{\theta}$ as its right end-point. That is, there exists a unique $\overline{\theta} > \theta^* > \underline{\theta}$ such that $\Omega_p = \left(\theta^*, \overline{\theta}\right]$ with $\theta^* = \hat{\theta}_c(\theta^*)$ and $\tilde{C}(\theta^*) \in [0,1]$, and $\theta < \hat{\theta}_c(\theta_c)$ for all $\theta < \theta^*$.[30]

---

[30]Note that empty set is a connected set. In addition, $\Omega_p = \left[\underline{\theta}, \overline{\theta}\right]$ with no indifferent type, and

Proof: Appendix. ∎

The idea behind lemma 1 is very simple. If a superior candidate decides to challenge in an equilibrium, then all candidate-types with a higher quality will challenge as well, as they would win with an equal probability and they would produce a higher amount of public good if elected. On the other hand, if an inferior candidate finds it worthwhile to challenge, then not only higher quality inferior candidates would challenge due to similar motives, but superior candidates would challenge too as they have the added benefit of an higher chance of electoral victory. The fact that there should be a unique indifferent type connecting non-challengers and challengers is just a restatement of candidate's strategy being upper hemi-continuous in her type. The usefulness of lemma 1 is that it allows expressing candidate's strategy by a single cutoff quality θ^* in any PBE with challenger entry.

Following lemma shows the "irrelevance" of the indifferent candidate-type's mixture, as well as of the strategy followed by informed voters when they find out that the challenger's quality is identical to incumbent's quality.

Lemma 2 (Anything goes for candidate and informed voter when they are indifferent): Consider a PBE defined by a belief system $\beta(\theta_c \mid C = 1)$ and a strategy profile $(\tilde{C}(\theta_c), I^*, R_0^*, R_1(\theta_c))$ and with $\Omega \neq \emptyset$, assigning a particular value $k \in [0, 1]$ to $R_1(\theta_c = \theta_i)$. Then varying k in $[0, 1]$ and leaving everything else intact (including beliefs), one gets a continuum of PBE. If there is an indifferent candidate θ^*, then the same is true for $\tilde{C}(\theta^*) \in [0, 1]$, with beliefs remaining intact *almost everywhere*.[31]

Proof: Appendix. ∎

When discussing various equilibria in what follows, I will sometimes use the word "unique". This should be understood as unique in the lemma 2 sense, referring to uniqueness of media and uninformed voting strategies and the set of pure strategy challengers. Relatedly, when presenting equilibria with challengers, I will omit $\tilde{C}(\theta^*)$ and R_1 from the description.

Finally, the following lemma shows that there is a common lower bound for the set of positive probability challengers in any equilibria.

---

two corner cases where the indifferent type is at the boundaries of the support, i.e. $\Omega_p = \left(\underline{\theta}, \overline{\theta}\right]$ ($\implies \Omega = \left[\underline{\theta}, \overline{\theta}\right]$ or $\Omega = \left(\underline{\theta}, \overline{\theta}\right]$) and $\Omega_p = \emptyset$ with $\Omega = \{\overline{\theta}\}$ (or $\Omega = \emptyset$) are possible. Notice that Ω can contain or exclude the indifferent type, depending whether she challenges with positive probability or not.

[31] For the latter case, if $\theta^* < \overline{\theta}$, $\beta(\theta^* \mid C = 1)$ will change accordingly but for $\theta > \theta^*$, $\beta(\theta \mid C = 1)$ will remain being the truncation of f at θ^*.

Lemma 3 (Lower bound for threshold quality): Let $\hat{\underline{\theta}} = \theta_i - (r - c_E)$. Then in any PBE, $\hat{\underline{\theta}} \leq \theta^* = \inf \Omega$.[32]

Proof: Inequalities (58)-(60) show that $\hat{\theta}_{c,b} \leq \hat{\theta}_{c,e} \leq \hat{\theta}_{c,w}$. From (23), varying I^* and R_0^*, the smallest value $\hat{\theta}_{c,b}$ can get is $\hat{\underline{\theta}}$. It follows that in any equilibrium with $\Omega \neq \emptyset$, $\hat{\underline{\theta}} \leq \hat{\theta}_c(\theta) \leq \theta$ for any $\theta \in \Omega$. If $\Omega = \emptyset$, then $\inf \Omega = +\infty$. ∎

Notice that $\theta_c + r - c_E$ is the expected payoff of the challenger if she would win the election with certainty. Lemma 3 then simply states that if the *ex post* payoffs do not justify running for office ($\theta_c + r - c_E < \theta_i$ or equivalently $\theta_c < \hat{\underline{\theta}}$), candidates will stay out. $\hat{\underline{\theta}}$ represents the full adverse selection outcome, and it is an important parameter playing a key role in determining what sort of equilibrium prevails.[33]

To summarize the discussion so far; (23) and (24) tell that in a PBE, candidate's strategy will depend on media and voter strategies. As the expressions in (15), (16) and (20) show, these will in turn depend on their beliefs regarding challenger's quality. Beliefs should in turn be consistent with candidate's strategy, so they act as the glue that brings everything together. Depending on parameter configurations and the freedom one has in specifying beliefs, there are a variety of equilibria. The media will be active in some of them, and remain passive in others. I start with relatively uninteresting ones, where the incumbent goes uncontested.

2.3 Dictatorial Equilibria

One implication of lemma 3 is that payoff parameters can completely deter entry. In particular, if costs of challenging are too high compared to office benefits, then no candidate-type would challenge at all.

Proposition 1 (PBE - Parametric dictatorship): If $\hat{\underline{\theta}} > \overline{\theta}$, then there exists a set of pooling PBE where no candidate-type challenges with positive probability ($\Omega = \emptyset$), media and voters hold whatever beliefs they want as long as these are common and identical across non-singleton information sets, and they act according to (15), (16) and (20) given those beliefs. Furthermore, there are no PBE with $\Omega \neq \emptyset$.

Proof: The last statement follows from lemma 3. For the first statement, sup-

---

[32] Note that lemma 1 does not say that θ^* is the indifferent type (although it is usually the indifferent type). It rather says that if an indifferent type exists, then it is unique and it is θ^*.

[33] Strictly speaking, $\max\left\{\theta, \hat{\underline{\theta}}\right\}$ is the full adverse selection outcome.

pose no candidate-type challenges with positive probability, and media and voters play arbitrary strategies with $R_1(\theta_c)$ satisfying (15). Since beliefs are free as long as they satisfy the criteria mentioned in the proposition (which follows from the definition of PBE, as discussed previously), neither media nor voters would have any incentive to deviate as long as beliefs are specified in a way ensuring that (16) and (20) holds for given strategies. Furthermore, given media and voter strategies, no candidate-type would have an incentive to deviate as they would either be worse off (with a positive probability of victory) or indifferent. ∎

For the remainder of the paper, it is assumed that $\hat{\underline{\theta}} \leq \overline{\theta}$. Even under this assumption, there are dictatorial equilibria with no challengers.[34]

Proposition 2 (PBE - Trivial dictatorship): For any parameter configuration satisfying $\hat{\underline{\theta}} \leq \overline{\theta}$, the following describes a continuum of pooling PBE: No candidate-type challenges ($\Omega = \emptyset$), and in case of a counterfactual challenge, media and voters hold common beliefs such that their strategies pinned down by (16) and (20) ensure that $\hat{\theta}_{c,b} > \overline{\theta}$ holds (where $\hat{\theta}_{c,b}$ is defined in (23)), and informed voters play according to (15).

Proof: Suppose media and uninformed voter strategies ensure that $\hat{\theta}_{c,b} > \overline{\theta}$. Then inequalities (58)-(60) imply that $\hat{\theta}(\theta) > \overline{\theta}$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, so no candidate-type challenges from (24). Conversely, suppose that no candidate-type challenges. Then $\Omega = \emptyset$ so beliefs are free. Therefore, one can always find beliefs which would ensure that the strategies implied by (16) and (20) under such beliefs would guarantee that $\hat{\theta}_{c,b} > \overline{\theta}$. One such example is beliefs that are degenerate at $\underline{\theta}$, which ensure that $I^* = 0$ and $R_0^* = 1$, so that no challenger has a positive chance of winning the elections.[35] ∎

The reason for labeling the no-challenge pooling equilibria described in proposition 2 as "trivial" is that their existence completely depends on the indeterminacy of beliefs off the equilibrium path and that they can occur under any parameter configuration.

---

[34]One can argue that the model is dictatorial by construction as even if $\Omega \neq \emptyset$, as long as $\Omega \neq [\underline{\theta}, \overline{\theta}]$, there will be instances (random draws by Nature) where the incumbent goes uncontested. However, in a game-theoretical sense, an equilibrium with $\Omega \neq \emptyset$ is not dictatorial on the equilibrium path. Moreover, it is not hard to establish a repeated game argument where the incumbent will always be challenged and the same equilibria will prevail. One trivial example is as follows. Suppose that the game is infinitely repeated at the pre-election stage but ends after the election. That is, the nature keeps drawing candidates sequentially until one of them challenges. In that case, if Nature's draws are independent over time and candidates discount fully in between consecutive draws, then the same set of equilibria will prevail.

[35]If $\theta_i = \underline{\theta}$, then these are the only possible beliefs.

Under the assumption that $\hat{\underline{\theta}} \leq \overline{\theta}$, given appropriate strategies, some candidate-types would find it worthwhile to challenge and might make voters better off in case they get elected. Furthermore, given that some candidate-types challenge, stronger candidates would have stronger incentives to challenge. This latter point suggests that upon observing a counterfactual challenger entry, it might be sensible for the media and voters to eliminate some types and form their out of equilibrium beliefs based on that elimination. This elimination can either be on the basis that some types would never challenge, or they would be "less likely" to challenge. These are the ideas behind the two common dominance based refinement criteria which are applied next, under which the pooling equilibria described in proposition 2 do not survive.

Lemma 4 (Equilibrium refinements): Assume $\hat{\underline{\theta}} \leq \overline{\theta}$. Cho and Kreps' (1987) Intuitive Criterion eliminates all trivial pooling equilibria if $\hat{\underline{\theta}} > \theta_i$ but it fails to eliminate them if $\hat{\underline{\theta}} \leq \theta_i$. On the other hand, Banks and Sobel's (1987) Divinity Criterion eliminates all trivial pooling equilibria with no challengers.

Proof: Appendix. ∎

Both of the criteria mentioned in lemma 4 first eliminate a subset of candidate-types on certain grounds of equilibrium dominance, then restrict off the equilibrium path beliefs to the undominated portion of the quality space. No further restrictions are imposed on beliefs. They then check whether there exists some candidate-types who would prefer to challenge under any such beliefs, i.e. given any media and voter best-responses implied by those beliefs. Both criteria coincide in the second step but differ in the way they refine the set of potential challengers. In particular, Intuitive Criterion (IC) compares the best payoff that a candidate can get by challenging to her equilibrium payoff θ_i. As such, it nullifies the possibility of counterfactual challengers arising from the set $[\underline{\theta}, \hat{\underline{\theta}})$, restricting off the equilibrium path beliefs to $[\hat{\underline{\theta}}, \overline{\theta}]$. If $\hat{\underline{\theta}} > \theta_i$, this is sufficient to eliminate all trivial equilibria, as in that case, any belief system constrained to $[\hat{\underline{\theta}}, \overline{\theta}]$ implies the challenger being always strictly superior to the incumbent. This makes an electoral victory a certainty, which incentivizes candidate-types with qualities above $\hat{\underline{\theta}}$ to challenge, destroying trivial equilibria. On the other hand, if $\hat{\underline{\theta}} \leq \theta_i$, then IC, say, allows for beliefs degenerate at $\hat{\underline{\theta}} \leq \theta_i$, which allows for media and voter best-responses implying a sure loss, preventing any candidate from challenging. This is where the Divinity Criterion (DC) comes into picture. DC contrasts sets of media and voter strategies under which candidates can be made better off by challenging compared to staying out and getting θ_i. In particular,

DC eliminates a candidate with quality θ if there exists another candidate-type θ' who is not only strictly better off under all media and voter strategies that make θ better off (or indifferent) conditional on a challenge, but who is also strictly better off under some strategies which leave θ strictly worse off. This captures the idea of some candidates being "more likely" to challenge, and eliminates all candidates in $[\hat{\underline{\theta}}, \theta_i]$, restricting off the equilibrium path beliefs to $(\theta_i, \overline{\theta}]$, i.e. to strictly superior politicians. Therefore, it eliminates all trivial pooling equilibria, including those that are missed by the IC.[36]

Next, I turn attention to more interesting equilibria with equilibrium challengers.

2.4 Equilibria With Challengers

This section presents equilibria where some candidate-types challenge and elections take place on the equilibrium path. Most of them are unique in the lemma 2 sense and whichever prevails depends on the particular configuration of payoff parameters. The presentation starts with less interesting ones with a passive media and moves gradually to equilibria with different levels of media activation.

To begin with, lemma 5 shows that beliefs have a very simple structure in any equilibrium with challengers.

Lemma 5 (Equilibrium beliefs): In any PBE with $\Omega \neq \emptyset$, equilibrium beliefs are described as follows when Ω is non-singleton.

$$\beta(\theta_c \mid C = 1) = \begin{cases} \tilde{C}(\theta_c) f(\theta_c \mid \theta^* \leq \theta_c), & (\theta_c = \theta^*) \\ f(\theta_c \mid \theta^* \leq \theta_c), & (\theta_c \in (\theta^*, \overline{\theta}]) \end{cases}, \tag{25}$$

where $\theta^* \in [\underline{\theta}, \overline{\theta})$ is either the unique candidate-type who is indifferent between challenging and staying out, or $\theta^* = \underline{\theta}$ and she strictly prefers to challenge. If Ω is singleton, then $\theta^* = \overline{\theta}$ and the beliefs are described as follows.

$$\beta(\theta_c \mid C = 1) = \begin{cases} 1, & (\theta_c = \overline{\theta}) \\ 0, & (\theta_c \in [\underline{\theta}, \overline{\theta})) \end{cases}. \tag{26}$$

Proof: Follows directly from (14) and lemma 1. ∎

As expected, connectedness of the challenger set implies that beliefs are simply

---

[36]This is reminiscent of how IC fails to eliminate some inefficient equilibria in a Spence model (Spence, 1973) with more than two types, whereas DC eliminates all but the *Riley equilibrium* (Riley, 1979).

truncated priors as long as the set of challengers is of non-zero measure. Note that one should not be too concerned with the exact belief concerning θ^* (except when it is the only challenger). Given a positive measure of challengers, this belief has no effect on relevant expectations, as the beliefs, being truncations of continuous priors, are continuous themselves.[37]

2.4.1 Equilibria With a Passive Media

When political payoffs are structured in a manner ensuring that only superior candidate-types challenge, media won't be needed at all.

Proposition 3 (PBE - Passive Media: No adverse selection): Assume $\theta_i \leq \hat{\underline{\theta}} \leq \overline{\theta}$. Then the following is the unique (in the lemma 2 sense) PBE with challengers:

$\Omega_p = (\theta^*, \overline{\theta}]$ with $\theta^* = \hat{\underline{\theta}}$; $I^* = 0$; $R_0^* = 0$.

Proof: First, assume that $\theta_i < \overline{\theta}$, and suppose that $\Omega \neq \emptyset$. Suppose $\Omega = \{\overline{\theta}\}$. Then $\beta(\overline{\theta} \mid C = 1) = 1$, so $\mathbb{E}[\theta_c \mid C = 1] = \overline{\theta}$ and $\mathbb{E}[\max\{\theta_i - \theta_c, 0\} \mid C = 1] = 0$, which imply $I^* = 0$ and $R_0^* = 0$. But then, (23), (24) and the fact that $\theta_i < \overline{\theta}$ imply that all candidate-types with $\theta_c \in [\theta_i, \overline{\theta})$ challenge, a contradiction. It follows that $\Omega = (\theta^*, \overline{\theta}]$ for some unique $\theta^* < \overline{\theta}$, and beliefs are pinned down by (25). Since f is full-support, this implies $\mathbb{E}[\theta_c \mid C = 1] > \theta_i$ and $\mathbb{E}[\max\{\theta_i - \theta_c, 0\} \mid C = 1] = 0$, so unique media and voter best-responses are $I^* = 0$ and $R_0^* = 0$ respectively. From (23) and (24), this implies that $\theta^* = \hat{\underline{\theta}} = \theta_i - (r - c_I)$. Next, assume that $\theta_i = \overline{\theta}$. Then it must be that $\hat{\underline{\theta}} = \theta_i = \overline{\theta}$ (otherwise, Ω would be empty). This implies $\beta(\overline{\theta} \mid C = 1) = 1$, so $\mathbb{E}[\theta_c \mid C = 1] = \overline{\theta} = \theta_i$ and $\mathbb{E}[\max\{\theta_i - \theta_c, 0\} \mid C = 1] = 0$, to which the unique media best-response is $I^* = 0$ and voter best-response is any $R_0^* \in [0, 1]$. Yet, from the expression for $\hat{\theta}_{c,e}$ in (23) and the fact that $I^* = 0$, the only voter strategy that would lead to a non-empty set of challengers is $R_0^* = 0$, in which case $\Omega = \{\overline{\theta}\}$ (and $\Omega_p = \emptyset$). ∎

Notice how $\hat{\underline{\theta}} \geq \theta_i$ implies that $r \leq c_E$. That is, office rents are no greater than costs of challenging. This implies that for any candidate-type who decides to challenge, public good considerations should dominate office rents because rents are not enough to compensate for costs of challenging even under a certain prospective victory. Consequently, any challenger should be quality-wise superior to the

---

[37]In fact, the specification in (25) implicitly assumes that receivers cannot distinguish between mixed and pure actions when they are played (i.e. they can only observe the realization if there is any mixing). If one were to assume the opposite, then (25) would be the straightforward truncation of f, independent of $\tilde{C}(\theta^*)$.

incumbent, which follows from lemma 3. In other words, whenever participation costs are higher then office rents, the mere act of challenging is a sufficiently strong signal of candidate quality, which allows uninformed voters to confidently elect the challenger. This eliminates the intervention need of the media against the possibility of voters making an *ex post* mistake. That is, the magnitude of challenging costs allow superior candidates to self-select without requiring media activity and there is no adverse selection.

When $\hat{\underline{\theta}} < \theta_i$, office rents exceed participation costs ($r > c_E$). This opens up the possibility of adverse selection, while at the same time indicates that challenging is not a sufficiently strong signal to imply superiority of the challenger with certainty. Yet in equilibrium, the challenger should still be better than the incumbent on average. This is because in any PBE, media and voters best-respond to their beliefs regarding challenger's quality, which turn out to be correct as they should be consistent with candidate's strategy.

Lemma 6 (Expected equilibrium non-inferiority of challengers): If $\hat{\underline{\theta}} < \theta_i$, then in any PBE with $\Omega \neq \emptyset$, beliefs should satisfy the following.

$$\mathbb{E}\left[\theta_c \mid C = 1\right] \geq \theta_i. \tag{27}$$

Proof: Suppose not, i.e. $\mathbb{E}\left[\theta_c \mid C = 1\right] < \theta_i$. Then from (16), $R_0^* = 1$. From (23), this implies $\hat{\theta}_{c,w} > \theta_i$. (24) then implies that no inferior ($\theta_c < \theta_i$) candidate-type challenges in equilibrium. But if $\Omega \neq \emptyset$, beliefs should then satisfy (27), a contradiction. ∎

Lemma 6 is very intuitive. A minimum requirement for a positive probability of an electoral victory for a challenger is voters having beliefs suggesting that she is better than incumbent on average. Voters in turn have these beliefs because they can improve average challenger quality by threatening to reelect the incumbent. One thing that lemma 6 does not imply is the signalling strength of challenging. It is rather the opposite. Because the trade-off between office rents and costs of challenging cannot prevent inferior candidate-types from challenging, media and voters become obliged to impose a threat-based selection upon them via their strategies, at least up to the point where the average challenger becomes non-inferior than the incumbent. A related result shows the limits to this strategic selection effect.

Lemma 7 (Limits to strategic selection): If $\hat{\underline{\theta}} < \theta_i$, then in any PBE with $\Omega \neq \emptyset$, the following should hold regarding the set of challengers.

1. $\theta_i = \overline{\theta} \implies \Omega_p = \emptyset$ and $\Omega = \{\overline{\theta}\}$.

2. $\theta_i = \underline{\theta} \implies \Omega_p = [\underline{\theta}, \overline{\theta}]$.

3. $\theta_i \in (\underline{\theta}, \overline{\theta}) \implies \Omega_p = (\theta^*, \overline{\theta}]$ with $\underline{\theta} \leq \theta^* < \theta_i$ or $\Omega_p = [\underline{\theta}, \overline{\theta}]$.

Proof: Appendix. ∎

The first two are boundary cases and are not very important. The last one, however, is important. It tells that whenever office rents exceed challenging costs, not only the signalling strength of a costly challenge is hindered due to a divergence between preferences of the candidate and the electorate, but also that this hindrance cannot be fully compensated by media and the voter efforts for inducing strategic selection. In simpler terms, if $r > c_E$, then inferior challengers cannot be fully excluded. The intuition behind this result will be clearer subsequently, but the main reason should be more or less clear: costly and imperfect technology of journalism. Although voters can induce some improvement in selection by sometimes blind-voting for the incumbent, they are constrained by the consistency of their beliefs: if they were to exclude all inferior challengers, say, by fully committing to reelect the incumbent in case of remaining uninformed (under such commitment, challengers win only under a successful investigation, i.e. only if they are superior to the incumbent), then they would no longer have the incentive to do so, as only superior candidate-types would challenge.

Corollary 1 (Impossibility of unconditional commitment to the incumbent): If $\hat{\underline{\theta}} < \theta_i$, then in any PBE with $\Omega \neq \emptyset$, $R_0^* < 1$.

This is precisely the reason voters need the media for: to mop up inferior challengers missed due to incredibility of their voting threat. Yet, because investigation is costly and sometimes fails, media itself is limited in the extent of its willingness to attempt in exposing inferior challengers.

Before moving on to presenting other equilibria, I will further elaborate on the equilibrium best-response correspondences of the media and voters. This will not only help with the presentation of equilibria later on, but will also facilitate comparative statics. First, notice that lemma 1 and lemma 5 imply that one can limit attention to simple truncated expectations when considering voter best-responses. This gives a simple cutoff rule that characterizes uninformed voters' decisions under any equilibrium with challengers.

Lemma 8 (Equilibrium voter best-response under uncertainty): Consider an equilibrium with $\Omega \neq \emptyset$. If $\theta_i < \mathbb{E}[\theta_c]$, then $R_0^* = 0$. If $\theta_i \geq \mathbb{E}[\theta_c]$, then

there exists a unique $\theta_V(\theta_i)$ such that the following holds.

$$
R_0^* \in \begin{cases} \{0\}, & (\theta^* > \theta_V(\theta_i)) \\ [0,1], & (\theta^* = \theta_V(\theta_i)) , \\ \{1\}, & (\theta^* < \theta_V(\theta_i)) \end{cases} \tag{28}
$$

where $\theta_V(\theta_i) \in [\underline{\theta}, \theta_i]$ with $\theta_V(\theta_i) \in \{\underline{\theta}, \theta_i\}$ only if $\theta_i = \mathbb{E}[\theta_c]$ or if $\theta_i = \overline{\theta}$ respectively, and $\theta^* = \inf \Omega$.

Proof: Suppose that $\Omega \neq \emptyset$ and define $\theta^* = \inf \Omega$. Notice that lemma 5 implies the following.

$$
\mathbb{E}[\theta_c \mid C = 1] = \mathbb{E}[\theta_c \mid \theta^* \leq \theta_c]. \tag{29}
$$

It follows that as f is full-support, $\mathbb{E}[\theta_c \mid C = 1]$ is strictly increasing in θ^* as long as $\theta^* < \overline{\theta}$. First, assume that $\theta_i < \mathbb{E}[\theta_c]$. Then the following holds.

$$
\theta_i < \mathbb{E}[\theta_c] \leq \mathbb{E}[\theta_c \mid C = 1], \tag{30}
$$

where the second inequality is an equality only if $\theta^* = \underline{\theta}$. Using (16), this immediately implies that $R_0^* = 0$. Next, assume that $\theta_i \geq \mathbb{E}[\theta_c]$. Consider the following equation.

$$
\theta_i - \mathbb{E}[\theta_c \mid \theta_V \leq \theta_c] = 0. \tag{31}
$$

Since the left-hand side is strictly decreasing in θ_V, the fact that $\theta_V(\theta_i) \in \{\underline{\theta}, \theta_i\}$ whenever $\theta_i = \mathbb{E}[\theta_c]$ or $\theta_i = \overline{\theta}$ respectively immediately follows. Now fix a θ_i satisfying $\mathbb{E}[\theta_c] < \theta_i < \overline{\theta}$. If $\theta_V = \underline{\theta}$, then the left-hand side of (31) is strictly positive and if $\theta_V = \theta_i$, it is strictly negative. Moreover, since the density f is continuous, the conditional expectation in (31) is continuous in θ_V when θ_V is in $[\underline{\theta}, \overline{\theta})$ from the fundamental theorem of calculus. Combining this with the fact that the left-hand side is strictly decreasing in θ_V, the intermediate value theorem implies that there exists a unique $\theta_V(\theta_i) \in (\underline{\theta}, \theta_i)$ solving (31). (28) directly follows from this and (29), along with (16). ∎

In a signalling model, when one is using PBE as an equilibrium concept, best-responses of receivers are defined as best-responses to their beliefs. In (16), those beliefs were embedded into the conditional expectation, which is the decision-relevant statistic from a voter's perspective. Using lemma 1 and its follow-up

lemma (lemma 5) allows one to summarize beliefs in any PBE with challengers using a single number: $\theta^* = \inf \Omega$. Whether this number implies that average challenger is superior to the incumbent depends not only on the incumbent's quality, but also on the shape of f. The exact requirements are buried within $\theta_V(\theta_i)$, which has a nice interpretation: it is the minimum challenger quality that voters are willing to accept before they commit on reelecting the incumbent in case they remain uninformed. Combining lemma 8 with corollary 1 results in the following.

Corollary 2 (Another lower bound for threshold quality): If $\hat{\underline{\theta}} < \theta_i$, then in any PBE with $\Omega \neq \emptyset$, one must have $\theta_V(\theta_i) \leq \theta^* = \inf \Omega$.

Corollary 2 tells that *ex ante*, voters get what they want in equilibrium. More precisely, given the monotonicity of candidate's participation decision and their own risk-neutrality, they are always satisfied on average in equilibrium. A similar exercise can be carried out for the media's decision.

Lemma 9 (Equilibrium media best-response): Consider an equilibrium with $\Omega \neq \emptyset$ and $\hat{\underline{\theta}} < \theta_i$. For any θ_i, there exists a unique unit cost $K(\theta_i)$ such that when $\frac{c_I}{\Psi} > K(\theta_i)$, $I^* = 0$, and when $\frac{c_I}{\Psi} \leq K(\theta_i)$, there exists a unique $\theta_M(\Psi, c_I, \theta_i)$ such that the following holds.

$$
I^* \in \begin{cases} \{0\}, & (\theta^* > \theta_M(\Psi, c_I, \theta_i)) \\ [0,1], & (\theta^* = \theta_M(\Psi, c_I, \theta_i)) \\ \{1\}, & (\theta^* < \theta_M(\Psi, c_I, \theta_i)) \end{cases} \tag{32}
$$

where $\theta_M(\Psi, c_I, \theta_i) \in [\underline{\theta}, \theta_i]$, with $\theta_M(\Psi, c_I, \theta_i) \in \{\underline{\theta}, \theta_i\}$ only if $\frac{c_I}{\Psi} = K(\theta_i)$ or $c_I = 0$ respectively, and $\theta^* = \inf \Omega$.

Proof: Suppose $\Omega \neq \emptyset$ and $\hat{\underline{\theta}} < \theta_i$. Let $\theta^* = \inf \Omega$. Since $\hat{\underline{\theta}} < \theta_i$, lemma 6 immediately implies that the media's equilibrium strategy given in (20) can be simplified as follows.

$$
I^* \in \begin{cases} \{0\}, & \left(\mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] < \frac{c_I}{\Psi}\right) \\ [0,1], & \left(\mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] = \frac{c_I}{\Psi}\right) \\ \{1\}, & \left(\mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] > \frac{c_I}{\Psi}\right) \end{cases} \tag{33}
$$

Consider the conditional expectation in (33). Lemma 5 allows one to simplify it as follows.

$$\mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] = \int_{\underline{\theta}}^{\theta_i} (\theta_i - \theta_c) f(\theta_c \mid C = 1)\, d\theta_c,$$

$$= \int_{\underline{\theta}}^{\theta_i} (\theta_i - \theta_c) f(\theta_c \mid \theta^* \leq \theta_c)\, d\theta_c,$$

$$= \frac{1}{1 - F(\theta^*)} \int_{\theta^*}^{\theta_i} (\theta_i - \theta_c) f(\theta_c)\, d\theta_c,$$

$$= \frac{F(\theta_i) - F(\theta^*)}{1 - F(\theta^*)} \{\theta_i - \mathbb{E}\left[\theta_c \mid \theta^* \leq \theta_c < \theta_i\right]\}. \quad (34)$$

Now consider the following equation.

$$\frac{F(\theta_i) - F(\theta_M)}{1 - F(\theta_M)} \{\theta_i - \mathbb{E}\left[\theta_c \mid \theta_M \leq \theta_c < \theta_i\right]\} - \frac{c_I}{\Psi} = 0. \quad (35)$$

The fundamental theorem of calculus implies that this function is continuous on $[\underline{\theta}, \theta_i]$ and differentiable on $(\underline{\theta}, \theta_i)$.[38] Differentiating the left-hand side with respect to θ_M, one gets the following.

$$\frac{f(\theta_M)(F(\theta_i) - 1)}{(1 - F(\theta_M))^2} \{\theta_i - \mathbb{E}\left[\theta_c \mid \cdot\right]\} - \frac{F(\theta_i) - F(\theta_M)}{1 - F(\theta_M)} \frac{\partial}{\partial \theta_M} \mathbb{E}\left[\theta_c \mid \cdot\right] < 0, \quad (36)$$

for $\theta_M \in (\underline{\theta}, \theta_i)$. It follows that (35) is strictly decreasing in θ_M when θ_M is in $[\underline{\theta}, \theta_i]$. This implies that for a given θ_i, $\mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right]$ will be at its lowest when $\theta^* = \underline{\theta}$. So there exists a unique unit cost $K(\theta_i)$ satisfying:

$$F(\theta_i)\{\theta_i - \mathbb{E}\left[\theta_c \mid \underline{\theta} \leq \theta_c < \theta_i\right]\} - K(\theta_i) = 0, \quad (37)$$

such that whenever $\frac{c_I}{\Psi} > K(\theta_i)$, left-hand side of (35) is strictly negative for all θ_M and whenever $\frac{c_I}{\Psi} = K(\theta_i)$, only $\theta_M = \underline{\theta}$ satisfies (35). So suppose that $\frac{c_I}{\Psi} < K(\theta_i)$. Then whenever evaluated at $\underline{\theta}$, left-hand side of (35) is strictly positive, and when $\theta_M \to \theta_i$, left-hand side goes to $-\frac{c_I}{\Psi}$. From the intermediate value theorem and (33), the result follows. ∎

Note that unlike lemma 8, the assumption $\underline{\hat{\theta}} < \theta_i$ is necessary to obtain lemma 9.[39] This allows one to invoke lemma 6, which identifies challenger as the potentially inferior politician to be elected in case the uncertainty does not resolve. This is because corollary 2 implies that if voters remain uncertain, they will "weakly" elect the challenger in any equilibrium with $\underline{\hat{\theta}} < \theta_i$. Other than that, the general logic remains the same. Lemma 1 and the fact that it implies truncated beliefs

---

[38]This can be more clearly seen from the third line of (34).
[39]Imposing this assumption to lemma 8 resulted in corollary 2.

in any equilibrium with challengers allow to embed beliefs into a single number θ^*. The decision-relevant statistic $\mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right]$, as well as the functional form of the prior distribution then determine the form of the investigation threshold $\theta_M(\cdot)$. Notice that this threshold not only depends on incumbent's quality, but also on the cost and strength of the media's investigation technology. $\theta_M(\cdot)$ is the minimum challenger quality that the media is willing to accept before "waking up". Intuitively, one would expect that a decrease in investigation cost or an increase in its strength would make the media more selective, which is indeed true as will be seen later on. However, investigation cost and journalistic strength are not isomorphic in their equilibrium effects as one would expect, which will also be seen later on.

Besides sufficient self-selection, there are two ways media can remain passive in equilibrium. I start with the most obvious one, already mentioned in lemma 9.

Proposition 4 (PBE - Passive media: Prohibitive costs): Assume that $\hat{\underline{\theta}} < \theta_i$ and $\Omega \neq \emptyset$. Assume further that $\frac{c_I}{\Psi} > K(\theta_i)$. Define $\theta^* = \inf \Omega$. Then the following equilibria prevail and they are unique in the lemma 2 sense.

1. If $\theta_i < \mathbb{E}[\theta_c]$, then $I^* = R_0^* = 0$ with $\theta^* = \max\{\underline{\theta}, \hat{\theta}\}$.

2. If $\theta_i \geq \mathbb{E}[\theta_c]$ and $\theta_V(\theta_i) \leq \hat{\underline{\theta}}$, then $I^* = R^* = 0$ with $\theta^* = \max\{\underline{\theta}, \hat{\underline{\theta}}\}$.

3. If $\theta_i \geq \mathbb{E}[\theta_c]$ and $\theta_V(\theta_i) > \hat{\underline{\theta}}$, then $I^* = 0$. Furthermore, there exists a unique $R_0^* \in \left(0, 1 - \frac{c_E}{r}\right]$ such that $\theta^* = \theta_V(\theta_i)$.

Proof: Suppose $\hat{\underline{\theta}} < \theta_i$, $\frac{c_I}{\Psi} > K(\theta_i)$ and $\Omega \neq \emptyset$. $I^* = 0$ follows directly from lemma 9.

1. If $\theta_i < \mathbb{E}[\theta_c]$, then this follows directly from lemma 8, the fact that $I^* = 0$ and (23)-(24).

2. Assume $\theta_i \geq \mathbb{E}[\theta_c]$ and $\theta_V(\theta_i) \leq \hat{\underline{\theta}}$ and suppose instead $\theta^* > \max\{\underline{\theta}, \hat{\theta}\}$. Since $\theta_V(\theta_i) \leq \hat{\underline{\theta}}$, lemma 8 directly implies $R_0^* = 0$. But then (23)-(24), along with $I^* = 0$ imply that $\theta^* = \max\{\underline{\theta}, \hat{\theta}\}$, a contradiction. Moreover, if $\theta^* = \max\{\underline{\theta}, \hat{\theta}\}$ ($\theta^* < \hat{\underline{\theta}}$ is not allowed due to lemma 3), $R_0^* = 0$ is the unique blind-voting strategy (given $I^* = 0$) consistent with this fact, which follows again from (23)-(24).

3. Assume $\theta_i \geq \mathbb{E}[\theta_c]$ and $\theta_V(\theta_i) > \hat{\underline{\theta}}$ and suppose instead $\theta^* > \theta_V(\theta_i)$. From lemma 8, this implies that $R_0^* = 0$. But then, (23)-(24) imply

$\theta^* = \max\{\underline{\theta}, \hat{\theta}\}$. If $\hat{\theta} \geq \underline{\theta}$, this is a direct contradiction. If not, then it is a contradiction because lemma 8 has shown that $\underline{\theta} \leq \theta_V(\theta_i)$. Combining this with corollary 2 yields $\theta^* = \theta_V(\theta_i)$. From (28), this implies that voters are indifferent when uninformed. But given $I^* = 0$ and the expressions in (23), there exists a unique R_0^* that solves the following:

$$\theta_V(\theta_i) = \theta_i - r + (1 - R_0^*)^{-1} c_E, \tag{38}$$

where the range $R_0^* \in \left(0, 1 - \frac{c_E}{r}\right]$ follows from the range given in lemma 8: $\theta_V(\theta_i) \in [\underline{\theta}, \theta_i]$. Note that $\hat{\theta} < \theta_i$ implies $\frac{c_E}{r} < 1$. ∎

Other than stating the most obvious reason for media passivity, proposition 4 also offers insights on the selection mechanism used by receivers to improve challenger quality. When the media is out of the picture due to too high costs or a too weak investigation technology, it is up to voters to improve challenger selection by exploiting the self-fulfilling nature of their rational expectations. In the first equilibria, they never want to reelect the incumbent because her quality is so low that even if all candidate-types were challenging, they would still be better off replacing the incumbent with one of them.[40] The second equilibria is a case of sufficient self-selection, not sufficient in the sense that it ensures *ex post* superiority of the challenger if she were to be elected as in proposition 3, but sufficient in the sense that the lowest quality candidate-type that would be elected is better than the lowest quality which voters are willing to face. As shown later on, a low incumbent quality would facilitate the prevalence of such "low standards". In the final equilibria, voters' standards are high enough, so that they induce some improvement in challenger selection by threatening to sometimes vote for the incumbent when they are uninformed. The next set of equilibria is related to this idea of standards.

Proposition 5 (PBE - Passive media: Low media standard): Assume that $\hat{\theta} < \theta_i$ and $\Omega \neq \emptyset$. Assume further that $\frac{c_I}{\Psi} \leq K(\theta_i)$. Define $\theta^* = \inf \Omega$. Then the following equilibria prevail and they are unique in the lemma 2 sense.

1. If $\theta_i < \mathbb{E}[\theta_c]$ and $\theta_M(\Psi, c_I, \theta_i) \leq \hat{\theta}$, then $I^* = R_0^* = 0$ and $\theta^* = \max\{\underline{\theta}, \hat{\theta}\}$.

2. If $\theta_i \geq \mathbb{E}[\theta_c]$ and $\max\{\theta_V(\theta_i), \theta_M(\Psi, c_I, \theta_i)\} \leq \hat{\theta}$, then $I^* = R_0^* = 0$ and $\theta^* = \max\{\underline{\theta}, \hat{\theta}\}$.

---

[40]Note that I am referring to these as "equilibria", because they are only unique in the lemma 2 sense.

3. If $\theta_i \geq \mathbb{E}[\theta_c]$, $\theta_M(\Psi, c_I, \theta_i) < \theta_V(\theta_i)$ and $\theta_V(\theta_i) > \hat{\underline{\theta}}$, then $I^* = 0$. Furthermore, there exists a unique $R_0^* \in \left(0, 1 - \frac{c_E}{r}\right]$ such that $\theta^* = \theta_V(\theta_i)$.

Proof: Suppose $\hat{\underline{\theta}} < \theta_i$, $\frac{c_I}{\Psi} \leq K(\theta_i)$ and $\Omega \neq \emptyset$.

1. Assume $\theta_i < \mathbb{E}[\theta_c]$ and $\theta_M(\Psi, c_I, \theta_i) \leq \hat{\underline{\theta}}$ and suppose instead $\theta^* > \max\{\underline{\theta}, \hat{\underline{\theta}}\}$. Since $\theta_i < \mathbb{E}[\theta_c]$, lemma 8 implies $R_0^* = 0$. Since $\theta_M(\Psi, c_I, \theta_i) \leq \hat{\underline{\theta}}$, lemma 9 implies $I^* = 0$. But then (23)-(24) imply that $\theta^* = \max\{\underline{\theta}, \hat{\underline{\theta}}\}$, a contradiction. Combining this with lemma 3 then implies $\theta^* = \max\{\underline{\theta}, \hat{\underline{\theta}}\}$. Given $R_0^* = 0$, $I^* = 0$ is the unique investigation strategy consistent with this fact, which follows again from (23)-(24).

2. Assume $\theta_i \geq \mathbb{E}[\theta_c]$ and $\max\{\theta_V(\theta_i), \theta_M(\Psi, c_I, \theta_i)\} \leq \hat{\underline{\theta}}$ and suppose instead that either $I^* = 0$ or $R^* = 0$ (or both). (23)-(24), along with lemma 1 then implies that $\theta^* > \hat{\underline{\theta}}$. But then, lemmas 8 and 9 imply that $R_0^* = 0$ and $I^* = 0$, a contradiction. It follows that in equilibrium, one must have $I^* = R_0^* = 0$ and the only θ^* consistent with beliefs which would justify these media and voter strategies is $\theta^* = \max\{\underline{\theta}, \hat{\underline{\theta}}\}$.

3. Assume $\theta_i \geq \mathbb{E}[\theta_c]$, $\theta_M(\Psi, c_I, \theta_i) < \theta_V(\theta_i)$ and $\theta_V(\theta_i) > \hat{\underline{\theta}}$. Suppose instead $\theta^* > \theta_V(\theta_i)$. Then by assumption, we also have $\theta^* > \theta_M(\Psi, c_I, \theta_i)$. From lemmas 8 and 9, these imply $R_0^* = 0$ and $I^* = 0$. But then, (23)-(24) imply $\theta^* = \max\{\underline{\theta}, \hat{\underline{\theta}}\} \leq \theta_V(\theta_i)$, a contradiction. Combining this with corollary implies that in equilibrium, one must have $\theta^* = \theta_V(\theta_i)$ so that voters are indifferent when uninformed. Furthermore, since $\theta_M(\Psi, c_I, \theta_i) < \theta_V(\theta_i)$, one must also have $I^* = 0$ from lemma 9. Finally from (23), there is a unique R_0^* satisfying (38), where the range $R_0^* \in \left(0, 1 - \frac{c_E}{r}\right]$ follows as before. ∎

Proposition 5 reveals two additional reasons for equilibrium media passivity besides sufficient self-selection and *a priori* prohibitive unit costs: absolutely, and relatively low media standard. The first two cases correspond to equilibria in which political incentives provided by *ex post* net office benefits are not too high relative to media's standard, so that the lowest candidate-type who would challenge under a certain victory is higher then the lowest type media is willing to accept before considering to use its journalistic tools. In the last case, media passivity arises due to its standard being low relative to voters' standards. Foreseeing that voters who are left to their own deductive means would already provide better selection (higher minimum challenger competence) compared to its own standard, it remains silent, choosing to avoid costs and risks associated with investigation. In other words, voters are "pickier", and knowing this, media stays

passive. Notice that within the current boundaries of the model, both absolute and relative media standards ultimately depend on two things. First, quality of the incumbent along with the specific unconditional distribution of types. Since the media and voters use different decision statistics, these would affect their cutoff qualities differently and hence could affect the ordering of their standards. For instance, an immediate look at equations (31) and (37) implies that if the quality distribution were to gain skewness to the left, i.e. started putting a higher mass on higher qualities, then *ceteris paribus*, media would have more reason to become active. This might seem counter-intuitive but it is the logical implication of the media being benevolent. If, conditional on a challenge threshold, average quality of the challenger were to increase, then voters would not be so aggressive in their incumbent retention threats. This would give the media more reason to be active, as there would be more *ex post* inferior challengers to expose. Second, media's investigation costs and its journalistic strength are the main determinants of the lowest quality challenger palatable to media before engaging in an investigation. These will be studied in more detail later on.

2.4.2 Equilibria With an Active Media

For presenting the equilibria in this section, it is useful to define the following parameter.

$$\hat{\underline{\theta}}_\Psi = \theta_i - \left(r - (1 - \Psi)^{-1}c_E\right). \tag{39}$$

$\hat{\underline{\theta}}_\Psi$ is the quality of the worst challenger when $R_0 = 0$ and $I = 1$, i.e. it is the maximin challenger quality that the media can achieve given that voters always elect the challenger if they remain uncertain. It reflects the maximum reach of media's journalistic powers, given that it receives no help from (uncertain) voters regarding selection. Notice that $\hat{\underline{\theta}} \leq \hat{\underline{\theta}}_\Psi$, with the inequality being strict unless $\Psi = 0$. As indicated previously, the main force incentivizing the media towards being active is its standard. Parameter configurations under which this standard will be high enough can be broadly classified under two categories. I start with the more obvious one.

Proposition 6 (PBE - Active media: Below average incumbent): Assume that $\hat{\underline{\theta}} < \theta_i$, $\frac{c_I}{\Psi} \leq K(\theta_i)$, and $\theta_M(\Psi, c_I, \theta_i) > \hat{\underline{\theta}}$, along with $\Omega \neq \emptyset$. Assume further that $\theta_i < \mathbb{E}[\theta_c]$. Define $\theta^* = \inf \Omega$. Then the following equilibria prevail and they are unique in the lemma 2 sense.

1. If $\Psi < 1 - \frac{c_E}{r}$ and $\theta_M(\Psi, c_I, \theta_i) < \hat{\underline{\theta}}_\Psi$, then $R_0^* = 0$ and there exists a unique

$I^* \in (0,1)$ such that $\theta^* = \theta_M(\Psi, c_I, \theta_i)$.

2. If $\Psi < 1 - \frac{c_E}{r}$ and $\theta_M(\Psi, c_I, \theta_i) \geq \hat{\underline{\theta}}_\Psi$, then $R_0^* = 0$, $I^* = 1$ and $\theta^* = \hat{\underline{\theta}}_\Psi$.

3. If $\Psi \geq 1 - \frac{c_E}{r}$, then $R_0^* = 0$ and there exists a unique $I^* \in (0, \Psi^{-1} - \frac{c_E}{\Psi r})$ such that $\theta^* = \theta_M(\Psi, c_I, \theta_i)$.

Proof: Suppose $\hat{\underline{\theta}} < \theta_i$, $\frac{c_I}{\Psi} \leq K(\theta_i)$, $\theta_M(\Psi, c_I, \theta_i) > \hat{\underline{\theta}}$, $\Omega \neq \emptyset$ and $\theta_i < \mathbb{E}[\theta_c]$. $R_0^* = 0$ directly follows from lemma 8.

1. Assume $\Psi < 1 - \frac{c_E}{r}$ and $\theta_M(\Psi, c_I, \theta_i) < \hat{\underline{\theta}}_\Psi$. Note that the former assumption implies $\hat{\underline{\theta}}_\Psi < \theta_i$ (and also note that this equilibrium could not occur when $c_I = 0$ since this would imply from lemma 9 that $\theta_M(\cdot) = \theta_i$). Suppose that $\theta^* < \theta_M(\Psi, c_I, \theta_i)$. Then from lemma 8, we have $I^* = 1$. Since $R_0^* = 0$, this implies from (23) that $\theta^* = \hat{\underline{\theta}}_\Psi > \theta_M(\Psi, c_I, \theta_i)$, a contradiction. Suppose instead $\theta^* > \theta_M(\Psi, c_I, \theta_i)$. Then from lemma 8, we have $I^* = 0$. Since $R_0^* = 0$, this implies from (23) that $\theta^* = \hat{\underline{\theta}} < \theta_M(\Psi, c_I, \theta_i)$, a contradiction. It follows that in equilibrium, one must have $\theta^* = \theta_M(\Psi, c_I, \theta_i)$. Given $R_0^* = 0$ and the expressions in (23), there exists a unique I^* that satisfies the following:

$$\theta_M(\Psi, c_I, \theta_i) = \theta_i - r + (1 - \Psi I^*)^{-1} c_E, \tag{40}$$

where the range $I^* \in (0, 1)$ is dictated from the fact that at this range, $\theta_M(\Psi, c_I, \theta_i) \in (\hat{\underline{\theta}}, \hat{\underline{\theta}}_\Psi)$.

2. Assume $\Psi < 1 - \frac{c_E}{r}$ and $\theta_M(\Psi, c_I, \theta_i) \geq \hat{\underline{\theta}}_\Psi$. Since $R^* = 0$, in any equilibrium, one must have $\theta^* \in [\hat{\underline{\theta}}, \hat{\underline{\theta}}_\Psi]$. This is because as I^* is varied in $[0, 1]$, that is the resulting range for θ^* from (23). So for any $\theta^* \neq \hat{\underline{\theta}}_\Psi$, one would get $I^* = 1$ from lemma 9, which would result in a contradiction of the form $\theta^* = \hat{\underline{\theta}}_\Psi$. It follows that $I^* = 1$ and $\theta^* = \hat{\underline{\theta}}_\Psi$.

3. Assume $\Psi \geq 1 - \frac{c_E}{r}$. Note that with this assumption; $\hat{\underline{\theta}}_\Psi \geq \theta_i$. Suppose instead that $\theta^* > \theta_M(\Psi, c_I, \theta_i)$. Then lemma 9 implies $I^* = 0$. This, along with (23)-(24) and the fact that $R_0^* = 0$ implies that $\theta^* = \hat{\underline{\theta}} < \theta_M(\Psi, c_I, \theta_i)$, a contradiction. Suppose instead $\theta^* < \theta_M(\Psi, c_I, \theta_i)$. Then lemma 9 implies $I^* = 1$. This, along with (23) and the fact that $R_0^* = 0$ implies $\theta^* = \hat{\underline{\theta}}_\Psi \geq \theta_i \geq \theta_M(\Psi, c_I, \theta_i)$ (the latter inequality follows from lemma 9), another contradiction. It follows that in equilibrium, challenger threshold should satisfy $\theta^* = \theta_M(\Psi, c_I, \theta_i)$. Lemma 9 has shown that if

$c_I > 0$, then $\theta_M(\cdot) < \theta_i$. So given $R_0^* = 0$ and the expressions in (23), there exists a unique I^* that satisfies the following:

$$\theta_M(\Psi, c_I, \theta_i) = \theta_i - r + (1 - \Psi I^*)^{-1} c_E, \tag{41}$$

where the range $I^* \in \left(0, \Psi^{-1} - \frac{c_E}{\Psi r}\right)$ follows from the observation that at this range, $\theta_M(\Psi, c_I, \theta_i) \in (\hat{\underline{\theta}}, \theta_i)$. ∎

When the incumbent's quality is below average (and investigation technology is not too weak and/or not too costly), media shines the most. Because voters are *ex ante* content with even the worst candidate-type challenging, this is the setting where the media has the most incentive to take the initiative and actively investigate. It does so in order to make sure that it is not an inferior challenger replacing the incumbent. In other words, bad incumbents attract bad challengers, which is when the media is most needed. Proposition 6 also reveals that the relation between the probability of a successful investigation and net (of costs of challenging) office rents plays a key role in determining the extent of media activity. Comparing these two is akin to contrasting the strength of media's positive selection powers to the potency of the negative selection effects of political payoffs. In the first two equilibria, the latter dominates so that even if the media was fully active ($I^* = 1$), it would be unable to deter all inferior candidate-types from challenging ($\hat{\underline{\theta}}_\Psi < \theta_i$). Yet if the chance of an investigation failure is sufficiently high, then the real costs of information would be high as well, which would translate into a low standard for the media, lower even than $\hat{\underline{\theta}}_\Psi$. So in the first equilibria, media never investigates with certainty but mixes up until the minimum challenger quality accommodates its low standard. Second equilibria occurs when the media strength is at its sweet spot regarding the intensity of its journalistic activity: when it is low relative to the severity of adverse selection, but not that low in absolute terms so that investigation is still cost efficient, its standard will be higher than the reach of its selection powers. Hence, it will have a strong incentive to engage actively in political journalism, implying $I^* = 1$. In the final case, the strength of investigation technology is sufficient to deter all inferior challengers ($\hat{\underline{\theta}}_\Psi \geq \theta_i$), so as long as $c_I > 0$, media will return to mixing.

Figure 2 depicts various ranges of Ψ associated with different regimes of equilibria when $\theta_i < \mathbb{E}[\theta_c]$. It has investigation strength on the x-axis and the associated values for θ_M and $\hat{\underline{\theta}}_\Psi$ on the y-axis for fixed values of $r, c_E, c_I, \theta_i, \mathbb{E}[\theta_c]$. The (solid) orange curve represents $\hat{\underline{\theta}}_\Psi$, which is equal to $\hat{\underline{\theta}} = \theta_i - r + c_E$ whenever $\Psi = 0$, is strictly increasing in Ψ, and has a vertical asymptote at $\Psi = 1$.
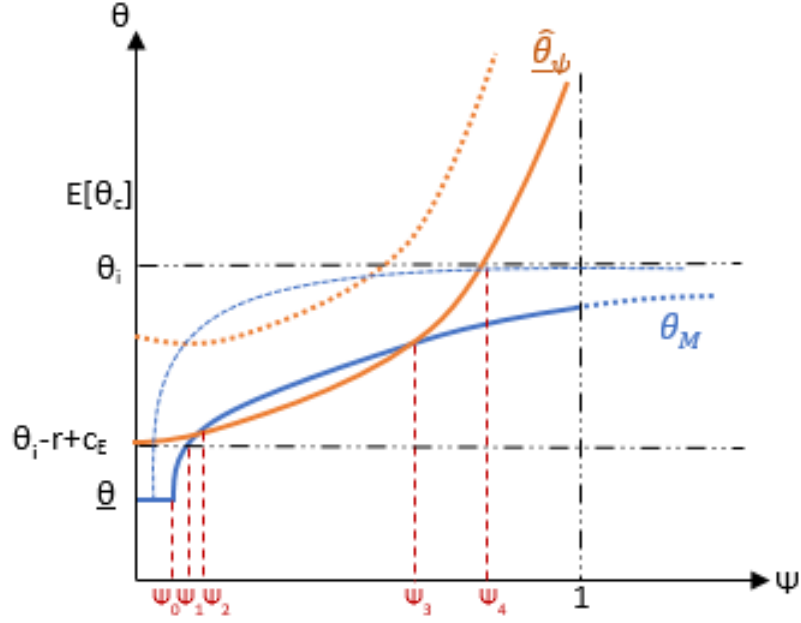
Figure 2: Media Strength and Equilibrium Transition - Below Average Incumbent

The (solid) blue curve represents $\theta_M(\cdot)$, which is equal to $\underline{\theta}$ when $\Psi \leq \Psi_0$, is increasing in Ψ and takes a specific value whenever $\Psi = 1$.[41] It's concave shape (for $\Psi > \Psi_0$) is strongly suggested by two facts: First, θ_M is strictly increasing in Ψ whenever $\Psi > \Psi_0$ (so its derivative with respect to Ψ is nowhere zero, see next section), and second, one can see from (35) that whenever $\Psi \to \infty$, $\theta_M \to \theta_i$, i.e. θ_M has a horizontal asymptote at $\theta = \theta_i$. Yet, it is possible that θ_M exhibits several "slanted" inflection points before turning fully concave and asymptoting towards θ_i. The way the figure is drawn, as Ψ is varied in $[0, 1]$, five regimes of equilibria will arise:

1. $\Psi \in [0, \Psi_0)$ - Proposition 4 case 1 (prohibitive costs): $I^* = 0$ and $\theta^* = \hat{\underline{\theta}}$.

2. $\Psi \in [\Psi_0, \Psi_1]$ - Proposition 5 case 1 (low standard): $I^* = 0$ and $\theta^* = \hat{\underline{\theta}}$.

3. $\Psi \in (\Psi_1, \Psi_2)$ - Proposition 6 case 1: $I^* \in (0, 1)$ and $\theta^* = \theta_M(\cdot)$.

4. $\Psi \in [\Psi_2, \Psi_3]$ - Proposition 6 case 2: $I^* = 1$ and $\theta^* = \hat{\underline{\theta}}_\Psi$.

5. $\Psi \in (\Psi_3, \Psi_4)$ - Proposition 6 case 1 again: $I^* \in (0, 1)$ and $\theta^* = \theta_M(\cdot)$.

6. $\Psi \in [\Psi_4, 1]$ - Proposition 6 case 3: $I^* \in \left(0, \Psi^{-1} - \frac{c_E}{\Psi r}\right)$ and $\theta^* = \theta_M(\cdot)$.

---

[41] Ψ_0 is the threshold for prohibitively low investigation strength and it is defined as $\Psi_0 = \frac{c_I}{K(\theta_i)}$. The fact that θ_M is increasing in Ψ is not hard to see but will nevertheless be shown in the following section.

There are three things noteworthy to mention about the regime switching behaviour when Ψ is varied. First, in the particular example given above, case 1 in proposition 6 occurs twice. This is not a peculiarity caused by the way the figure is drawn, nor does it depend on the assumed pictographic concavity of $\theta_M(\cdot)$ but rather a consequence of policy continuity, which is to be shown subsequently. If the equilibria in which media always investigates were to occur, then case 1 in proposition 6 should occur at least twice as Ψ is varied, because otherwise (if the passage was direct from case 2 to case 3), I^* would jump from 1 to $I^* \in \left(0, \Psi^{-1} - \frac{c_E}{\Psi r}\right)$ where $\Psi^{-1} - \frac{c_E}{\Psi r} \leq 1$. Second, in general, the equilibria in proposition 6 case 2 ($I^* = 1$) does not need to occur, since if the signalling strength of challenger entry is relatively high, then media would never be fully active, i.e. it would always mix. This can be seen from the dashed orange curve, which represents an upwards shift in $\hat{\underline{\theta}}_\Psi$ due to an increase in participation costs c_E or a decrease in office rents r. On the flip side, if the cost of investigation was lower, then the curve representing $\theta_M(\cdot)$ would shift up (dashed blue curve), implying that the media would be fully active for a wide range of parameter values reflecting its strength.[42] Notice that investigation cost is the sole (besides strength) determinant of media pickiness in this model. However, one can also assume that media gets additional benefits (besides altruism) from revealing the true quality of challengers by, say, giving it a reward b every time it successfully does so. An increase in b would have very similar effects to a decrease in c_I. Finally, again with respect to the example given above, the passage from case 2 to case 1 and 3 respectively implies that media becomes less active as its journalistic strength increases. Yet, this decrease in journalistic activity is never high enough to undermine the gains associated with the increase in Ψ. From figure 2, one can immediately see that θ^* is the lower envelope of the relevant curves, i.e.

$$\theta^* \left(\Psi \mid \theta_i < \mathbb{E}\left[\theta_c\right]\right) = \min\left\{\hat{\underline{\theta}}_\Psi, \max\left\{\hat{\underline{\theta}}, \theta_M(\Psi)\right\}\right\}, \tag{42}$$

and thus, it is increasing in Ψ. The intuition behind this is clear. The reason for reduced media activity when it is stronger is the increased threat it poses to inferior challengers. As such, the media no longer needs to be as active as before, as it can achieve a better selection effect by less journalism.

A below average incumbent is not the only reason for the media being actively involved in selection. If it has sufficiently low costs, it can also be active due to higher standards compared to voters.

---

[42]As $c_I \to 0$, $\theta_M(\cdot)$ converges *pointwise* to the horizontal line at θ_i. On the other hand, as $c_I \to \infty$, $\theta_M(\cdot)$ converges *pointwise* to the horizontal line at $\underline{\theta}$.

Proposition 7 (PBE - Active media: Higher media standard): Assume that $\hat{\underline{\theta}} < \theta_i$, $\frac{c_I}{\Psi} \leq K(\theta_i)$, and $\Omega \neq \emptyset$. Assume further that $\theta_i \geq \mathbb{E}[\theta_c]$ and that $\theta_M(\Psi, c_I, \theta_i) > \max\{\hat{\underline{\theta}}, \theta_V(\theta_i)\}$. Define $\theta^* = \inf \Omega$. Then the following equilibria prevail and they are unique in the lemma 2 sense.

1. If $\Psi < 1 - \frac{c_E}{r}$ and $\theta_M(\Psi, c_I, \theta_i) < \hat{\underline{\theta}}_\Psi$, then $R_0^* = 0$ and there exists a unique $I^* \in (0, 1)$ such that $\theta^* = \theta_M(\Psi, c_I, \theta_i)$.

2. If $\Psi < 1 - \frac{c_E}{r}$ and $\theta_M(\Psi, c_I, \theta_i) > \theta_V(\theta_i) \geq \hat{\underline{\theta}}_\Psi$, then $I^* = 1$, and there exists a unique $R_0^* \in [0, 1 - \frac{c_E}{(1-\Psi)r})$ such that $\theta^* = \theta_V(\theta_i)$.

3. If $\Psi < 1 - \frac{c_E}{r}$ and $\theta_M(\Psi, c_I, \theta_i) \geq \hat{\underline{\theta}}_\Psi > \theta_V(\theta_i)$, then $R_0^* = 0$, $I^* = 1$ and $\theta^* = \hat{\underline{\theta}}_\Psi$.

4. If $\Psi \geq 1 - \frac{c_E}{r}$, then $R_0^* = 0$ and there exists a unique $I^* \in (0, \Psi^{-1} - \frac{c_E}{\Psi r})$ such that $\theta^* = \theta_M(\Psi, c_I, \theta_i)$.

Proof: Suppose $\hat{\underline{\theta}} < \theta_i$, $\frac{c_I}{\Psi} \leq K(\theta_i)$, $\mathbb{E}[\theta_c] \leq \theta_i$, $\Omega \neq \emptyset$ and $\theta_M(\Psi, c_I, \theta_i) > \max\{\hat{\underline{\theta}}, \theta_V(\theta_i)\}$.

1. Almost identical to the proof of proposition 6 case 1, with the added fact that $\theta_V(\theta_i) < \theta_M(\Psi, c_I, \theta_i) = \theta^*$ implying $R_0^* = 0$.

2. Assume $\Psi < 1 - \frac{c_E}{r}$ and $\theta_M(\Psi, c_I, \theta_i) > \theta_V(\theta_i) \geq \hat{\underline{\theta}}_\Psi$. Suppose instead $\theta^* > \theta_V(\theta_i)$. Then lemma 8 implies $R_0^* = 0$. But then, for any $I^* \in [0, 1]$, one must have:

$$\theta^* = \theta_i - r + (1 - \Psi I^*)^{-1} c_E \leq \hat{\underline{\theta}}_\Psi \leq \theta_V(\theta_i),$$

a contradiction. Combining this with corollary 2 ensures that in any equilibrium, one must have $\theta^* = \theta_V(\theta_i)$, so voters are indifferent when uninformed. Since by assumption $\theta_M(\cdot) > \theta_V(\cdot)$, $I^* = 1$ follows directly. Moreover, from (23)-(24), the unique R_0^* should solve the following.

$$\theta_V(\theta_i) = \theta_i - r + [(1 - \Psi)(1 - R_0^*)]^{-1} c_E, \tag{43}$$

where the range $R_0^* \in [0, 1 - \frac{c_E}{(1-\Psi)r})$ follows from the fact that at this range, $\theta_V(\cdot) \in [\hat{\underline{\theta}}_\Psi, \theta_i)$. Note that $\theta_V(\cdot) = \theta_i$ can't occur in this equilibrium as long as $c_I > 0$, since from lemma 9 we have $\theta_M(\cdot) < \theta_i$.

3. Assume $\Psi < 1 - \frac{c_E}{r}$ and $\theta_M(\Psi, c_I, \theta_i) \geq \hat{\underline{\theta}}_\Psi > \theta_V(\theta_i)$. First, suppose instead $\theta^* < \hat{\underline{\theta}}_\Psi$. Then from lemma 9, $I^* = 1$. Then from (23)-(24), for any

97

$R_0^* \in [0, 1]$, one has:

$$\theta^* = \theta_i - r + [(1 - \Psi)(1 - R_0^*)]^{-1} c_E \geq \hat{\underline{\theta}}_\Psi, \qquad (44)$$

a contradiction. Next, suppose $\theta^* > \hat{\underline{\theta}}_\Psi$. Then from lemma 8, $R_0^* = 0$, so from (23)-(24), for any $I^* \in [0, 1]$, one has:

$$\theta^* = \theta_i - r + (1 - \Psi I^*)^{-1} c_E \leq \hat{\underline{\theta}}_\Psi, \qquad (45)$$

a contradiction. It follows that in equilibrium, $\theta^* = \hat{\underline{\theta}}_\Psi$. If $\theta_M(\cdot) > \hat{\underline{\theta}}_\Psi$, then $I^* = 1$ follows from lemma 9. If $\theta_M(\cdot) = \hat{\underline{\theta}}_\Psi$, then $I^* = 1$ follows from the definition of $\hat{\underline{\theta}}_\Psi$. In both cases, $R_0^* = 0$ follows from lemma 8.

4. Almost identical to the proof of proposition 6 case 3, with the added fact that $\theta_V(\theta_i) < \theta_M(\Psi, c_I, \theta_i) = \theta^*$ implying $R_0^* = 0$. ∎

Proposition 7 shows that even if the incumbent is an above average quality politician, which gives voters the incentive to improve the pool of potential challengers by fine-tuning their voting strategies under uncertainty, media can still play a role in selection if costs of investigation are sufficiently low. First, third and fourth cases are very similar to first, second and third cases of proposition 6 respectively, so they don't require additional discussion. Second case of proposition 7 is interesting however, because it involves the media aggressively investigating and voters threatening to sometimes reelect the incumbent simultaneously. Unlike other equilibria, which underline the substitutability between media and voter strategies in improving selection (as either one or the other is active, never both), these equilibria emphasize the complementarity between them. The maximum quality for the worst candidate achievable by media selection alone is bounded above by $\hat{\underline{\theta}}_\Psi$. Since this is below the minimum challenger quality acceptable to voters, one needs to introduce an appropriate probability of retaining the incumbent into their strategy under uncertainty to make sure that a satisfactory (and stable) expected challenger quality is attained. Notice that voters could get this minimum quality even though media was passive, i.e. even if $I^* = 0$. Since the media acts as an altruistic "extended arm" of voters in this model, why then do they still investigate aggressively and undertake costs of journalism? One can read this as a coordination failure result. An alternative, and a more appropriate reading would note that voters and media care about the same thing, but from different perspectives. While the voters care about the expected quality of the politician to be placed in the office, media cares about the expected *ex post* welfare

loss due to placing the wrong guy in the office. As long as costs of journalism are low and investigation strength is not too low, this latter concern would dominate the former in its standard for worst challenger quality. This observation follows from a simple comparison of media and voters' objectives in (31) and (35), which suggests that whenever c_I is low enough and Ψ is not too low, $\theta_M(\cdot) > \theta_V(\cdot)$.
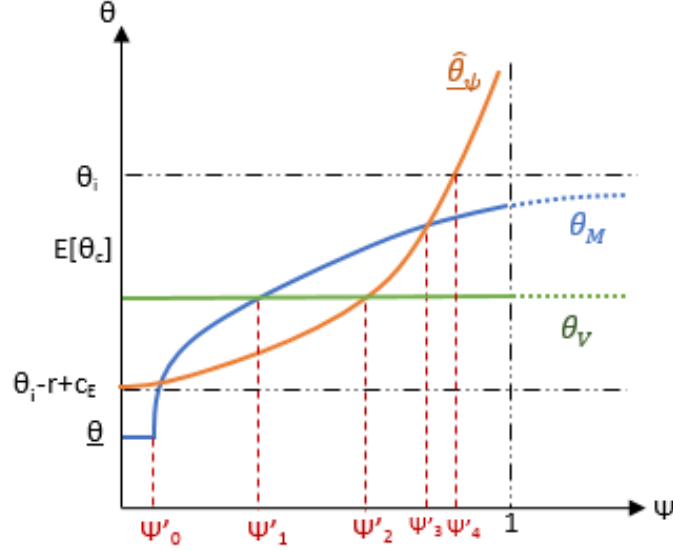

Figure 3: Media Strength and Equilibrium Transition - Above Average Incumbent

Figure 3 gives the figure 2 analogue of proposition 7, with $\theta_V(\theta_i)$ taking an appropriate value for case 2 to occur at a certain range of Ψ. The main difference between figure 3 and figure 2 is the existence of a $\theta_V(\theta_i) > \underline{\theta}$, which is indicated by $\theta_i > \mathbb{E}[\theta_c]$ as shown in lemma 8. Below, I list various regimes of equilibria associated with it.

1. $\Psi \in [0, \Psi_0')$ - Proposition 4 case 3 (prohibitive costs):
   $I^* = 0$, $R_0^* \in (0, 1 - \frac{c_E}{r})$, $\theta^* = \theta_V(\theta_i)$.[43]

2. $\Psi \in [\Psi_0', \Psi_1')$ - Proposition 5 case 3 (low media standard):
   $I^* = 0$, $R_0^* \in (0, 1 - \frac{c_E}{r})$, $\theta^* = \theta_V(\theta_i)$.

3. $\Psi \in (\Psi_1', \Psi_2']$ - Proposition 7 case 2 (higher media standard):
   $I^* = 1$, $R_0^* \in [0, 1 - \frac{c_E}{(1-\Psi)r})$, $\theta^* = \theta_V(\theta_i)$.

4. $\Psi \in (\Psi_2', \Psi_3']$ - Proposition 7 case 3 (higher media standard):
   $I^* = 1$, $R_0^* = 0$, $\theta^* = \hat{\underline{\theta}}_\Psi$.

---

[43]Note that R_0^*'s range does not include $1 - \frac{c_E}{r}$ unlike the statement of the proposition. This is because in the figure, $\theta_V(\theta_i) < \theta_i$.

5. $\Psi \in [\Psi'_3, \Psi'_4)$ - Proposition 7 case 1 (higher media standard):
   $I^* \in (0, 1)$, $R_0^* = 0$, $\theta^* = \theta_M(\cdot)$.

6. $\Psi \in [\Psi_4, 1]$ - Proposition 7 case 4 (higher media standard):
   $I^* \in \left(0, \Psi^{-1} - \frac{c_E}{\Psi r}\right)$, $R_0^* = 0$, $\theta^* = \theta_M(\cdot)$.

First, notice that the same equilibria prevail under the first two regimes in the list above. This is because $\theta^* = \theta_V$ in both of them and θ_V is independent of Ψ, which implies that the cost-benefit structure of the media is irrelevant in equilibrium. Second, non-monotonic response of media activity to improvements in journalistic strength can be clearly seen: up to Ψ'_1, media does not investigate at all; between Ψ'_1 and Ψ'_3, it always investigates; and between Ψ'_3 and 1, it sometimes investigates. Keep in mind that the equilibria does not remain the same between Ψ'_3 and 1 as at this range, $\theta^* = \theta_M(\cdot)$ where the latter depends on Ψ. Voters reelect the incumbent with positive probability under uncertainty even after the media becomes fully active, but cease to do so once the media strength attains a sufficiently high level and allows the media to provide a strong enough selection effect by itself. This level corresponds to Ψ'_2, where $\theta_V = \hat{\underline{\theta}}_\Psi$. Third, notice that despite the non-monotonic response of media activity, minimum challenger quality is (weakly) increasing in media strength, because when this is low and thus selection powers of the media is low, voters provide an adequate amount of selection through their blind-voting strategies.

$$\theta^* \left(\Psi \mid \theta_i > \mathbb{E}[\theta_c]\right) = \max \left\{ \theta_V(\theta_i), \min \left\{ \hat{\underline{\theta}}_\Psi, \theta_M(\Psi) \right\} \right\}. \tag{46}$$

Finally, notice that the equilibria that arise exactly at Ψ'_1 is not mentioned in the list. Furthermore, media policy have a discontinuity at that point, jumping from $I^* = 0$ to $I^* = 1$. This might seem contradictory with the previous remark about policy continuity under $\theta_i < \mathbb{E}[\theta_c]$. As will be shown in the next subsection, policy continuity depends on the lemma 2 sense of uniqueness, which no longer holds at Ψ'_1.

Proposition 7.1 (PBE - Indeterminate media activity: Knife-edge case):
Assume that $\hat{\underline{\theta}} < \theta_i$, $\frac{c_I}{\Psi} \leq K(\theta_i)$, and $\Omega \neq \emptyset$. Assume further that $\theta_i \geq \mathbb{E}[\theta_c]$ and that $\theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i) > \hat{\underline{\theta}}$. Define $\theta^* = \inf \Omega$. Then prevailing equilibria are no longer unique in the lemma 2 sense and consist of $\theta^* = \theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i)$ with any pair I^*, R_0^* such that the following holds.

$$\theta^* = \theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i) = \theta_i - r + \left[(1 - \Psi I^*)(1 - R_0^*)\right]^{-1} c_E. \tag{47}$$

Proof: Suppose that $\theta^* > \theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i)$, then lemmas 8 and 9 imply $R_0^* = I^* = 0$, which implies (from (23)-(24)) $\theta^* = \hat{\underline{\theta}}$, a contradiction. Suppose instead $\theta^* < \theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i)$, then lemmas 8 and 9 imply $R_0^* = I^* = 1$, which contradicts corollary 1. It follows that $\theta^* = \theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i)$. So the media and voters should both be indifferent in equilibrium, and from (23)-(24), the only θ^* that would satisfy such simultaneous indifference and candidate's sequential rationality is given by (47). Note that there is a continuum of I^*, R_0^* satisfying (47), hence no longer uniqueness. The exact allowable ranges for I^* and R_0^* will depend on the specific parameter configuration at which the indifference $\theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i)$ occurs. ∎

This concludes the presentation of equilibria in this model. Next, I will study the effects induced on equilibrium strategies by changes in payoff parameters with a particular emphasis on the media's journalistic strength and its investigation decision.

2.4.3 Equilibrium Effects of Changes in Payoffs

Payoff parameters in the model are: θ_i, c_E, r, c_I and Ψ. I start by studying the equilibrium selection effects of these parameters, i.e. how does the equilibrium challenge threshold respond to a change in one of these parameters. Following lemma partly answers this question.

Lemma 10 (Comparative statics - Media and voter cutoffs): Assume $\theta_V(\theta_i) \in (\underline{\theta}, \theta_i)$, then $\theta_V(\theta_i)$ is strictly increasing in θ_i. If $\theta_M(\Psi, c_I, \theta_i) \in (\underline{\theta}, \theta_i)$, then $\theta_M(\cdot)$ is strictly increasing in Ψ, θ_i and strictly decreasing in c_I.

Proof: Appendix. ∎

This is a very intuitive result and in line with previous discussions. A higher incumbent quality incentivizes both the media and voters to demand a higher minimum challenger quality. Furthermore, a fall in information costs or an increase in journalistic strength allow the media to be more aggressive in setting its standard. Next, I define two piecewise functions which will allow expressing the equilibrium challenge threshold in a simple form. For the voters:

$$\tilde{\theta}_V(\theta_i) = \begin{cases} \underline{\theta}, & (\theta_i < \mathbb{E}[\theta_c]) \\ \theta_V(\theta_i), & (\theta_i \geq \mathbb{E}[\theta_c]) \end{cases}, \tag{48}$$

with $\theta_V(\theta_i)$ is as defined in lemma 8. Notice that $\tilde{\theta}_V(\theta_i)$ is continuous, as $\theta_V(\theta_i) = \underline{\theta}$ whenever $\theta_i = \mathbb{E}[\theta_c]$, and $\theta_V(\theta_i)$ itself is continuous. For the media:

$$\tilde{\theta}_M(\Psi, c_I, \theta_i) = \begin{cases} \underline{\theta}, & \left(\frac{c_I}{\Psi} > K(\theta_i)\right) \\ \theta_M(\Psi, c_I, \theta_i), & \left(\frac{c_I}{\Psi} \leq K(\theta_i)\right) \end{cases}, \tag{49}$$

with $\theta_M(\cdot)$ is as defined in lemma 9. Notice that $\tilde{\theta}_M(\cdot)$ is continuous, as $\theta_M(\cdot) = \underline{\theta}$ whenever $\frac{c_I}{\Psi} = K(\theta_i)$, and $\theta_M(\cdot)$ itself is continuous. Combining all the different equilibria presented previously, following function can be seen to map payoff parameters to the unique equilibrium challenge threshold.[44]

$$\theta^*(\Psi, c_I, \theta_i, c_E, r) = \max\left\{\hat{\underline{\theta}}, \tilde{\theta}_V(\theta_i), \min\left\{\hat{\underline{\theta}}_\Psi, \tilde{\theta}_M(\Psi, c_I, \theta_i)\right\}\right\}. \tag{50}$$

The mapping defined in (50) is continuous, as all of its components are continuous. Moreover, $\hat{\underline{\theta}}$ and $\hat{\underline{\theta}}_\Psi$ are strictly increasing in θ_i, c_E and strictly decreasing in r.[45] Furthermore, $\hat{\underline{\theta}}_\Psi$ is strictly increasing in Ψ. Combining these with lemma 10, next corollary follows immediately.

Corollary 3 (Equilibrium response of minimum challenger quality): $\theta^*(\Psi, c_I, \theta_i, c_E, r)$ is non-decreasing in θ_i, Ψ, c_E and non-increasing in c_I, r.[46]

The reason for using the terms non-decreasing and non-increasing is that depending on the equilibrium, slight changes in some parameters might leave the challenge threshold intact. For instance, if $\hat{\underline{\theta}} > \max\left\{\tilde{\theta}_V(\theta_i), \tilde{\theta}_M(\Psi, c_I, \theta_i)\right\}$, then $I^* = R_0^* = 0$ and $\theta^* = \hat{\underline{\theta}}$ (proposition 5 case 2), and slight changes in Ψ or c_I have no impact on the value of θ^*. From (50), it can be seen that an increase in θ_i always increases θ^* unless $\theta^* = \underline{\theta}$: if $\theta^* \in \{\hat{\underline{\theta}}, \hat{\underline{\theta}}_\Psi\}$, then it increases it due to the fact that candidates themselves care about the quality of the politician to assume the office, and if $\theta^* \in \{\theta_V(\cdot), \theta_M(\cdot)\}$, then it increases it due to either the media or voters demanding a higher minimum quality from them. On the other hand, a slight change in Ψ or c_I affects the equilibrium only if the media was already actively involved in selection.[47] As expected from such a case, an increase in information costs worsens selection while an increase in media strength improves it. Finally, a change in political payoffs r and c_E affects the equilibrium minimum challenger quality only if $\theta^* \in \{\hat{\underline{\theta}}, \hat{\underline{\theta}}_\Psi\}$. While the direction of their impact is intu-

---

[44]Notice that even if an equilibrium is not unique in the lemma 2 sense (i.e. proposition 7.1), θ^* is still unique.

[45]As a reminder, $\hat{\underline{\theta}} = \theta_i - r + c_E$ and $\hat{\underline{\theta}}_\Psi = \theta_i - r + (1 - \Psi)^{-1}c_E$.

[46]Notice that this result does not depend on the concavity/convexity characteristics of $\theta_M(\cdot)$. It solely hinges on the monotonicity of thresholds. If $\theta_M(\cdot)$ were to intersect $\hat{\underline{\theta}}_\Psi$ multiple times, then the figures (4) and (5) would be different but the monotonicity would remain intact.

[47]Or about to be actively involved, i.e. at a regime switching point. This can occur when there is an equality relationship between $\theta_M(\cdot)$ and one of the remaining thresholds in (50).

itive, i.e. higher office rents attract worse challengers and higher challenge costs discourage them, a more interesting aspect about this result is that they don't influence θ^* whenever $\theta^* \in \{\theta_V(\theta_i), \theta_M(\Psi, c_I, \theta_i)\}$. Whenever voters or the media are able to achieve a minimum challenger quality up to their standards (and these standards are higher than the institutional lower bounds), their strategies neutralize negative selection effects of political payoffs in the margin. This can be read as an envelope result. For instance, whenever the media has a higher standard compared to voters but its ability to induce selection is limited by its strength and yet it is still sufficient to provide an adequate level of selection in the eyes of voters (i.e. proposition 7 case 3), then the constraint associated with $\hat{\underline{\theta}}_\Psi$ is active so a change in political payoffs affects the minimum challenger quality in equilibrium. Figures 4 and 5 plot challenge thresholds corresponding to figures 2 and 3 respectively.
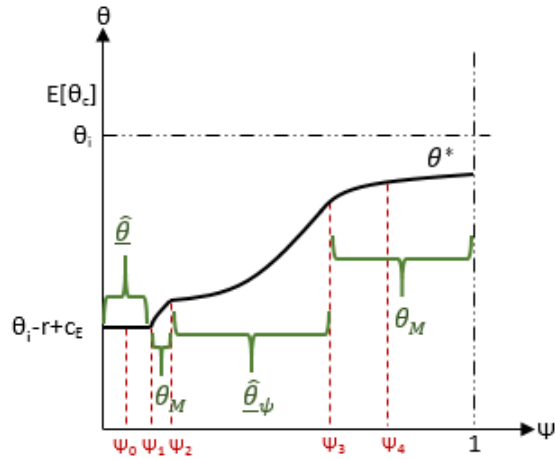

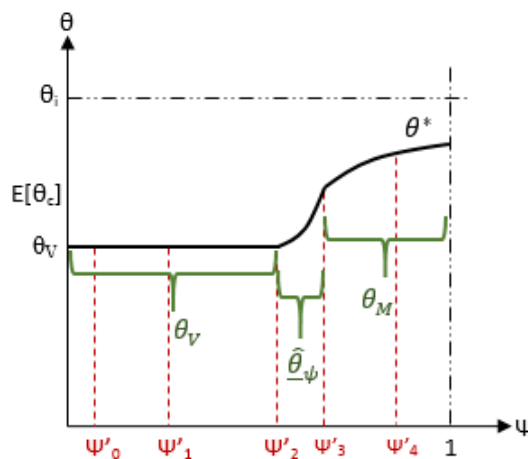Figure 4: Media Strength and Challenger Threshold - Below Average Incumbent

Figure 5: Media Strength and Challenger Threshold - Above Average Incumbent

Equilibrium responses of media and voter strategies (under uncertainty) to changes in payoffs are slightly more involved. I first start with a straightforward comparative statics result. Notice that both media's and uncertain voters' equilibrium strategies are differentiable in payoff parameters when the equilibrium is not at a regime switching point. This is because at an "interior" equilibria, they are unique and are either pure strategies equal to 0 or 1 in a sufficiently small open neighbourhood around the given payoff vector, or a completely mixed strategy satisfying a uniquely solvable equation (in such a neighbourhood) such as (41) made up of differentiable components $\theta_M(\cdot)$ or $\theta_V(\cdot)$.[48]

Lemma 11 (Comparative statics - Media and voter strategies): Assume that the given payoff vector $(\theta_i, c_E, r, c_I, \Psi)$ corresponds to an equilibrium at a non-regime switching point and satisfies $\hat{\underline{\theta}} < \theta_i$, $\theta_i \in (\underline{\theta}, \overline{\theta})$, $c_E, c_I > 0$ and $\Psi \in (0, 1)$. Then both I^* and R_0^* are differentiable in payoff parameters. If they are equal to 0 or 1, then their derivatives with respect to payoff parameters are zero. If the equilibrium is characterized by $\theta^* = \theta_V(\theta_i)$, then:

$$\frac{\partial R_0^*}{\partial r} > 0 \; ; \; \frac{\partial R_0^*}{\partial c_E} < 0 \; ; \; \frac{\partial R_0^*}{\partial \theta_i} \gtrless 0 \; ; \; \frac{\partial R_0^*}{\partial \Psi} \leq 0 \; ; \; \frac{\partial R_0^*}{\partial c_I} = 0, \tag{51}$$

where $\frac{\partial R_0^*}{\partial \theta_i} \geq (>)0$ if f is (strictly) log-concave, and $\frac{\partial R_0^*}{\partial \Psi} < 0$ only if the equilibrium occurs at proposition 7 case 2.[49] If the equilibrium is characterized by $\theta^* = \theta_M(\Psi, c_I, \theta_i)$, then:

$$\frac{\partial I^*}{\partial r} > 0 \; ; \; \frac{\partial I^*}{\partial c_E} < 0 \; ; \; \frac{\partial I^*}{\partial \theta_i} \gtrless 0 \; ; \; \frac{\partial I^*}{\partial \Psi} \gtrless 0 \; ; \; \frac{\partial I^*}{\partial c_I} < 0, \tag{52}$$

where $\frac{\partial I^*}{\partial \theta_i} \geq (>)0$ again if f is (strictly) log-concave.

Proof: Appendix. ∎

First thing to notice from lemma 11 is that payoff parameters (satisfying the

---

[48]By regime switching points, I mean parameter configurations at which slight perturbations to payoffs would result the new equilibria to occur under a new parameter regime. For instance a payoff vector leading to $\theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i) > \hat{\underline{\theta}}_\Psi$ corresponds to such a regime switching point, as a slight increase in Ψ would change the equilibria from proposition 7.1 to one under proposition 7 case 2. Another example is a configuration such that $\theta_i < \mathbb{E}[\theta_c]$ and $\hat{\underline{\theta}} = \theta_M(\Psi, c_I, \theta_i)$. Yet another example is proposition 7 case 3 with $\hat{\underline{\theta}}_\Psi = \theta_M(\Psi, c_I, \theta_i)$. The reason that differentiability of strategies can be violated at these points (besides possible continuity issues) is that if they correspond to indifference points where a slight perturbation would lead a receiver to switch from playing a pure strategy to a complete mixture, then they can introduce non-differentiable kinks into strategies.

[49]Notice that the non-regime switching point assumption requires $\theta_V(\theta_i) \neq \hat{\underline{\theta}}_\Psi$ if the equilibrium were to occur at proposition 7 case 2.

non-regime switching condition) affect media and voter strategies at the margin only if they are actively participating in selection and playing completely mixed strategies. This results from their action sets being finite and the related nature of equilibria in this model. If they are playing pure strategies, then there must be a particular strict ordering between various decision-relevant quality thresholds, which would keep holding if the payoff vector was slightly perturbed. Looking at the impact of marginal changes in payoff parameters, signs associated with r and c_E are intuitive. An increase in r or a decrease in c_E represents a weakening in the signalling power of a challenge because holding media and voter strategies constant, the quality of the worst candidate-type who challenges falls. Media and voters respond to this by tightening the leash. In case of media, this takes the form of investigating more frequently. In case of voters, it implies retaining the incumbent more often. Regarding θ_i, it is possible to distinguish two effects. Since candidates themselves care about the public good and hence the quality of the elected official, *ceteris paribus*, an increase in the incumbent quality attracts better candidates, which increases expected challenger quality (voters' decision statistic), as well as the expected inferior challenger quality (media's decision statistic). If these latter increases are not enough to compensate the higher benefit from reelecting the incumbent -which is the case when f is log-concave-then media has a stronger incentive to be active and voters have more reason to increase their threat of reelecting the incumbent. Regarding c_I, an increase in information costs never influence an uncertain voter's decision as long as the non-regime switching condition is satisfied. But as expected, it makes the media less likely to be active given that it is not constrained by its strength. Finally, the interesting fact about media strength Ψ is that it can influence voting decision under uncertainty. If the equilibrium is as described in proposition 7 case 2, i.e. the media has a higher quality standard compared to voters ($\theta_M(\cdot) > \theta_V(\cdot)$) but not only it is constrained by its strength when it comes to satisfying its own standard ($\theta_M(\cdot) > \hat{\underline{\theta}}_\Psi$), it is also constrained in the sense that the minimum challenger quality it can guarantee by being fully active is not sufficient to satisfy voters ($\theta_V(\cdot) > \hat{\underline{\theta}}_\Psi$). In that case, an increase in media strength translates into a higher minimum challenger quality as the media is fully active. This allows voters to reduce their threat of retaining the incumbent. On the other hand, an increase in Ψ has an ambiguous effect on media activation whenever it is not constrained by its strength and is actively involved in selection. This is in line with the non-monotonic media activation pattern discussed in the previous subsection, associated with figures 2 and 3.

$$\frac{\partial I^*}{\partial \Psi} = \Psi^{-1} \frac{(1 - \Psi I^*)^2}{c_E} \frac{\partial \theta_M(\Psi, c_I, \theta_i)}{\partial \Psi} - \Psi^{-1} I^* \gtrless 0. \tag{53}$$

In (53), the first term (which is strictly positive from lemma 10) reflects the efficiency gain due to increased journalistic strength. Since an investigation is more likely to succeed following an increase in Ψ, undertaking costs associated with it is better justified, which is reflected in the higher media standard. The second term captures the boost to media's selection strength. For a given level of media activity I^*, a higher Ψ implies a higher minimum challenger quality. In other words, same (or even slightly higher) positive selection effect can be induced by a slightly lower level of activity, which would translate into expected cost savings. If the latter effect dominates the former, then media activity should decrease. However, this decrease would never be strong enough to fully erode the efficiency gain. This can be either deduced indirectly from corollary 3, or can be seen from the following.

$$\frac{\partial (\Psi I^*)}{\Psi} = I^* + \Psi \frac{\partial I^*}{\partial \Psi} > 0$$
$$\Longleftrightarrow \quad \frac{\partial I^*}{\partial \Psi} > -\Psi^{-1} I^*. \tag{54}$$

The inequality in (54) follows from the first term in (53) being strictly positive. Intuitively, one should expect the expression in (53) to be positive for relatively low values of Ψ and negative for high values of Ψ. In fact, if the continuity (or *almost everywhere* continuity) of I^* can be established, then one can show that I^* indeed possesses such non-monotonicity in its comparative static.

More generally, if I^* and R_0^* are continuous in payoff parameters, then one can generalize the comparative statics results in lemma 11 to the entire set of payoff parameters satisfying $\hat{\underline{\theta}} < \theta_i$. The following result is the first step in establishing such generalization.

Lemma 12 (Payoff upper hemicontinuity of media and voters' equilibrium mixed strategy correspondences): Consider the following set.

$$\Upsilon = \left\{ (\theta_i, c_E, r, c_I, \Psi) : \theta_i \in [\underline{\theta}, \overline{\theta}], c_E > 0, r > 0, r > c_E, c_I > 0, \Psi \in [0, 1] \right\} \tag{55}$$

Equilibrium mixed strategy correspondences of voters and the media are upper hemicontinuous in payoff parameters on Υ.

Proof: Appendix. ∎

Propositions 4 to 7.1 have shown that equilibrium mixed strategy correspondences

are single-valued except at the locus of parameters satisfying the conditions for proposition 7.1. This implies that when particular selections I^* and R_0^* are made from these correspondences, they will be functions continuous everywhere on Υ except on the locus where such non-uniqueness (in the lemma 2 sense) is present.[50]

Corollary 4 (Almost everywhere continuity of equilibrium media and voter strategies): Let Υ_d be a subset of Υ defined as follows.

$$\Upsilon_d = \left\{ (\theta_i, c_E, r, c_I, \Psi) \in \Upsilon : \hat{\underline{\theta}} < \theta_i, \frac{c_I}{\Psi} \leq K(\theta_i), \theta_i \geq \mathbb{E}\left[\theta_c\right], \right.$$
$$\left. \theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i) \geq \hat{\underline{\theta}} \right\} \qquad (56)$$

Then a particular pair $(I^*, R_0^*) \equiv (I(\theta_i, c_E, r, c_I, \Psi), R_0(\theta_i, c_E, r, c_I, \Psi))$ of media and voter strategies is continuous on $\Upsilon \setminus \Upsilon_d$ and is discontinuous on Υ_d.[51]

One can select infinitely many pairs of functions (I^*, R_0^*) over Υ describing equilibrium strategies as long as they match the descriptions laid out in proposition 4 to 7.1. However, all those pairs will be identical except on Υ_d, due to lemma 2 sense of uniqueness. Despite the discontinuity, it turns out that it is still possible to partially generalize lemma 11 over Υ due to "monotonicity" of the jumps at points of discontinuity with respect to parameters characterizing media's technology.

Lemma 13 (Equilibrium response of media and voter strategies): Take a pair I^*, R_0^* of equilibrium strategies. Then in $\Upsilon \setminus \Upsilon_d$; R_0^* and I^* are non-decreasing in r and non-increasing in c_E. In Υ; R_0^* is non-decreasing in c_I, non-increasing in Ψ and I^* is non-increasing in c_I. Finally, in Υ_d; I^* is non-decreasing in Ψ.

Proof: Appendix. ∎

The intuition behind lemma 13 is simple. Continuity of I^* and R_0^* on $\Upsilon \setminus \Upsilon_d$ ensures that lemma 11 carries over to any regime-switching point in it. It was also shown that Υ_d consisted of payoff vectors leading to the knife-edge case described in proposition 7.1. At any knife-edge equilibrium, media and voter standards are exactly equal. So any increase in media strength or decrease in information costs lead to media's standard exceeding that of the voters, which introduces the

---

[50]This is because single-valued upper hemicontinuous correspondences are continuous when viewed as functions.

[51]The terminology *almost everywhere* follows from the fact that Υ_d is a subset of a four-dimensional subspace, so it has a Lebesgue measure zero on five dimensions, where the (restricted) domain of strategies Υ resides.

required monotonicity even when discontinuous jumps are involved. This monotonicity allows to match the Υ_d-behaviour of strategies in media parameters with the general pattern laid out in lemma 11. Shortcomings of lemma 13 are also not surprising. Regarding θ_i, one can see that the differential impact of θ_i on $\theta_V(\cdot)$ and $\theta_M(\cdot)$ may lead to knife's edge being broken in both ways ($\theta_M(\cdot) > \theta_V(\cdot)$ or vice versa). Whichever case obtains, this makes sure that the impact of θ_i on I^* and R_0^* in Υ_d is asymmetric, so that one can not generalize the sign conformity result obtained in lemma 11 for θ_i to Υ_d. Even if one is not interested in generalizing lemma 11 to θ_i but simply wants to know its direction of impact at Υ_d, this requires comparing $\frac{\partial \theta_M(\cdot)}{\partial \theta_i}$ with $\frac{\partial \theta_V(\cdot)}{\partial \theta_i}$ whenever $\theta_M = \theta_V$. Unfortunately, such comparison does not yield a clear-cut ordering, but strongly suggests a higher increase for θ_V under log-concavity. This would be in line with the result that the media is relatively more important when the incumbent is below average and would imply an upward impact (of an increase in θ_i) to R_0^* and vice versa. Regarding political payoffs, notice that whenever $\xi \in \Upsilon_d$ satisfies $\theta_M(\xi) = \theta_V(\xi) = \hat{\underline{\theta}}(\xi)$, a small increase in r or a decrease in c_E would lead to an upward jump in one or both media and voter strategies, so monotonicity holds at that point. However, if $\xi \in \Upsilon_d$ satisfies $\theta_M(\xi) = \theta_V(\xi) > \hat{\underline{\theta}}(\xi)$, then a sufficiently small increase in r or a decrease in c_E would leave media and voter standards intact and would *ceteris paribus* worsen the minimum challenger quality. This would require additional selection efforts by media and/or voters.[52] But this required improvement in selection can be achieved in any possible way. For instance, voters can slightly reduce their threat of retaining the incumbent with media becoming much more active, or vice versa, or they can both become more aggressive in their selection strategies. This implies that such a monotonicity-in-Υ_d result cannot be generally obtained for political payoffs and would be dependent on the particular strategies I^*, R_0^* in hand.

2.4.4 Information Technology and Incentives for Media Activity

Notice that lemmas 11 and 13 left the strategic impact of media strength on media activity ambiguous besides suggesting a general non-monotonic pattern in line with the previously discussed equilibrium transition patterns. This section retakes the issue in a more detailed manner by building on the continuity and jump-monotonicity results presented in the previous section. The presentation is based on two cases already introduced in figures 2 and 3. These are not the only possible transition patterns mapping changes in media strength to equilibrium

---

[52]This follows from (47).

behaviour, but are the most interesting ones. A necessary condition for such full pictures is not too high information costs and a sufficiently strong *a priori* negative selection due to relatively high office rents and/or low participation costs. In addition, the influence of improvements in media strength will be contrasted with the monotonic impact of (information) cost reductions, and whenever incumbent is of above average quality, voters' equilibrium responses will be included in the picture.

I start with the case when the incumbent is relatively weak, i.e. when its quality is below the unconditional average ($\theta_i < \mathbb{E}[\theta_c]$). Figure 2 is reproduced below with additional media standard curves corresponding to different levels of investigation costs.
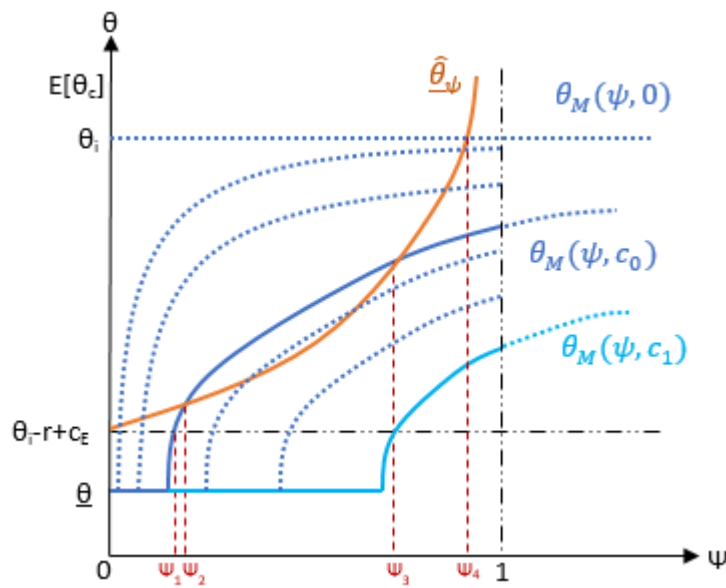
Figure 6: Media Strength, Information Costs and Equilibrium Transition - Below Average Incumbent

As investigation costs decrease down to zero, media strictly prefers investigating as long as the minimum challenger quality is below the incumbent, so its standard becomes θ_i. As costs increase from zero, two effects are observed. First, media's θ_M-curve (its non-constant portion) shifts down, reflecting the fall in its standard for a given level of journalistic strength. Second, the minimum strength above which the media considers becoming active (the prohibitively low level of strength) increases. When $\theta_i < \mathbb{E}[\theta_c]$, strategies are continuous in c_I and Ψ. This is a direct consequence of corollary 4. Figure 7 gives media and voters' strategic responses as costs are increased from zero when other parameters are fixed. The

solid purple line assumes that media strength is fixed at Ψ_3 from figure 6.[53]

Figure 7: Strategic Effects of Information Costs - Below Average Incumbent

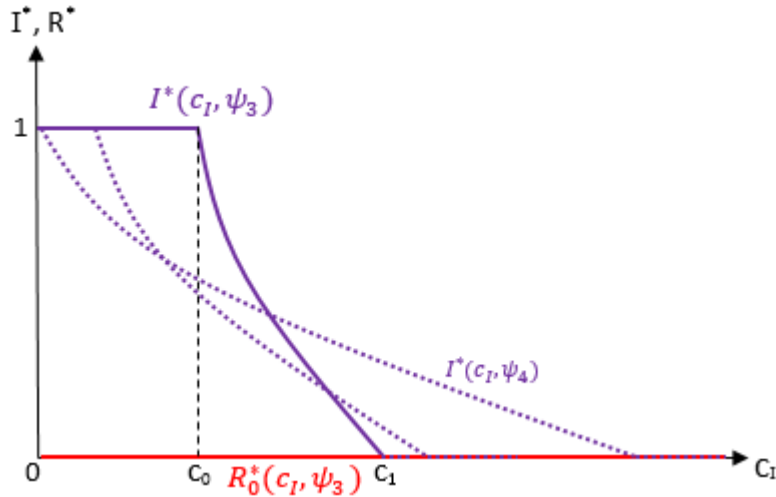As shown before, when incumbent's quality is below average, voters always vote for the challenger when uninformed and let media deal with selection. When media's strength is not high enough to fully eliminate the adverse selection effects of high net political payoffs ($\Psi < \Psi_4 = 1 - \frac{c_E}{r}$), it is fully active until costs become sufficiently high (c_0). Further increases in information costs gradually reduce media activation, reflecting the increasing cost-inefficiency of inducing selection, which is captured by the decreasing media standard. When costs become too high (c_1), media ceases to be active and becomes obliged to accept the adverse selection in its full extent. Figure 7 also shows that increases in media strength has a dual effect on the investigation curve in the (c_I, I) space. First, it shifts the full activation cutoff leftward. The reason behind this is the fact that a higher strength leads to a higher deterrence power for a given level of media activation, which implies that it becomes less efficient to be fully active at a given cost. Relatedly, whenever $\Psi = \Psi_4$, media ceases full activation unless costs are zero, because a fully active media would imply a complete eradication of adverse selection, which is never justified under positive costs. Second, it tilts the decreasing portion of I^* rightward. This reflects gains in cost efficiency, which ensures that the media can remain active for a wider range of investigation costs.

To see the effects of increasing media strength, fix costs to $c_I = c_0$. Previous discussions showed that when $\Psi \in [0, \Psi_1]$, media is completely passive and when

---

[53]The shape of investigation response in figure 7 does not depend on concavity of θ_M in Ψ for $\Psi > \Psi_0$ as depicted in figure 6. Even if $\theta_M(\cdot)$ intersects $\hat{\theta}_\Psi$ more than twice as Ψ is varied, for a given Ψ, as c_I is increased from 0, it will intersect it at most once in the (c_I, θ) plane.

$\Psi \in [\Psi_2, \Psi_3]$, it is fully active. Furthermore, from the continuity of I^*, one can deduce that it should be increasing at some right-neighbourhood of Ψ_1, increasing at some left-neighbourhood of Ψ_2, and decreasing at some right-neighbourhood of Ψ_3. This result does not depend on the above-Ψ_0 ($\Psi_0 = \frac{c_I}{K(\theta_i)}$) concavity of $\theta_M(\cdot)$, as the following lemma shows.

Lemma 14 (Irregularly non-monotonic strategic response to changes in media strength): For any given θ_i, r, c_E, c_I satisfying $\theta_i < \mathbb{E}[\theta_c]$, if $\theta_M(\cdot)$ intersects $\hat{\underline{\theta}}_\Psi$ at least once as Ψ is varied from 0 to 1, then it should intersect it an even number of times. Suppose it does and let $\Psi^1, \ldots, \Psi^{2n}$ denote the ordered set of intersection points for some positive integer n. Also let Ψ^0 denote where $\theta_M(\Psi^0) = \hat{\theta}$. Then I^* is increasing at some right-neighbourhood of Ψ^0. Furthermore, for any $k \in \{1, \ldots, 2n\}$. If k is odd then I^* is increasing at some left-neighbourhood of Ψ^k and if k is even then I^* is decreasing at some right-neighbourhood of Ψ^k. Finally $\Psi^{2n} < 1 - \frac{r}{c_E}$.

Proof: Follows directly from continuity of I^* and from $\theta_M(\cdot) < \theta_i$. ∎

Assuming that the media ever finds itself constrained by its strength ($\theta_M(\cdot) > \hat{\underline{\theta}}_\Psi$) at some range of Ψ, lemma 14 suggests an irregular flat-topped inverse-U shaped equilibrium strategic response by media to increases in its journalistic strength. One can say that I^* is "on average" increasing for low values of Ψ and decreasing for high values of Ψ.
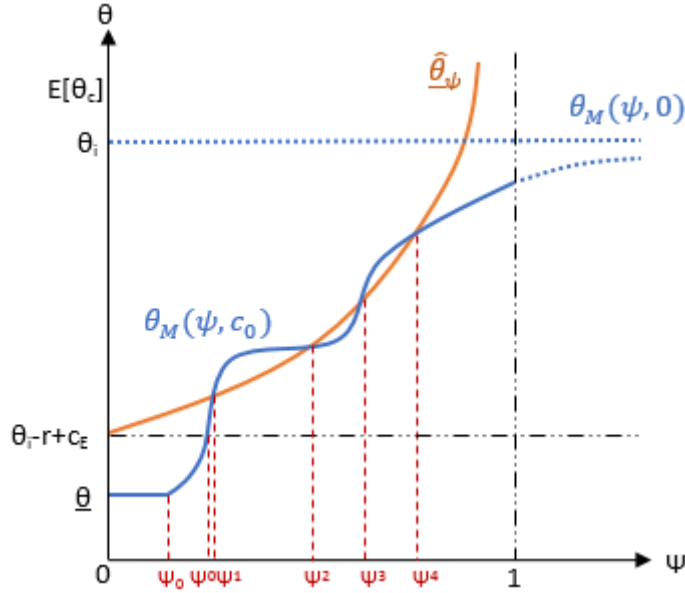

Figure 8: Media Strength and Equilibrium Transition - Below Average Incumbent (Irregular Case)
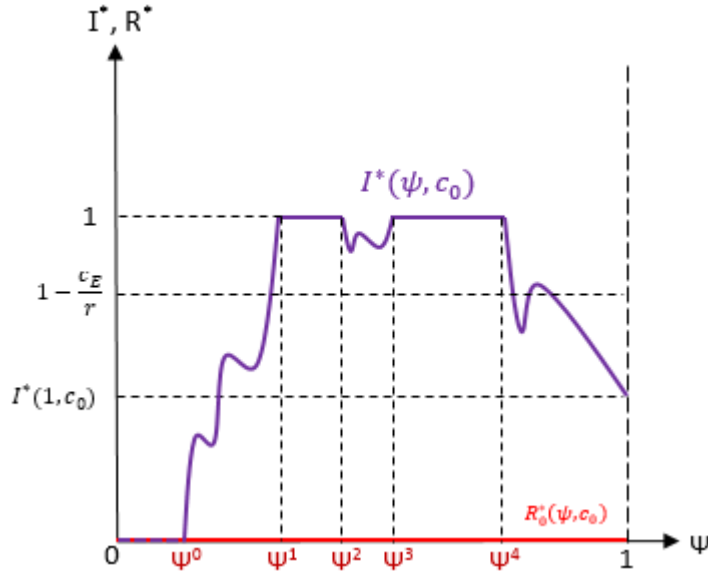
Figure 9: (Irregular) Strategic Effects of Media Strength - Below Average Incumbent

Figure 8 shows an irregular $\theta_M(\cdot)$ which starts convex (as Ψ exceeds $\Psi_0 = \frac{c_I}{K(\theta_i)}$), then bounces back and forth between convexity and concavity while intersecting $\hat{\underline{\theta}}_\Psi$ 4 times as Ψ is varied before turning fully concave.[54]. Lemma 14 allows one to establish the increasing behaviour of I^* at the immediate right-neighbourhood of Ψ^0, immediate left-neighbourhoods of Ψ^1 and Ψ^3 and decreasing behaviour of it at immediate right-neighbourhoods of Ψ^2 and Ψ^4. Meanwhile in the background, R_0^* is always zero due to weak incumbent quality. Notice that figure 9 also shows that as $\Psi \to 1$, I^* should fall below $1 - \frac{c_E}{r}$. This was already mentioned in previous propositions where various regimes of equilibria were presented, and the reason is that if $I^* \geq 1 - \frac{c_E}{r}$ when $\Psi = 1$, then adverse selection is fully eliminated, which we saw to be incompatible with costly investigation. Following lemma shows that the irregular behaviour of I^* disappears if $\theta_M(\cdot)$ is strictly concave in Ψ for $\Psi > \Psi_0$ as depicted in figure 6.

Lemma 15 (Regularly non-monotonic strategic response to changes in media strength): Suppose that $\theta_M(\cdot)$ is strictly concave for $\Psi > \Psi_0$. Then if it intersects $\hat{\underline{\theta}}_\Psi$, it should intersect it exactly twice.[55] Furthermore, given a parametric configuration satisfying the transition pattern depicted in figure 6 and assuming that $c_I = c_0$; $I^* = 0$ for $\Psi \in [0, \Psi_1]$, it is continuously increasing

---

[54]Note the difference between Ψ_0 (prohibitively low strength) and Ψ^0 (strength at which θ_M becomes high enough to make sure media is no longer satisfied by full adverse selection outcome $\hat{\underline{\theta}}$)

[55]Except possibly if there is a tangency.

for $\Psi \in (\Psi_1, \Psi_2]$, $I^* = 1$ for $\Psi \in [\Psi_2, \Psi_3]$, and it is continuously decreasing for $\Psi \in (\Psi_3, 1]$.

Proof: Appendix.　　　　　　　　　　　　　　　　　　　　　　　■

In the current model, concavity/convexity behaviour of $\theta_M(\cdot)$ (above Ψ_0) can be isolated down to shape of the distribution governing θ_c. For instance, if θ_c is uniformly distributed, then $\theta_M(\cdot)$ is indeed strictly concave at its increasing portion as the figure 6 suggests. More generally, the last part of the proof of lemma 10 shows that $\theta_M(\cdot)$ is concave for $\Psi > \Psi_0$ if (but not only if) $\frac{\partial^2 H(\theta_M, \theta_i)}{\partial \theta_M^2} \leq 0$ for all $\theta_M \leq \theta_i$ for a given θ_i, where $H(\cdot)$ is given as follows.

$$H(\theta_M, \theta_i) = \frac{1}{1 - F(\theta_M)} \int_{\theta_M}^{\theta_i} (\theta_i - \theta_c) f(\theta_c) \, \mathrm{d}\theta_c. \tag{57}$$

A sufficient condition for this is the distribution of θ_c satisfying *convex mean residual life* (CMRL) property.[56] Concavity of H amounts to saying that the decrease in the average quality gap between the incumbent and inferior challengers becomes more pronounced as the minimum challenger quality increases. As a result, required increases for media standard regarding minimum challenger quality following strength gains gradually dampen down. The figure below gives the equilibrium strategic responses corresponding to figure 6 by fixing information costs at $c_I = c_0$ and varying Ψ from 0 to 1.
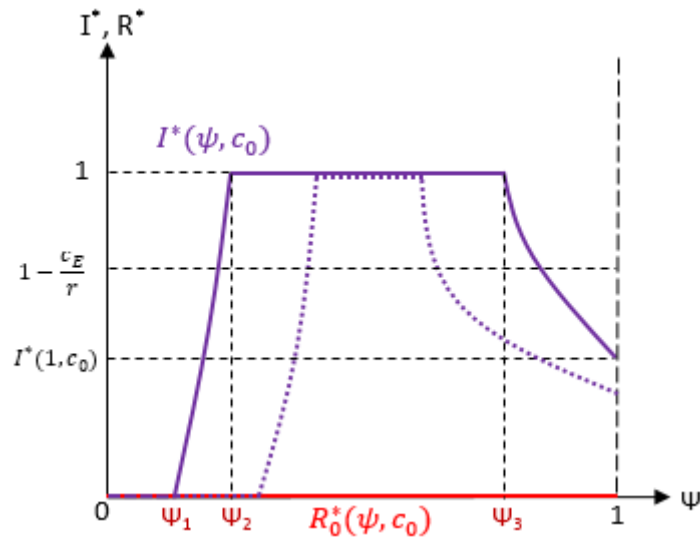

Figure 10: (Regular) Strategic Effects of Media Strength - Below Average Incumbent

---

[56]See Belzunce and Shaked (2001). Examples of commonly used distributions satisfying CMRL property are exponential, Pareto and Weibull.

The reason for the disappearance of irregularities when concavity of $\theta_M(\cdot)$ holds lies in a fact revealed in the proof of lemma 15. A key driver of $\frac{\partial I^*}{\partial \Psi}$ are the relative magnitudes of $\frac{\partial \theta_M(\cdot)}{\partial \Psi}$ and $\frac{\partial \hat{\theta}_\Psi}{\partial \Psi}$. The former represents the marginal increase in media standard due to gains in cost-efficiency, and the latter describes the marginal gains in deterrence power by threat of exposure. If the former is non-smaller than the latter, then I^* is necessarily increasing. When $\theta_M(\cdot)$ is concave at its increasing range, this ensures that the relative ranking of these two marginal effects change at most once, which allows one to conclude that I^* should either be always increasing (e.g. high information costs such as c_1 in figure 6) or it should be (weakly) increasing up to a point and then decreasing for all Ψ above that point. Imposing a parametric configuration which ensures that an equilibrium with $I^* = 1$ occurs then acts as an identification condition, which recovers the flat-topped inverse-U shaped strategic media response to gains in journalistic strength depicted in figure 10. Figure 10 also demonstrates the impact of an increase in information costs, which "squeezes" media's equilibrium strategy schedule, making it activate at a higher strength and start reducing its activity at a lower strength.

Next, I consider the case when incumbent quality is above average. For simplicity, I only consider the regular case, but a slightly reworded version of lemma 14 still holds (and if the media ever becomes active there will still be a single jump as $\theta_M(\cdot)$ never decreases). Figure 3 is reproduced below with additional media standard curves corresponding to different levels of investigation costs.
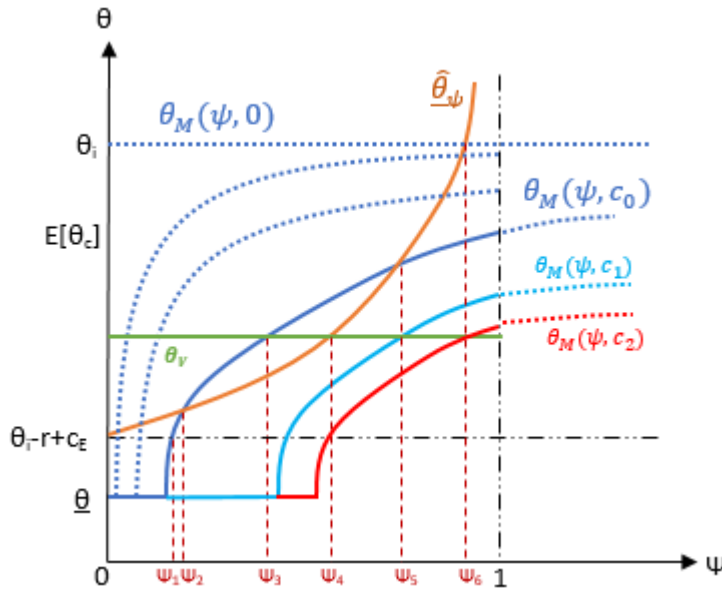

Figure 11: Media Strength, Costs and Equilibrium Transition - Above Average Incumbent

The only difference between figure 11 and figure 6 is the presence of an above average incumbent quality, which ensures $\theta_V(\cdot) \geq \underline{\theta}$. In the figure, it is assumed that θ_i is sufficiently high so that $\theta_V(\cdot) > \underline{\theta}$. Again, I start by fixing several different levels of media strength and let the information costs vary to see the strategic responses of the media and voters.
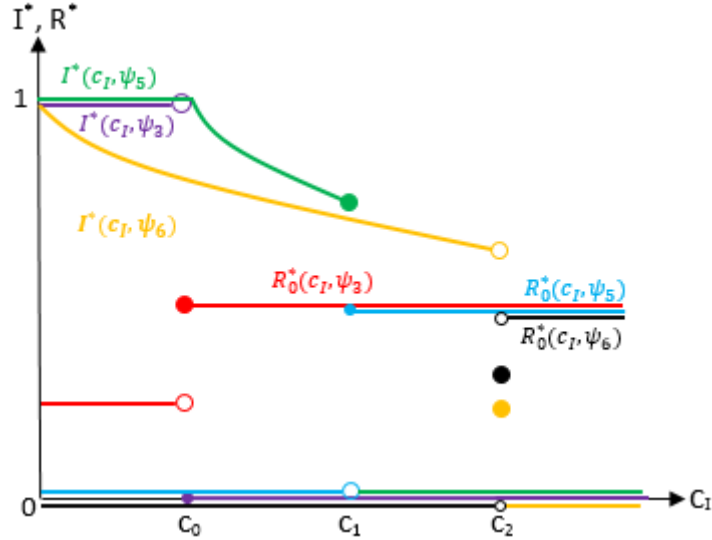
Figure 12: Strategic Effects of Information Costs - Above Average Incumbent

Figure 12 depicts strategic effects of varying information costs when media strength is fixed at three different levels from figure 11: Ψ_3, Ψ_5, Ψ_6. Effects at Ψ_5 and Ψ_6 are straightforward. For the former, media is fully active up to c_0, for that this is the level of information costs below which its standard is above the minimum quality it can achieve using its investigation technology. As costs exceed c_0, media starts monotonically decreasing its level of activity due to decreased cost-efficiency of eliminating adverse selection. Once costs exceed c_1, its standard becomes lower than the minimum challenger quality acceptable to voters, so the roles switch and voters start dealing with adverse selection by assigning a positive probability to retaining the incumbent. Effects of cost increases under Ψ_6 is similar, except for the fact that media never becomes fully active due to an already sufficiently strong investigation technology, and that due to cost-efficiency gains, the cost level above which its standard falls below that of the voters increases. Under both Ψ_5 and Ψ_6, media's selection constraint exceeds the standard of voters, i.e. the minimum challenger quality that the media can induce accommodates voters' preferences. This changes when $\Psi = \Psi_3$. When media strength is that low, even if the media is always active, it cannot provide voters with the high quality challenger pool they demand ($\theta_V(\cdot) > \hat{\underline{\theta}}_{\Psi_3}$). Consequently, up to c_0,

media is fully active and voters retain the incumbent with positive probability. Once investigation costs exceed c_0, media ceases activation, which is when voters increase their threat of retaining the incumbent to the level necessary when media is passive. In line with corollary 4, both media's and voters' strategies exhibit a single jump, which occurs when their standards are identical. Furthermore, directions of jumps satisfy the directions indicated in lemma 13. In figure 12, I consider various jump patterns emphasizing their indeterminacy, except for their monotonicity and the requirement that they should satisfy equation (47) in proposition 7.1.
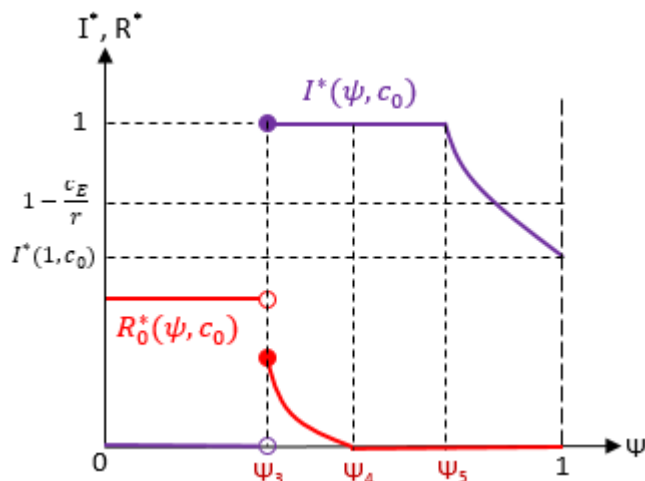
Figure 13: Strategic Effects of Media Strength - Above Average Incumbent

Figure 13 shows strategic effects of increasing media strength when information costs are fixed at $c_I = c_0$ in figure 11. There are two important differences between figure 13 and figure 10. As in figure 10, improvements in media strength first increase, then decrease media activity. But unlike figure 10, the increase is sudden in figure 13 and takes the form of a jump to full activation instead of a gradual increase. The second difference concerns voters' behaviour. Unlike the case when incumbent quality is below average, voters have to induce some positive selection by threatening to retain the incumbent when the media is unable to match their standards due to cost-inefficiency of its information technology. Even if media strength becomes sufficiently high so that the standards of media and voters match, voters do not immediately cease their threat of retaining the incumbent, as at first, media's ability to induce selection is still limited. So between Ψ_3 and Ψ_4, as the media becomes stronger, voters gradually reduce their threat of keeping the incumbent as they become more and more satisfied with the minimum challenger quality provided by a fully active media. Also, when media's

investigation technology becomes perfect, media selection and voter selection become perfect substitutes. In that case, $I^*(1, c_0) > R_0^*(\Psi, c_0)$ for $\Psi < \Psi_3$ reflects the overall higher standard of the media relative to voters ($\theta_M(\cdot) > \theta_V(\theta_i)$) whenever $\Psi = 1$.

Before concluding, it is worth considering the nature of the results regarding the response of media activity to improvements in its investigation technology and reflecting on how they would hold up against relaxation of some of the assumptions of the model. The assumption that candidate-types face positive costs for challenging implies that they care about the probability of winning and losing. In turn, this makes sure that the media can increase or decrease the extent of adverse selection continuously by mixing its investigation strategy, which is interpreted as corresponding to different journalistic activity levels. This setup, in essence, makes the problem of the media analogous to a problem where it is choosing a continuous activity level (having a certain probability of success) between zero and one with the goal of maximizing its (expected) payoff by equating its marginal benefits to its marginal costs. Parameter configurations under which the media is playing a pure strategy can be interpreted as parametric regions where certain constraints bind (either related to absolute prohibitiveness of its costs, weakness of its technology, or voters having relatively higher standards), forcing media to a corner solution where such marginal equalization does not hold. Whenever such corner solutions arise, changes in investigation strength does not affect media's behaviour on the margin, but can lead to discontinuous upward jumps by relaxing previously binding constraints. If the media is playing a mixed strategy, it is operating on the margin and changes in media strength exhibit a continuous response by affecting media's marginal benefit relative to its constant marginal cost c_I. While a reduction in information costs always (weakly) increase media activity, a downturn in media activity can be observed in response to an improvement in journalistic strength when marginal gains in deterrence power (at the given level of activity) exceed marginal net (taking costs into account) benefit of an increased investigation success.

Costly challenge is key in this picture. If challenging was costless ($c_E = 0$), then it is always possible to construct an equilibrium where every candidate-type with $\theta_c \geq \theta_i - r$ challenges even if they have no chance of winning at all. In that case, depending on how the average challenger stands out against the incumbent, uninformed voters would either retain the incumbent or vote for the challenger. Contingent on voters' subsequent behaviour, media would then base its decision either on the average gain from exposing inferior challengers or the average gain

from revealing a superior challenger. If costs are not too high, then media's behaviour would be characterized by complete passiveness up to a certain level of Ψ, above which it would be fully active.

In addition, one could have assumed that political rents are increasing in candidate quality. This assumption would not alter the nature of equilibria, nor the qualitative conclusions as long as the type-monotonicity of challenging decision remains and voters still have an incentive to vote for the politician with the highest quality. For instance, suppose that a challenger receives $\theta_c(1 + r) - c_E$ if she wins the election (i.e. a higher productivity also implies higher rents). Then one can check that adverse selection will be present as long as $\frac{c_E}{r} < \theta_i$ and that the decision to challenge preserves its upward monotonicity in quality levels (i.e. if a candidate-type finds it worthwhile to challenge, so does a higher quality candidate-type). As a result, starting with lemma 1, all of the previous results would still be valid after appropriate modifications accommodating the new form of candidate payoffs and the resulting new cutoffs. More generally, as long as candidate payoffs are single crossing in $(\theta_c, \tilde{C})$ ($\tilde{C} \in \{0, 1\}$ is the challenge decision) and at the end of the day the highest quality candidate produces the highest amount of public good regardless of the rents she receives, the set of equilibria would be isomorphic to the current one and the conclusions would remain intact.

Moreover, the model can be modified slightly to maintain its signalling structure while allowing for multiple candidates deciding on electoral participation. By modifying media's investigation technology so that it only partitions the set of challengers into inferior and superior (relative to the incumbent) challengers if successful and letting voters randomize uniformly over superior challengers when informed and over all challengers when uninformed (if they decide to replace the incumbent under uncertainty), the only difference one would get would be a rescale effect in challengers' individual probabilities of victory.[57]

Furthermore, dropping the assumption that voters are sophisticated enough to make Bayesian deductions would only make the media more relevant (because this opens the possibility of *ex ante* mistakes, in addition to *ex post* wrong decisions) as well as keeping the non-monotonic activity pattern intact. For instance, suppose that voters base their voting decisions on unconditional averages alone, i.e. always retain an above average incumbent and replace a below average one.

---

[57]This would also constitute an intuitive modification as with a large set of challengers, it wouldn't be realistic to expect the media revealing all individual qualities with precision. In fact, two technologies are effectively identical when there is a single candidate. If one were to assume that all qualities are revealed with multiple challengers, then one would have to take into account that even if a challenger is superior compared to the incumbent, she can still be inferior compared to another challenger.

Media behaviour concerning the latter case would be identical to its behaviour in the below average incumbent equilibria in the current model. In the former case, only superior candidate-types would challenge but they would be elected if and only if media successfully revealed their type. So a higher journalistic activity would imply a larger challenger pool and media's strength would constrain its ability to attract superior challengers (so a weak media would be able to attract only the top end of the distribution). Media would compare expected benefit from introducing a superior politician to voters with costs of journalism, which would determine its optimal lower bound for a superior challenger pool. Notice that unlike the case with adverse selection at the challenger front, a case where voters default to incumbent would imply adverse selection when the incumbent is retained, so the media would want to make sure that the probability of a candidate-type challenging is high enough. The set of equilibria would be similar and one would again potentially observe a non-monotonic media response due to binding-relaxing technology constraint.

Finally, suppose that the media receives an additional exogenous benefit for successfully revealing challenger quality (reputation, or advertisement revenues for high quality journalism). One can easily implement this into the current framework by letting $c_I = \tilde{c}_I - \Psi b$ where b is the benefit. In that case, there are two possibilities. If $\tilde{c}_I > b$, then the set of equilibria would be identical up to a scale. On the other hand, if there exists a Ψ^* where $\frac{\tilde{c}_I}{\Psi^*} = b$, media would always investigate for $\Psi > \Psi^*$. Depending on the magnitude of Ψ^*, it would then be possible to get a media response similar to the one in figure 10 featuring an upwards jump at $\Psi = \Psi^*$, assuming relatively weak challenger signalling.[58] In other words, media activity would display a hump-shaped response to improvements in its strength: first increasing, then decreasing, then increasing again by an upwards jump.

3   Conclusion

This chapter has built a pure adverse selection model of political entry featuring an idealized media with limited screening powers to study two interrelated questions: how would a stronger media affect *ex ante* politician selection and how would media itself respond to improvements in its screening technology. A motivating factor was to establish a framework highlighting media's neglected role in dealing with *ex ante* selection in politics. One of the most obvious ways the media

---

[58]The reason for the jump would be a "true indifference" occurring at $\frac{\tilde{c}_I}{\Psi^*} = b$, where media's equilibrium mixed best-response correspondence would no longer be unique.

can alleviate adverse selection is by exposing inferior politicians, allowing voters to make an informed decision. Endowing an idealized media with the technology of doing so but allowing that technology to be imperfect provides a simple way of mapping media strength to the quality of political participation within an endogenous information provision framework. Relatedly, a second motivation was to provide an explicit account for the mechanism via which selection by media threat operates, and to see whether any non-trivial pattern of journalistic activity arises in response to improvements in media's information technology. The reason for having an a priori suspicion for that it might stemmed from the nature of a media who acts as voters' extended arm. If the extent of the adverse selection is the sole factor that makes a media tick, then there is a theoretical possibility of a strong but relatively passive media capitalizing on the threat it poses to low quality challengers.

Several results have arisen from the model, including those that confirm this intuition. First, even if voters are sophisticated in their beliefs and vote correctly on average under uncertainty, media will still be relevant as long as its costs are low relative to the scale of adverse selection because the technology it possesses allows it to spare voters from a retrospective mistake. Second, an improvement in media's technology, whether it consists of cost reductions or gains in journalistic strength, always improves selection by shrinking the set of inferior candidate-types who challenge in equilibrium. The mechanism through which this selection effect operates involves the media adjusting its activity level to make sure that challenger quality is made public more frequently. This lowers the probability of an electoral victory for inferior challengers (and inferior challengers alone), reducing the magnitude of adverse selection. Relatedly, media is always (weakly) more active when adverse selection is stronger, which ultimately depends on the structure of political payoffs, i.e. how high the political rents are compared to costs of political participation. Finally, strategic effects of an improvement in media technology is considered. A lower cost of journalism always translates to a more active media, as expected. However, improvements to media's investigation strength has a non-monotonic effect on its activity. For a relatively weak media, it leads to an increase in journalistic activity and for a relatively strong media, it leads to a more passive ("asleep") media. The reason behind this asymmetry between costs and strength is that while cost reductions only affect the net benefits of journalism, strength has a dual effect. It not only increases net (expected) benefits of an investigation, but also increases the threat posed by the media to an inferior challenger at a given level of activity. This latter effect implies that

120

if the media is already quite strong, it can reduce its costly journalistic activities without undermining deterrence, an effect that can be best called the paradox of good (strong) journalism.

Extending the boundaries of the current model to incorporate elements of moral hazard is no trivial task. Besides rethinking the incumbent as an active player instead of a mere payoff parameter, it would involve issues such as introducing a type-dependant tradeoff between rents and public good due to an underlying choice variable reflecting the effort allocated to public services in contrast to rent extraction (and either suppression of politicians' public good motivation altogether or making rent extraction more beneficial), as well as recognition of the fact that media's beliefs and strategies would have to take incumbent's output signals and challenger's post-election behaviour into account. However, there are several potentially fruitful extensions that can be considered while remaining within a pure adverse selection framework. First, the current model can easily be extended to incorporate a non-singleton set of candidates simultaneously deciding whether to challenge or not in the way discussed in the last part of the final section. A more interesting approach would be to let the media (and only media) observe a set of noisy signals correlated with challenger qualities and decide on which candidate to investigate basing on that set of signals. Even if media's investigation fails to resolve the uncertainty, the fact that media has singled out a specific candidate can affect voter behaviour, and in particular cause them to vote for that specific challenger (if they are not voting for the incumbent). This would give an additional strategic motive for media activation and in turn affect candidate behaviour, presumably in a form that excludes a subset of inferior challengers. Furthermore, it would add an additional dimension of media strength (rumor/pre-screening accuracy), which can affect media activity not only in isolation but also by its interaction with media's investigative strength. Next, one of the major reasons why the need for a media arises in the current model (besides payoff-induced adverse selection) is the restricted nature of the message space. For instance, if it was possible for challengers to choose their campaign expenditures and if spending $c_E \geq r$ was a feasible option, then one would expect the existence of a semi-separating equilibrium with all non-inferior candidate-types pooling at $c_E = r$ and all inferior candidate-types staying out (or mounting a cheap challenge with $c_E = 0$), leading to full resolution of negative selection. In essence, this underlines the fact that media's technology of type discovery and candidates' signalling technology are substitutable in voters' eyes. However, an electorate rarely possesses tools to costlessly and accurately distinguish between

different campaign types. Under a setting of unobservable/partially observable (for voters) actions, media can potentially provide a valuable service of signal transmission by undertaking a costly investigation to discover and validate a challenger's participation costs and perhaps packaging its discovery in the form of an endorsement. Its ability of doing so would in turn affect candidates' strategies. This line of reasoning points out towards potential complementarities between candidate signalling and media activity, which could be studied within the current framework after appropriate modifications. Finally, the proposed framework of media selection can be extended in a variety of ways in order to shed light on numerous possible interactions between the structure of media industry, (political) journalism and politician selection. For instance, suppose that there are two media outlets and their journalistic activities are complementary in the sense that both outlets investigating makes truth discovery more likely. Such a setting, especially if both outlets possess sufficiently strong investigative technologies, would possibly give rise to a problem of free-riding, not only creating an additional mechanism for media passivity under increased journalistic strength, but also threatening to undermine selection. This could in turn underline factors such as journalistic benefits and media exclusives as important determinants of media activity, and thus selection. These issues are left for future research.

Appendix

Proof of Lemma 1 (Connectedness of the set of strict challengers and uniqueness of the indifferent candidate): Fix a PBE with $I^*, R_0^*, R_1(\theta_c)$ and Ω_p. If $\Omega_p = \emptyset$, then it is connected. Consider $\hat{\theta}_c(\theta_c)$ given by (22) and its various evaluations according to θ_c's position relative to θ_i given in (23). Following inequalities are obvious and hold for any $(I, R_0, R_1(\theta_c = \theta_i))$.

$$\hat{\theta}_{c,b} \leq \hat{\theta}_{c,e} \tag{58}$$

$$\hat{\theta}_{c,e} \leq \hat{\theta}_{c,w} \tag{59}$$

$$\hat{\theta}_{c,b} < \hat{\theta}_{c,w} \tag{60}$$

Suppose that there is an indifferent type θ^* in this equilibrium. Take an arbitrary $\theta \in [\underline{\theta}, \theta^*)$. If $\theta^* > \theta_i$, then $\theta^* = \hat{\theta}_{c,b}$, so from inequalities (58)-(60), $\theta < \hat{\theta}_c(\theta)$. If $\theta^* = \theta_i$, then $\theta^* = \hat{\theta}_{c,e}$, so $\theta < \theta_i = \theta^* = \hat{\theta}_{c,e} \leq \hat{\theta}_{c,w}$ and hence $\theta < \hat{\theta}_c(\theta)$. If $\theta^* < \theta_i$, then $\theta^* = \hat{\theta}_{c,b}$, so $\theta < \theta_i$ and $\theta < \hat{\theta}_{c,b}$, which imply $\theta < \hat{\theta}_c(\theta)$. Now take

an arbitrary $\theta' \in \left(\theta^*, \overline{\theta}\right]$. If $\theta^* > \theta_i$, then $\theta^* = \hat{\theta}_{c,b}$ so $\theta' > \hat{\theta}_c(\theta')$. If $\theta^* = \theta_i$, then $\theta^* = \hat{\theta}_{c,e}$, so $\theta' > \theta^* = \hat{\theta}_{c,e} \geq \hat{\theta}_{c,b}$ from the inequality in (58), from which it follows that $\theta' > \hat{\theta}_c(\theta')$. If $\theta^* < \theta_i$, then $\theta^* = \hat{\theta}_{c,w}$ and (60) implies that $\theta' > \hat{\theta}_c(\theta')$. Since θ and θ' were arbitrary, it follows that if it exists, θ^* is unique and a candidate with a higher quality (then θ^*) strictly prefers challenging.

Now, suppose that $\Omega_p \neq \emptyset$. Suppose first that $\Omega_p = \{\overline{\theta}\}$. Then, either $\overline{\theta} > \theta_i$ and $\overline{\theta} > \hat{\theta}_{c,b}$, or $\overline{\theta} = \theta_i$ and $\overline{\theta} > \hat{\theta}_{c,e}$. For the former, let $\tilde{\theta} = \overline{\theta} - \epsilon$ where $0 < \epsilon < \min\left\{\overline{\theta} - \theta_i, \overline{\theta} - \hat{\theta}_{c,b}\right\}$. Then $\tilde{\theta} > \theta_i$ and $\tilde{\theta} > \hat{\theta}_{c,b}$, so $\tilde{\theta} \in \Omega_p$, a contradiction. For the latter, note that if $\Omega_p = \{\overline{\theta}\}$ and $\overline{\theta} = \theta_i$, then either $\overline{\theta}$ is the only type who challenges with a positive probability, in which case (14) requires $\mathbb{E}\left[\theta_c \mid C = 1\right] = \theta_i$ and $\mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] = 0$, or there are some $\theta_c \in \left[\underline{\theta}, \overline{\theta}\right)$ who challenge with a strictly positive probability that is less than one, in which case (14) requires $\mathbb{E}\left[\theta_c \mid C = 1\right] < \theta_i$ and $\mathbb{E}\left[\max\{0, \theta_c - \theta_i\} \mid C = 1\right] = 0$, as the prior f has full-support. In both cases, (20) implies that $I^* = 0$, so from (23), $\hat{\theta}_{c,e} = \hat{\theta}_{c,w}$, and hence from the definition of Ω_p, $\overline{\theta} = \theta_i > \hat{\theta}_{c,e} = \hat{\theta}_{c,w}$. Let $\tilde{\theta} = \overline{\theta} - \epsilon$ where $0 < \epsilon < \overline{\theta} - \hat{\theta}_{c,w}$. Then $\tilde{\theta} < \theta_i$ and $\tilde{\theta} > \hat{\theta}_{c,w}$, so from (24), $\tilde{\theta} \in \Omega_p$, another contradiction. It follows that if $\Omega_p \neq \emptyset$, then $\Omega_p \neq \{\overline{\theta}\}$, i.e. $\Omega_p \cap \left[\underline{\theta}, \overline{\theta}\right) \neq \emptyset$. Pick an arbitrary $\theta_c \in \left(\Omega_p \cap \left[\underline{\theta}, \overline{\theta}\right)\right)$, and pick a θ'_c satisfying $\theta_c < \theta'_c \leq \overline{\theta}$. There are three cases to consider.

1. Suppose that $\theta_c < \theta_i$. Then by definition of Ω_p, $\theta_c > \hat{\theta}_{c,w}$. It follows that $\theta'_c > \hat{\theta}_{c,w}$. So if $\theta'_c < \theta_i$, then $\theta'_c \in \Omega_p$. If $\theta'_c = \theta_i$, then from (59), $\theta'_c > \hat{\theta}_{c,w} \geq \hat{\theta}_{c,e}$, so $\theta'_c \in \Omega_p$. If $\theta'_c > \theta_i$, then from (60), $\theta'_c > \hat{\theta}_{c,w} > \hat{\theta}_{c,b}$, so $\theta'_c \in \Omega_p$.

2. Suppose that $\theta_c = \theta_i$. Then by definition of Ω_p, $\theta_c > \hat{\theta}_{c,e}$. It follows that $\theta'_c > \hat{\theta}_{c,e}$. Moreover, $\theta'_c > \theta_i$ and from (58), $\theta'_c > \hat{\theta}_{c,e} \geq \hat{\theta}_{c,b}$, so $\theta'_c \in \Omega_p$.

3. Suppose that $\theta_c > \theta_i$. Then by definition of Ω_p, $\theta_c > \hat{\theta}_{c,b}$. It follows that $\theta'_c > \theta_i$ and $\theta'_c > \hat{\theta}_{c,b}$, so $\theta'_c \in \Omega_p$.

Since both θ_c and θ'_c were arbitrary, this proves that if Ω_p is non-empty, then it is an interval either of form $\left[\theta^*, \overline{\theta}\right]$ or of form $\left(\theta^*, \overline{\theta}\right]$, with $\theta^* < \overline{\theta}$.

Now suppose that Ω_p is a strict subset of $\left[\underline{\theta}, \overline{\theta}\right]$, i.e. $\inf \Omega_p = \theta^* > \underline{\theta}$. Note that $\theta > \theta^*$ implies $\theta \in \Omega_p$. I will show that the candidate with θ^* is indifferent between challenging and staying out, which will prove that $\Omega_p = \left(\theta^*, \overline{\theta}\right]$. Again, there are three cases to consider.

1. Suppose that $\theta^* < \theta_i$. First, assume that $\theta^* > \hat{\theta}_{c,w}$. Define $\tilde{\theta} = \theta^* - \epsilon$,

where $0 < \epsilon < \min\left\{\theta^* - \hat{\theta}_{c,w}, \theta^* - \underline{\theta}\right\}$.[59] Then $\underline{\theta} < \tilde{\theta} < \theta^* < \theta_i$ and $\tilde{\theta} > \hat{\theta}_{c,w}$, which (from (24)) implies that $\tilde{\theta} \in \Omega_p$, which contradicts with the fact that $\theta^* = \inf \Omega_p$. Next, assume that $\theta^* < \hat{\theta}_{c,w}$. Define $\tilde{\theta} = \theta^* + \epsilon$, where $0 < \epsilon < \min\left\{\theta_i - \theta^*, \hat{\theta}_{c,w} - \theta^*\right\}$. Since $\tilde{\theta} > \theta^*$, $\tilde{\theta} \in \Omega_p$. But $\tilde{\theta} < \theta_i$ and $\tilde{\theta} < \hat{\theta}_{c,w}$, so $\tilde{\theta} \notin \Omega_p$, a contradiction. It follows that if $\theta^* < \theta_i$, then $\theta^* = \hat{\theta}_{c,w}$.

2. Suppose that $\theta^* > \theta_i$. First, assume that $\theta^* > \hat{\theta}_{c,b}$. Define $\tilde{\theta} = \theta^* - \epsilon$, where $0 < \epsilon < \min\left\{\theta^* - \theta_i, \theta^* - \hat{\theta}_{c,b}\right\}$. Then $\tilde{\theta} < \theta^*$. Moreover, $\tilde{\theta} > \theta_i$ and $\tilde{\theta} > \hat{\theta}_{c,b}$, which implies (from (24)) that $\tilde{\theta} \in \Omega_p$, contradicting with $\theta^* = \inf \Omega_p$. Next, assume that $\theta^* < \hat{\theta}_{c,b}$. Define $\tilde{\theta} = \theta^* + \epsilon$, where $0 < \epsilon < \hat{\theta}_{c,b} - \theta^*$. Then $\tilde{\theta} > \theta^*$, so $\tilde{\theta} \in \Omega_p$. But we also have $\tilde{\theta} > \theta_i$ and $\theta^* < \hat{\theta}_{c,b}$, which (from (24)) implies that $\tilde{\theta} \notin \Omega_p$, a contradiction. It follows that if $\theta^* > \theta_i$, then $\theta^* = \hat{\theta}_{c,b}$.

3. Suppose that $\theta^* = \theta_i$. First, assume that $\theta^* > \hat{\theta}_{c,e}$. From (59), we know that $\hat{\theta}_{c,e} \leq \hat{\theta}_{c,w}$. Suppose that $\hat{\theta}_{c,e} = \hat{\theta}_{c,w}$. Then $\theta^* > \hat{\theta}_{c,w}$, and one obtains a contradiction identical to the one obtained in the first part of the point 1 above. Suppose instead $\hat{\theta}_{c,e} < \hat{\theta}_{c,w}$. If $\theta^* > \hat{\theta}_{c,w} > \hat{\theta}_{c,e}$, one obtains the same contradiction. If $\hat{\theta}_{c,w} \geq \theta^* > \hat{\theta}_{c,e}$, then note that for any $\theta < \theta^*$, one has $\theta < \hat{\theta}_c(\theta) = \hat{\theta}_{c,w}$, i.e. from (24), nobody below θ^* challenges. So from (14) and the fact that f is full support, one must have $\mathbb{E}\left[\theta_c \mid C = 1\right] > \theta_i$ and $\mathbb{E}\left[\max\left\{\theta_i - \theta_c, 0\right\} \mid C = 1\right] = 0$. Hence from (20), $I^* = 0$, and thus from (23), $\hat{\theta}_{c,e} = \hat{\theta}_{c,w}$, which contradicts with $\hat{\theta}_{c,e} < \hat{\theta}_{c,w}$. Next, assume $\theta^* < \hat{\theta}_{c,e}$. From (59), we know that $\hat{\theta}_{c,b} \leq \hat{\theta}_{c,e}$. If $\hat{\theta}_{c,b} = \hat{\theta}_{c,e}$, then $\theta^* < \hat{\theta}_{c,b}$, and one gets a contradiction identical to the one obtained in the second part of the point 2 above. So suppose instead $\hat{\theta}_{c,b} < \hat{\theta}_{c,e}$. If $\theta^* < \hat{\theta}_{c,b} < \hat{\theta}_{c,e}$, one gets the exact same contradiction. If $\hat{\theta}_{c,b} \leq \theta^* < \hat{\theta}_{c,e}$, then for any $\theta < \theta^* = \theta_i$, one has $\theta < \hat{\theta}_c(\theta)$, as (59) implies $\hat{\theta}_{c,e} \leq \hat{\theta}_{c,w}$. Since for any $\theta > \theta^* = \theta_i$, one has $\theta \in \Omega_p$, (14) and the fact that f is full-support implies that one must have $\mathbb{E}\left[\theta_c \mid C = 1\right] > \theta_i$ and $\mathbb{E}\left[\max\left\{\theta_i - \theta_c, 0\right\} \mid C = 1\right] = 0$. So from (20), $I^* = 0$, and hence from (23), $\hat{\theta}_{c,b} = \hat{\theta}_{c,e}$, which is another contradiction. It follows that if $\theta^* = \theta_i$, then $\theta^* = \hat{\theta}_{c,e}$.

This shows that when $\inf \Omega_p = \theta^* > \underline{\theta}$, then $\theta^* = \hat{\theta}_c(\theta^*)$, i.e. θ^* is indifferent between challenging and staying out. Since Ω_p is the set of candidates strictly

---

preferring to challenge, it should therefore have the form $\Omega_p = \left(\theta^*, \overline{\theta}\right]$. ∎

Proof of Lemma 2 (Anything goes for the candidate and informed voters when they are indifferent): Take a PBE defined by $\beta(\theta_c \mid C = 1)$ and a strategy profile $(\tilde{C}(\theta_c), I^*, R_0^*, R_1(\theta_c))$ and assume that $\Omega \neq \emptyset$. The latter implies that beliefs are pinned down by (14). Varying $R_1(\theta_c = \theta_i) = k \in [0, 1]$ will potentially destroy this equilibrium only if there exists an indifferent type $\theta^* = \theta_i$. Suppose there is. Then from lemma 1, either $\Omega_p = \emptyset$ and $\Omega = \{\theta_i\}$ (which is possible only if $\theta_i = \overline{\theta}$), or $\Omega_p = \left(\theta_i, \overline{\theta}\right]$ and $\Omega = \left[\theta_i, \overline{\theta}\right]$. In both cases, beliefs should impose $\mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] = 0$, which implies $I^* = 0$ (as long as $c_I > 0$), which implies that $\hat{\theta}_{c,e}$ is independent of $R_1(\theta_c = \theta_i)$ from (23). Since the candidate's strategy does not depend on it, neither do beliefs. Therefore, with a $k' \neq k$, (10)-(14) should still be satisfied and hence varying k along $[0, 1]$ results in a continuum of PBE. Now suppose that the given PBE features an indifferent type θ^* and let $g = \tilde{C}(\theta^*)$. Either $\theta^* = \overline{\theta}$, in which case $g \in (0, 1]$ and beliefs put all the mass on $\overline{\theta}$, in which case they are independent of g, or $\theta^* < \overline{\theta}$. In the latter case, lemma 1 and (14) implies that beliefs on Ω_p are simply f truncated at θ^* and a change in g would have no effect on that. Hence, a change in g would have no effect on the expectations in (16) and (20) and thus lead to no deviation incentives for the media and voters. It follows that varying g along $[0, 1]$ will simply result in a continuum of PBE, with only $\beta(\theta^* \mid C = 1)$ changing to reflect changes in $g = \tilde{C}(\theta^*)$. ∎

Proof of Lemma 4 (Equilibrium Refinements): Assume $\hat{\underline{\theta}} \leq \overline{\theta}$. First suppose that $\hat{\underline{\theta}} > \theta_i$ and consider some trivial pooling equilibrium as described in proposition 2. Note that this implies $r < c_E$, i.e. costs of participation cannot be fully recovered in case of an electoral victory. Consider the following sets of types.

$$\Phi = [\underline{\theta}, \hat{\underline{\theta}}); \tag{61}$$
$$\Phi' = [\hat{\underline{\theta}}, \overline{\theta}]. \tag{62}$$

Equilibrium payoff of the candidate is θ_i. Consider arbitrary candidate-types $\theta_c \in \Phi$ and $\theta_c' \in \Phi'$. Then following inequalities are true by definitions of Φ, Φ' and $\hat{\underline{\theta}}$:

$$\theta_c + r - c_E < \theta_i, \tag{63}$$
$$\theta_c' + r - c_E \geq \theta_i. \tag{64}$$

Furthermore, from (3), the maximum payoff that a candidate can get by playing off the equilibrium strategy of challenging is $\max\{\theta_c + r - c_E, \theta_i - c_E\}$.[60,61] It follows that the set Φ is equilibrium dominated in the Intuitive Criterion sense, while the set Φ' is not. So upon observing a challenge, the beliefs should restrict candidate quality to Φ'.[62] Due to $\hat{\underline{\theta}} > \theta_i$, this implies $\mathbb{E}\left[\theta_c \mid C = 1\right] > \theta_i$ and $\mathbb{E}\left[\max\{\theta_i - \theta_c, 0\} \mid C = 1\right] = 0$. If the beliefs satisfy these, then $I^* = 0$ and $R_0^* = 0$, i.e. the media never investigates and uninformed voters replace the incumbent with certainty. If this is the case, then a candidate with quality $\theta_c' \in \Phi'$ would get $\theta_c' + r - c_E$ if she were to challenge, which is higher than her equilibrium payoff θ_i. So the equilibrium is eliminated by the Intuitive Criterion. Since the set of equilibria that survive the Divinity Criterion is a subset of the set of equilibria that survive the Intuitive Criterion, this equilibrium is also eliminated by the Divinity Criterion.[63]

Next, suppose that $\hat{\underline{\theta}} \leq \theta_i$ and consider some trivial pooling equilibrium as described in proposition 2. Note that this implies $r \geq c_E$. Also note that Intuitive Criterion is no longer enough to eliminate this equilibrium, as, say, it allows for a belief density degenerate at $\hat{\underline{\theta}} \leq \theta_i$. However, it still refines the set $\Phi = [\underline{\theta}, \hat{\underline{\theta}})$ away from the above procedure, and so does the Divinity Criterion. Now define the following sets.

$$\tilde{\Phi} = [\hat{\underline{\theta}}, \theta_i]; \tag{65}$$

$$\Phi'' = \Phi' \setminus \tilde{\Phi} = (\theta_i, \overline{\theta}]. \tag{66}$$

I will show that Divinity Criterion refines $\tilde{\Phi}$ away but leaves Φ'' intact. Define the sets $\mathcal{D}(\theta \mid C = 1)$ and $\mathcal{D}_s(\theta \mid C = 1)$ as the sets of media and voter strategies (both mixed and pure) that yields non-smaller and strictly greater (expected)

---

[60]Strictly speaking, Intuitive Criterion requires that when computing the maximum payoff a candidate can get by playing an off the equilibrium strategy, one should assume that receivers best respond to such strategy by pure strategies and by assuming that $\theta \in [\underline{\theta}, \overline{\theta}]$. This latter assumption requires receivers' (media and voters) pure strategies to be constrained to the "set of best-responses to $[\underline{\theta}, \overline{\theta}]$" (see the next footnote). Here, that set contains all possible actions by the media and uninformed voters.

[61]In general, a set of best responses to some set Λ is defined by fixing some belief under the sole assumption that the sender (challenger) is in Λ, finding the set of best-responses to that belief, and taking the union of such sets of best responses over all feasible beliefs under that assumption. Technically, such feasible beliefs are required to be measurable distributions assigning a positive measure only to measurable subsets of Λ, and belief densities to be Radon-Nikodym derivatives of such distributions. As such, there are uncountably many of them as long as Λ is a set of positive measure.

[62]And that is the only restriction on beliefs. So any belief distribution assigning a positive measure only to subsets of Φ' is allowed.

[63]See Muñoz-García and Espínola-Arredondo (2011).

payoffs respectively compared to the equilibrium payoff θ_i for the candidate with quality θ who came to play the off-the-equilibrium strategy of challenging.[64] A type $\tilde{\theta}$ is eliminated under Divinity Criterion if there exists another type θ'' such that the following holds.

$$\mathcal{D}(\tilde{\theta} \mid C = 1) \subset \mathcal{D}_s(\theta'' \mid C = 1). \tag{67}$$

Fix a $\theta'' \in \Phi''$ and consider an arbitrary $\tilde{\theta} \in \tilde{\Phi}$. First, assume that $\tilde{\theta} = \hat{\underline{\theta}}$. This candidate's quality satisfies $\theta_c + r - c_E = \theta_i$, so the only receiver strategies that don't make her worse off are the ones giving her a probability one of winning the election. But those strategies would also give a probability one of winning to θ'', making her better off as $\theta'' > \tilde{\theta}$. Furthermore, since $\theta''_c + r - c_E > \theta_i$, θ'' can be made strictly better off by strategies giving her a probability of victory less than one (this follows from rearranging the third line in (23)). This eliminates $\tilde{\theta} = \hat{\underline{\theta}}$. Next, assume that $\hat{\underline{\theta}} < \tilde{\theta} < \theta_i$. Consider any pair (R_0^*, I^*) that makes $\tilde{\theta}$ no worse off by giving her a victory probability $(1 - R_0^* + \Psi I^*(R_0^* - 1))$. But then the victory probability of θ'' is $(1 - R_0^* + \Psi I^*(R_0^* - 0))$, which is strictly greater, so she should be strictly better off. Now consider a strategy that consists of $R_0^* = 1$ and $I^* = 1$. This gives a certain victory to θ'' and a sure loss to $\tilde{\theta}$, making the former better off and the latter worse off. This eliminates any $\tilde{\theta} \in (\hat{\underline{\theta}}, \theta_i)$. Next, assume that $\tilde{\theta} = \theta_i$. If $\theta_i = \hat{\underline{\theta}}$, then this is eliminated so assume $\theta_i > \hat{\underline{\theta}}$. Consider any triplet $(R_0^*, I^*, R_1(\theta_c = \theta_i))$ that makes $\tilde{\theta}$ no worse off. From (23), one can see that these strategies will give a greater or equal victory probability to θ''. Furthermore, $\theta'' > \theta_i$, so she should be strictly better off. Finally consider a strategy that consists of $R_0^* = 1$, $I^* = 1$ and $R_1(\theta_c = \theta_i) = 1$. This gives a certain victory to θ'' and a sure loss to $\tilde{\theta} = \theta_i$, making the former better off and the latter worse off. So $\tilde{\Phi}$ is equilibrium dominated in the Divinity Criterion sense. In addition, pick some $\theta''' \in \Phi''$. Since both θ'' and θ''' are better than the incumbent, they have the same payoffs and hence have identical $\mathcal{D}$'s and $\mathcal{D}_s$'s. It follows that upon observing a challenge, the media and voters should conclude that the challenger belongs to $\Phi'' = (\theta_i, \overline{\theta}]$. If this is the case, then any beliefs given this restriction should satisfy $\mathbb{E}[\theta_c \mid C = 1] > \theta_i$ and $\mathbb{E}[\max\{\theta_i - \theta_c, 0\} \mid C = 1] = 0$, implying $I^* = 0$ and $R_0^* = 1$. But then, any candidate with a type $\theta > \hat{\underline{\theta}}$ would be strictly better off challenging, so the equilibrium does not survive Divinity Criterion. ∎

---

[64]Again, Divinity Criterion in reality constrains voter and media strategies in $\mathcal{D}$ and $\mathcal{D}_s$ to be best-responses to $[\underline{\theta}, \overline{\theta}]$. But as indicated by the discussion in footnotes 60-61, this is a "slack constraint", as any feasible belief over $[\underline{\theta}, \overline{\theta}]$ is allowed in the definition of these sets, and so they allow any strategy to be considered.

Proof of Lemma 7 (Limits to strategic selection): Assume $\hat{\underline{\theta}} < \theta_i$ and consider a PBE with $\Omega \neq \emptyset$.

1. Suppose $\theta_i = \bar{\theta}$. Ω being assumed non-empty, assume instead that $\inf \Omega < \bar{\theta}$. Then lemma 1 and f being full-support along with lemma 5 implies that beliefs should satisfy $\mathbb{E}\left[\theta_c \mid C = 1\right] < \theta_i$ and $\mathbb{E}\left[\max\left\{0, \theta_c - \theta_i\right\} \mid C = 1\right] = 0$, implying $R_0^* = 1$ and $I^* = 0$. But from (24), this would imply $\Omega = \emptyset$, a contradiction. So from lemma 3, $\Omega = \left\{\bar{\theta}\right\}$ and $\Omega_p = \emptyset$. Moreover, because the media does not investigate under this scenario ($I^* = 0$), R_0^* should adjust to make sure that $\theta_c = \bar{\theta}$ is indifferent between challenging and staying out, i.e. so that $\hat{\theta}_{c,e} = \theta_i = \bar{\theta}$ in equilibrium. Rearranging this using (23) and the fact that $I^* = 0$, one can see that there exists a unique $R_0^* \in (0, 1)$ so that this is true. R_0^* should in turn satisfy the following.

$$\frac{r}{c_E} = \frac{1}{1 - R_0^*}, \tag{68}$$

   because $r > c_E$ (which is implied by $\hat{\underline{\theta}} < \theta_i$).

2. Suppose $\theta_i = \underline{\theta}$. Then if $\Omega \neq \emptyset$, from lemma 1 and f being full-support, beliefs should imply $\mathbb{E}\left[\theta_c \mid C = 1\right] > \theta_i$ and $\mathbb{E}\left[\max\left\{\theta_i - \theta_c, 0\right\} \mid C = 1\right] = 0$, implying $I^* = R_0^* = 0$ from (16) and (20). From (23), this implies $\hat{\theta}_{c,w} < \underline{\theta}$, so (24) implies that all candidates challenge, i.e. $\Omega_p = \left[\underline{\theta}, \bar{\theta}\right]$.

3. Suppose $\theta_i \in \left(\underline{\theta}, \bar{\theta}\right)$, and suppose instead $\theta^* \geq \theta_i$. Again, if $\Omega \neq \emptyset$, lemma 1 and f being full-support makes sure that beliefs induce $I^* = R_0^* = 0$. But then from (23) and (24); $\inf \Omega = \max\left\{\underline{\theta}, \hat{\theta}\right\}$, a contradiction. ∎

Proof of Lemma 10 (Comparative statics - Media and voter cutoffs): Assume that $\theta_V(\theta_i) \in (\underline{\theta}, \theta_i)$ solves the following equation:

$$\theta_i - \mathbb{E}\left[\theta_c \mid \theta_V \leq \theta_c\right] = 0. \tag{69}$$

Such $\theta_V(\theta_i)$ exists and is unique when $\theta_i > \mathbb{E}\left[\theta_c\right]$ from lemma 8. Since the density f is continuous, $\mathbb{E}\left[\theta_c \mid \theta_V \leq \theta_c\right]$ is continuously differentiable in $(\underline{\theta}, \theta_i)$ from the fundamental theorem of calculus. So an application of the implicit function theorem gives:

$$\frac{\partial \theta_V(\theta_i)}{\partial \theta_i} = \left(\frac{\partial \mathbb{E}\left[\theta_c \mid \theta_V \leq \theta_c\right]}{\partial \theta_V}\right)^{-1} > 0, \tag{70}$$

where the derivative at the right-hand side is evaluated at $\theta_V = \theta_V(\theta_i)$ and its sign follows from the fact that f is full-support. Furthermore, if f is (strictly) log-concave, then $\frac{\partial \mathbb{E}[\theta_c | \theta_V \leq \theta_c]}{\partial \theta_V} \in (0, 1] \, (\in (0, 1))$.[65] So if f is (strictly) log-concave, then the following holds.

$$\frac{\partial \left(\theta_V(\theta_i) - \theta_i \right)}{\partial \theta_i} \geq (>)0. \tag{71}$$

Similarly, assume that $\theta_M(\Psi, c_I, \theta_i) \in (\underline{\theta}, \theta_i)$ solves the following equation.

$$\frac{F(\theta_i) - F(\theta_M)}{1 - F(\theta_M)} \left\{ \theta_i - \mathbb{E}\left[\theta_c \mid \theta_M \leq \theta_c < \theta_i\right] \right\} - \frac{c_I}{\Psi} = 0. \tag{72}$$

Such $\theta_M(\Psi, c_I, \theta_i)$ exists and is unique when $0 < \frac{c_I}{\Psi} < K(\theta_i)$ from lemma 9. Let $H(\theta_i, \theta_M)$ denote the first term at the left-hand side of (72). Since the density f is continuous, $H(\theta_i, \theta_M)$ is continuously differentiable in $(\underline{\theta}, \theta_i)$ from the fundamental theorem of calculus. Proof of lemma 9 has already shown that $\frac{\partial H(\theta_i, \theta_M)}{\partial \theta_M} < 0$. So applying the implicit function theorem three times to the equality in (72) yields:

$$\frac{\partial \theta_M(\Psi, c_I, \theta_i)}{\partial c_I} = \left(\frac{\partial H(\theta_i, \theta_M)}{\partial \theta_M} \right)^{-1} < 0; \tag{73}$$

$$\frac{\partial \theta_M(\Psi, c_I, \theta_i)}{\partial \Psi} = -\left(\frac{c_I}{\Psi^2} \right) \left(\frac{\partial H(\theta_i, \theta_M)}{\partial \theta_M} \right)^{-1} > 0; \tag{74}$$

$$\frac{\partial \theta_M(\Psi, c_I, \theta_i)}{\partial \theta_i} = -\left(\frac{F(\theta_i) - F(\theta_M)}{1 - F(\theta_M)} \right) \left(\frac{\partial H(\theta_i, \theta_M)}{\partial \theta_M} \right)^{-1} > 0, \tag{75}$$

where the first term in (75) follows from differentiating the right-hand side of the third line in (34) with respect to θ_i and evaluating the resulting expression at $\theta^* = \theta_M$.[66] The expressions at right-hand sides of (73)-(75) are evaluated at the given Ψ, c_I, θ_i and the corresponding $\theta_M(\Psi, c_I, \theta_i)$ which solves the equation (72). Now assume that the prior density f is (strictly) log-concave and consider the following equivalent expression for $\frac{\partial \theta_M(\Psi, c_I, \theta_i)}{\partial \theta_i}$ obtained via differentiating (72) with respect to θ_i.

$$\frac{\partial \theta_M(\Psi, c_I, \theta_i)}{\partial \theta_i} = \frac{\frac{f(\theta_i)}{1 - F(\theta_M)} \left\{ \theta_i - \mathbb{E}\left[\theta_c \mid \cdot\right] \right\} + \frac{F(\theta_i) - F(\theta_M)}{1 - F(\theta_M)} \left(1 - \frac{\partial}{\partial \theta_i} \mathbb{E}\left[\theta_c \mid \cdot\right] \right)}{\frac{f(\theta_M)(1 - F(\theta_i))}{(1 - F(\theta_M))^2} \left\{ \theta_i - \mathbb{E}\left[\theta_c \mid \cdot\right] \right\} + \frac{F(\theta_i) - F(\theta_M)}{1 - F(\theta_M)} \frac{\partial}{\partial \theta_M} \mathbb{E}\left[\theta_c \mid \cdot\right]}. \tag{76}$$

---

[65]This fact is shown in Heckman and Honoré (1990), proposition 1.

[66]Differentiating the first term in (72) with respect to θ_i is equivalent to this, as the first term in (72) is simply a reexpression of (34).

Since f is (strictly) log-concave, its hazard rate is (strictly) increasing (An, 1998), i.e.

$$\frac{f(\theta_i)}{1 - F(\theta_i)} \geq (>)\frac{f(\theta_M)}{1 - F(\theta_M)}$$

$$\Longleftrightarrow \quad \frac{f(\theta_i)}{1 - F(\theta_M)} \geq (>)\frac{f(\theta_M)(1 - F(\theta_i))}{(1 - F(\theta_M))^2}, \tag{77}$$

because $\theta_i > \theta_M$. So the first term of the numerator in (76) is no smaller (strictly greater) than the first term of the denominator. Furthermore, due to log-concavity of f, for any arbitrary $\delta > 0$, the following ratio should be increasing in θ_c:

$$\frac{f(\theta_c)}{f(\theta_c + \delta)}, \tag{78}$$

whenever $\theta_c \in (\underline{\theta}, \overline{\theta})$.[67] As Shaked and Shantikumar (2007) show, this is equivalent to random-variable θ_c dominating the random variable $\theta_c - \delta$ in the *likelihood ratio order*, since $\delta > 0$ is arbitrary.[68] This implies that in any interval, the conditional expectation of θ_c should be no smaller than the conditional expectation of $\theta_c - \delta$, i.e.

$$\mathbb{E}\left[\theta_c - \delta \mid \theta_M \leq \theta_c - \delta < \theta_i\right] \leq \mathbb{E}\left[\theta_c \mid \theta_M \leq \theta_c < \theta_i\right]. \tag{79}$$

Taking δ out and rearranging yields the following.

$$\mathbb{E}\left[\theta_c \mid \theta_M + \delta \leq \theta_c < \theta_i + \delta\right] - \mathbb{E}\left[\theta_c \mid \theta_M \leq \theta_c < \theta_i\right] \leq \delta. \tag{80}$$

Dividing both sides by δ and taking the limit as $\delta \to 0$ gives:

$$\frac{\partial}{\partial \theta_M}\mathbb{E}\left[\theta_c \mid \cdot\right] + \frac{\partial}{\partial \theta_i}\mathbb{E}\left[\theta_c \mid \cdot\right] \leq 1$$

$$\Longleftrightarrow \quad 1 - \frac{\partial}{\partial \theta_i}\mathbb{E}\left[\theta_c \mid \cdot\right] \geq \frac{\partial}{\partial \theta_M}\mathbb{E}\left[\theta_c \mid \cdot\right]. \tag{81}$$

It follows that the second term of the numerator in (76) is no smaller than the second term of the denominator. Hence, one must have:

$$\frac{\partial\left(\theta_M(\Psi, c_I, \theta_i) - \theta_i\right)}{\partial \theta_i} \geq (>)0, \tag{82}$$

---

[67]See lemma 1 in An (1998).
[68]See Shaked and Shantikumar (2007), p. 66.

whenever f is (strictly) log-concave. Finally, differentiating (74) once more with respect to θ_M yields the following condition for concavity.

$$\frac{\partial^2 \theta_M(\Psi, c_I, \theta_i)}{\partial \Psi^2} = 2\left(\frac{c_I}{\Psi^3}\right)\left(\frac{\partial H(\theta_i, \theta_M)}{\partial \theta_M}\right)^{-1}$$
$$+ \left(\frac{c_I}{\Psi^2}\right)\left(\frac{\partial H(\theta_i, \theta_M)}{\partial \theta_M}\right)^{-2}\left(\frac{\partial \theta_M(\Psi, c_I, \theta_i)}{\partial \Psi}\right)\left(\frac{\partial^2 H(\theta_i, \theta_M)}{\partial \theta_M^2}\right) \leq 0.$$
(83)

The term in the first line is strictly negative and the first three terms in the second line are strictly positive. Keeping these in mind, plugging in for $\frac{\partial \theta_M(\Psi, c_I, \theta_i)}{\partial \Psi}$ from (74) and rearranging, the condition in (83) boils down to the following expression.

$$2\Psi\left(\frac{\partial H(\cdot)}{\partial \theta_M}\right)^2 \geq \frac{\partial^2 H(\cdot)}{\partial \theta_M^2}.$$
(84)

For this condition to hold for $\Psi > \Psi_0$, a sufficient condition is the non-positivity of the right-hand side, i.e. concavity of $H(\cdot)$ in θ_M. ∎

Lemma 11 (Comparative statics - Media and voter strategies): Take a payoff vector $(\theta_i, c_E, r, c_I, \Psi)$ corresponding to an equilibrium at a non-regime switching point and satisfying $\hat{\underline{\theta}} < \theta_i$, $\theta_i \in (\underline{\theta}, \overline{\theta})$, $c_E, c_I > 0$ and $\Psi \in (0, 1)$. Then there are several possible cases corresponding to different regimes. The cases are listed below.

1. **Case 1:** $R_0^* = I^* = 0$ with pure strategies: proposition 4 cases 1 and 2; proposition 5 cases 1 and 2. Since by assumption the equilibrium is not at a regime-switching point, there is some open neighbourhood around the given payoff vector where the equilibrium satisfies the same parametric condition specific to the particular regime at which the equilibrium is occurring and thus R_0^*, I^* are still zero. This constancy implies differentiability with derivatives being equal to zero.

2. **Case 2:** Passive media with voters doing the selection: proposition 4 case 3 and proposition 5 case 3. Then $I^* = 0$, it is differentiable and its derivatives are zero for the same reasons given in case 1. Furthermore, $R_0^* \in \left(0, 1 - \frac{c_E}{r}\right)$ (notice how the non-regime switching point assumption implies that $R_0^* \neq 1 - \frac{c_E}{r}$) and satisfies:

$$\theta_V(\theta_i) = \theta_i - r + (1 - R_0^*)^{-1}c_E,$$
(85)

on a sufficiently small neighbourhood around the given payoff vector. Solving this yields:

$$R_0^* = 1 - [\theta_V(\theta_i) - \theta_i + r]^{-1} c_E. \tag{86}$$

Differentiating with respect to payoff parameters yields the following:

$$\frac{\partial R_0^*}{\partial c_I} = 0; \tag{87}$$

$$\frac{\partial R_0^*}{\partial \Psi} = 0; \tag{88}$$

$$\frac{\partial R_0^*}{\partial r} = [\theta_V(\theta_i) - \theta_i + r]^{-2} c_E > 0; \tag{89}$$

$$\frac{\partial R_0^*}{\partial c_E} = -[\theta_V(\theta_i) - \theta_i + r]^{-1} < 0; \tag{90}$$

$$\frac{\partial R_0^*}{\partial \theta_i} = [\theta_V(\theta_i) - \theta_i + r]^{-2} c_E \frac{\partial (\theta_V(\theta_i) - \theta_i)}{\partial \theta_i} \gtrless 0, \tag{91}$$

where the sign of (90) follow from the fact that $\theta_V(\theta_i) - \theta_i + r > 0$. Furthermore, if f is (strictly) log-concave, then (91) is non-negative (strictly positive) from (71) in the proof of lemma 10.

3. **Case 3:** Active media with voters participating in selection: proposition 7 case 2. Then $I^* = 1$, it is differentiable and its derivatives are zero for the same reasons given in case 1. Furthermore, $R_0^* \in \left(0, 1 - \frac{c_E}{(1-\Psi)r}\right)$ (notice how the non-regime switching point assumption implies $R_0^* \neq 0$) and satisfies:

$$\theta_V(\theta_i) = \theta_i - r + [(1 - \Psi)(1 - R_0^*)]^{-1} c_E, \tag{92}$$

on a sufficiently small neighbourhood around the given payoff vector. Solving this yields:

$$R_0^* = 1 - (1 - \Psi)^{-1} [\theta_V(\theta_i) - \theta_i + r]^{-1} c_E. \tag{93}$$

Differentiating with respect to payoff parameters yields the following:

$$\frac{\partial R_0^*}{\partial c_I} = 0; \tag{94}$$

$$\frac{\partial R_0^*}{\partial \Psi} = -(1 - \Psi)^{-2} [\theta_V(\theta_i) - \theta_i + r]^{-1} c_E < 0; \tag{95}$$

$$\frac{\partial R_0^*}{\partial r} = (1 - \Psi)^{-1} [\theta_V(\theta_i) - \theta_i + r]^{-2} c_E > 0; \tag{96}$$

$$\frac{\partial R_0^*}{\partial c_E} = -(1 - \Psi)^{-1} \left[\theta_V(\theta_i) - \theta_i + r \right]^{-1} < 0; \tag{97}$$

$$\frac{\partial R_0^*}{\partial \theta_i} = (1 - \Psi)^{-1} \left[\theta_V(\theta_i) - \theta_i + r \right]^{-2} c_E \frac{\partial \left(\theta_V(\theta_i) - \theta_i \right)}{\partial \theta_i} \gtreqless 0, \tag{98}$$

where the signs in (95),(97) follow from the fact that $\theta_V(\theta_i) - \theta_i + r > 0$. Furthermore, if f is (strictly) log-concave, then (98) is non-negative (strictly positive) from (82) in the proof of lemma 10.

4. **Case 4:** Active media with uncertain voters strictly preferring the challenger: proposition 6 cases 1, 2, 3, and proposition 7 cases 1, 3 and 4. Then $R_0^* = 0$, it is differentiable and its derivatives are zero for the same reason given in case 1. If the equilibrium occurs at proposition 6 case 2 or proposition 7 case 3, where the media is (strictly) constrained by its strength, then $I^* = 1$ and the same kind of differentiability and zero-derivatives argument applies. If not, then I^* satisfies the following:

$$\theta_M(\Psi, c_I, \theta_i) = \theta_i - r + (1 - \Psi I^*)^{-1} c_E, \tag{99}$$

on a sufficiently small neighbourhood around the given payoff vector. Solving this yields:

$$I^* = \Psi^{-1} - \Psi^{-1} \left[\theta_M(\Psi, c_I, \theta_i) - \theta_i + r \right]^{-1} c_E. \tag{100}$$

Differentiating with respect to payoff parameters yields the following:

$$\frac{\partial I^*}{\partial c_I} = \Psi^{-1} \left[\theta_M(\cdot) - \theta_i + r \right]^{-2} c_E \frac{\partial \theta_M(\cdot)}{\partial c_I} < 0; \tag{101}$$

$$\frac{\partial I^*}{\partial \Psi} = -\Psi^{-1} I^* + \Psi^{-1} \frac{(1 - \Psi I^*)^2}{c_E} \frac{\partial \theta_M(\cdot)}{\partial \Psi} \gtreqless 0 \tag{102}$$

$$\frac{\partial I^*}{\partial r} = \Psi^{-1} \left[\theta_M(\cdot) - \theta_i + r \right]^{-2} c_E > 0; \tag{103}$$

$$\frac{\partial I^*}{\partial c_E} = -\Psi^{-1} \left[\theta_M(\cdot) - \theta_i + r \right]^{-1} c_E < 0; \tag{104}$$

$$\frac{\partial I^*}{\partial \theta_i} = \Psi^{-1} \left[\theta_M(\cdot) - \theta_i + r \right]^{-2} c_E \frac{\partial \left(\theta_M(\cdot) - \theta_i \right)}{\partial \theta_i} \gtreqless 0, \tag{105}$$

where the signs in (101) and (104) follow from the fact that $\theta_M(\cdot) - \theta_i + r > 0$. Furthermore, if f is (strictly) log-concave, then (105) is non-negative (strictly positive) from (82) in the proof of lemma 10. Finally, the expression in (102) follows from a straightforward (re)substitution for I^*. ∎

Proof of Lemma 12 (Payoff upper hemicontinuity of media and voters' equilibrium mixed strategy correspondences): First, define the following set.

$$\Upsilon^o = \left\{ (\theta_i, c_E, r, c_I, \Psi) : \theta_i \in \left(\underline{\theta}, \overline{\theta}\right), c_E > 0, r > 0, r > c_E, c_I > 0, \Psi \in (0,1) \right\}. \tag{106}$$

Let $\xi = (\theta_i, c_E, r, c_I, \Psi) \in \Upsilon^o$. For any $\xi \in \Upsilon^o$, the equilibrium challenge threshold is given by a unique $\theta^*(\xi)$ defined in (50). This completely pins down beliefs in a Bayesian manner from (25).[69] This and lemma 6 in turn imply that the equilibrium best-response correspondence of the media is given by the following.

$$BR_M(\xi) = \begin{cases} \{1\}, & (\Phi_M(\xi) > 0) \\ [0,1], & (\Phi_M(\xi) = 0) , \\ \{0\}, & (\Phi_M(\xi) < 0) \end{cases} \tag{107}$$

where $\Phi_M : \Upsilon \to \mathbb{R}$ is a function satisfying the following:

$$\Phi_M(\xi) = \theta^*(\Psi, c_I, \theta_i, c_E, r) - \tilde{\theta}_M(\Psi, c_I, \theta_i), \tag{108}$$

with $\tilde{\theta}_M(\Psi, c_I, \theta_i)$ as defined in (49) and $\theta^*(\Psi, c_I, \theta_i, c_E, r)$ as defined in (50). From the facts that both $\theta^*(\cdot)$ and $\tilde{\theta}_M(\cdot)$ are continuous, Φ_M is continuous in ξ. First, I will show that $BR_M(\xi)$ is UHC in ξ. Let BR_M^u denote the upper inverse (image) of BR_M, i.e.

$$BR_M^u(E) = \{\xi \in \Upsilon^o : BR_M(\xi) \subset E\}, \tag{109}$$

where E is some set in $\mathbb{R}$. $BR_M(\xi)$ is UHC if and only if BR_M^u maps open sets to open sets. Let E be an arbitrary open set. There are four cases to consider.

1. $E \subset [0,1]$: $BR_M^u(E) = \emptyset$, open.

2. $[0,1] \subset E$: $BR_M^u(E) = \Phi_M^{-1}(\mathbb{R}) \cap \Upsilon^o$, open because Φ_M is continuous so its inverse image maps open sets to open sets and Υ^o is open and intersections of open sets are open.

3. $\{0\} \subset E$ and $\{1\} \not\subset E$: $BR_M^u(E) = \Phi_M^{-1}(\mathbb{R}_{--}) \cap \Upsilon^o$, open due to same reason above.

---

[69]Note that I am limiting attention to the set of non-refinable equilibria, i.e. equilibria with challengers.

4. $\{0\} \not\subset E$ and $\{1\} \subset E$: $BR_M^u(E) = \Phi_M^{-1}(\mathbb{R}_{++}) \cap \Upsilon^o$, open due to same reason above.

It follows that $BR_M(\xi)$ is UHC on Υ^o. It is also UHC in Υ because one can "open" Υ by replacing domains for θ_i and Ψ in it with $(\underline{\theta} - \epsilon, \overline{\theta} + \epsilon)$ and $(-\epsilon, 1 + \epsilon)$ respectively and repeat the above argument. Now take some sequence $\xi_n \to \xi^*$ and let the associated sequence of equilibrium media strategies be denoted by $I_n^* \to I^*$.[70] I will show that I^* is an equilibrium strategy given ξ^*. Suppose not. Then $I^* \notin BR_M(\xi^*)$. By assumption, I_n^* is an equilibrium strategy for ξ_n for all n, i.e. $I_n^* \in BR_M(\xi_n)$ for all n. Furthermore, BR_M is UHC and its range ($[0, 1]$) can be contained in some compact set $A \subset \mathbb{R}$. So the closed-graph characterization of upper hemicontinuity implies $I^* \in BR_M(\xi)$, a contradiction. So I^* is an equilibrium strategy, which completes the proof. The proof of upper hemicontinuity for voters' equilibrium mixed strategy correspondence is analogous. ∎

Proof of Lemma 13 (Equilibrium response of media and voter strategies): Take a pair I^*, R_0^* of equilibrium strategies. The fact that in $\Upsilon \setminus \Upsilon_d$, R_0^* and I^* are non-decreasing in r and non-increasing in c_E follows from lemma 11 and corollary 4. Using the same two results, we can see that I^* is non-increasing in c_I, R_0^* is non-decreasing in it (in fact independent of it), and that R_0^* is non-increasing in Ψ, again in $\Upsilon \setminus \Upsilon_d$. Finally, pick a payoff vector $\xi \in \Upsilon_d$. If $\theta_M(\xi) = \theta_V(\xi) = \hat{\underline{\theta}}(\xi)$, then $I^* = R_0^* = 0$ (proposition 5, case 2), which is unique in the lemma 2 sense. Any marginal increase in c_I would result in $\theta_M(\cdot) < \theta_V(\xi) = \hat{\underline{\theta}}(\xi)$, which would again lead to proposition 5, case 2, so the equilibrium is unique in the lemma 2 sense.[71] Any marginal decrease in c_I would result in $\theta_M(\cdot) > \theta_V(\xi) = \hat{\underline{\theta}}(\xi)$ and yield a unique equilibrium under proposition 7 case 1. So from lemma 12, I^*, R_0^* are continuous in Ψ, c_I, and hence lemma 11 applies.[72] Finally, assume $\theta_M(\xi) = \theta_V(\xi) > \hat{\underline{\theta}}(\xi)$. Then from proposition 7.1, $I^* = I(\xi)$ and $R_0^* = R_0(\xi)$ take particular values to ensure that the following is satisfied.

$$\theta^* = \theta_M(\Psi, c_I, \theta_i) = \theta_V(\theta_i) = \theta_i - r + [(1 - \Psi I^*)(1 - R_0^*)]^{-1} c_E, \qquad (110)$$

with $I^* \in [0, 1]$, $R^* \in (0, 1)$. Now consider an arbitrarily small increase in c_I. Then from lemma 10, $\theta_V(\xi) > \theta_M(\xi) > \hat{\underline{\theta}}(\xi)$. So the equilibrium switches to case 3 in proposition 4 with $I^* = 0$ and R_0^* uniquely solving (38), which is just

---

[70]If I_n^* is not unique, then pick arbitrarily. Existence is guaranteed because it was shown.

[71]This follows from lemma 10.

[72]The reason for including this point to the set of discontinuities is that strategies are possibly discontinuous in r and c_E, which is why one cannot generalize lemma 11 for them over Υ_d.

(110) with $I^* = 0$. So R_0^* should not decrease, i.e. should jump up. Similarly, I^* should not increase, i.e. jump down. Considering an arbitrarily small decrease in c_I would lead to a mirror image of this, with a passage to an equilibria under proposition 7 case 1, 2, 3 or 4. If it is case 2 or 3, then one can see that I^* either jumps up to 1 (full investigation), or it was already equal to 1 so that it does not decrease, which translates into a non-increasing R_0^*. If it is case 1 or 4, then one can see that R_0^* either jumps down or stays constant, which implies an upward jump, i.e. a non-decreasing behaviour by I^*. Finally, the regime transition effect of an arbitrarily small increase in Ψ (when the payoff vector Ψ belongs to Υ_d) is identical to the regime switching effect caused by an arbitrarily small decrease in c_I. ■

Proof of Lemma 15 (Regularly non-monotonic strategic response to changes in media strength): The fact that $\theta_M(\cdot)$ and $\hat{\underline{\theta}}_\Psi$ intersect exactly twice if they intersect without being tangent follows immediately from strict concavity of $\theta_M(\cdot)$, strict convexity of $\hat{\underline{\theta}}_\Psi$, $\hat{\underline{\theta}}_\Psi > \theta_i$ for $\Psi > 1 - \frac{c_E}{r}$ and $\theta_M(\cdot) < \theta_i$ for all Ψ. The behaviour of I^* for $\Psi \in [0, \Psi_1]$ and $\Psi \in [\Psi_2, \Psi_3]$ are already shown. Consider the expression for the derivative of I^* with respect to Ψ whenever $\Psi \in (\Psi_1, \Psi_2) \cup (\Psi_3, 1)$:

$$\frac{\partial I^*}{\partial \Psi} = \Psi^{-1} \left[\frac{\partial \theta_M(\Psi)}{\partial \Psi} \frac{(1 - \Psi I^*)^2}{c_E} - I^* \right]. \tag{111}$$

Since I^* is continuous and is strictly smaller than 1 for $\Psi > \Psi_3$, this expression should be negative when one is arbitrarily close to Ψ_3. Indeed, continuity of I^* implies that $\lim_{\Psi \downarrow \Psi_3} I^* = 1$. Combining this with the continuous differentiability of $\theta_M(\Psi)$, one gets the following.

$$\lim_{\Psi \downarrow \Psi_3} \frac{\partial I^*}{\partial \Psi} = \Psi_3^{-1} \left[\left(\frac{\partial \theta_M(\Psi)}{\partial \Psi} \right)_{\Psi = \Psi_3} \frac{(1 - \Psi_3)^2}{c_E} - 1 \right] < 0. \tag{112}$$

The sign in (112) follows from a simple fact. Because $\theta_M(\Psi)$ is strictly concave in $[\Psi_0, 1]$ and $\hat{\underline{\theta}}_\Psi$ is strictly convex in $[0, 1]$; because they are both strictly increasing over these ranges; and because $\theta_M(\Psi_3) = \hat{\underline{\theta}}_{\Psi_3}$, $\theta_M(\cdot) > \hat{\underline{\theta}}_\Psi$ for $\Psi \in (\Psi_2, \Psi_3)$ and $\theta_M(\cdot) < \hat{\underline{\theta}}_\Psi$ for $\Psi \in (\Psi_3, 1]$; $\hat{\underline{\theta}}_\Psi$ should be steeper than $\theta_M(\Psi)$ at $\Psi = \Psi_3$. This can also be clearly seen in figure 6. This implies:

$$\left(\frac{\partial \theta_M(\Psi)}{\partial \Psi} \right)_{\Psi = \Psi_3} \left(\frac{\partial \hat{\underline{\theta}}_\Psi}{\partial \Psi} \right)_{\Psi = \Psi_3}^{-1} = \left(\frac{\partial \theta_M(\Psi)}{\partial \Psi} \right)_{\Psi = \Psi_3} \frac{(1 - \Psi_3)^2}{c_E} < 1. \tag{113}$$

From this, (112) immediately follows. Furthermore, since $\theta_M(\Psi)$ is continuously differentiable, so is I^* on $(\Psi_3, 1)$. It follows that there is some $\epsilon < 1 - \Psi_3$ such that $\frac{\partial I^*}{\partial \Psi} < 0$ when $\Psi \in (\Psi_3, \Psi_3 + \epsilon)$. Now suppose that I^* is increasing somewhere on $(\Psi_3, 1)$. Then its derivative should be strictly positive somewhere on this interval. Due to continuous differentiability of I^* on $(\Psi_3, 1)$, this implies the existence of some δ satisfying $\Psi_3 + \epsilon < \delta < 1$ where $\frac{\partial I^*}{\partial \Psi} = 0$ whenever $\Psi = \delta$ and $\frac{\partial I^*}{\partial \Psi} > 0$ for a Ψ slightly higher than δ. If this is the case, then (111) implies that at $\Psi = \delta$, the following should be satisfied.

$$\left(\frac{\partial \theta_M(\Psi)}{\partial \Psi} \right)_{\Psi = \delta} \frac{(1 - \delta I^*)^2}{c_E} = I^*. \tag{114}$$

Moreover, for an infinitesimally higher (than δ) strength, $\frac{\partial I^*}{\partial \Psi}$ is strictly positive, i.e.

$$\frac{\partial \theta_M(\Psi)}{\partial \Psi} \frac{(1 - \delta I^*)^2}{c_E} > I^*. \tag{115}$$

Because $\theta_M(\cdot)$ is strictly concave, and because ΨI^* is increasing everywhere due to (54), such an infinitesimal increase implies that the left-hand side of (114) becomes smaller than the right-hand side holding the right-hand side constant. So the only way for the inequality in (115) to hold is if the right-hand side falls even more. But this would imply that I^* is decreasing, which is a contradiction. Hence there is no such δ and I^* is (strictly) decreasing on $(\Psi_3, 1)$. The proof for (Ψ_1, Ψ_2) is similar. One first shows that:

$$\lim_{\Psi \uparrow \Psi_2} \frac{\partial I^*}{\partial \Psi} = \Psi_2^{-1} \left[\left(\frac{\partial \theta_M(\Psi)}{\partial \Psi} \right)_{\Psi = \Psi_2} \frac{(1 - \Psi_2)^2}{c_E} - 1 \right] > 0, \tag{116}$$

which follows from the fact that $\theta_M(\cdot)$ is steeper than $\underline{\hat{\theta}}_\Psi$ at $\Psi = \Psi_2$. Regarding (Ψ_1, Ψ_2), continuity of I^* implies that there exists some $\epsilon_1 \leq \Psi_2 - \Psi_1$ where I^* is increasing in $(\Psi_1, \Psi_1 + \epsilon_1)$, and some $\epsilon_2 \leq \Psi_2 - \Psi_1$, where I^* is increasing in $(\Psi_2 - \epsilon_2, \Psi_2)$. If these intervals are overlapping, then I^* is increasing everywhere on (Ψ_1, Ψ_2). Suppose not, and without loss of generality, suppose that there exists only a subinterval $(a, b) \subset (\Psi_1 + \epsilon_1, \Psi_2 - \epsilon_2)$ where I^* is decreasing. Then $\frac{\partial I^*}{\partial \Psi} = 0$ must be true at $\Psi = b$. But then, the previous argument shows that it should be decreasing for all $\Psi > b$, a contradiction. This also establishes that if I^* is increasing, then it is non-decreasing everywhere in the left (increasing everywhere in the left down to Ψ_0), and if it is decreasing, it is decreasing everywhere in the right. ∎

References

[1] **Alesina, A. and H. Rosenthal (2000)**, "Polarized policies and moderate policies with checks and balances", *Journal of Public Economics*, 75(1): 1-20.

[2] **An, Y. (1998)**, "Logconcavity versus logconvexity: a complete characterization", *Journal of Economic Theory*, 80(2): 350-369.

[3] **Ashworth, S. and E. Bueno de Mesquita (2008)**, "Electoral selection, strategic challenger entry and the incumbency advantage", *Journal of Politics*, 70(4): 1006-1025.

[4] **Banks, J. S. and D. R. Kiewiet (1989)**, "Explaining patterns of candidate competition in congressional elections", *American Journal of Political Science*, 33(4): 997-1015.

[5] **Banks, J. S. and R. K. Sundaram (1990)**, "Incumbents, challengers and bandits: Bayesian learning in a dynamic choice model", RCER Working Papers, 235, Univesity of Rochester, New York (US).

[6] **Banks, J. S. and R. K. Sundaram (1989)**, "Optimal retention in agency problems", *Journal of Economic Theory*, 82(2): 293-323.

[7] **Belzunce, F. and M. Shaked (2001)**, "Stochastic comparisons of mixtures of convexly ordered distributions with applications in reliability theory", *Statistics and Probability Letters*, 53: 363-372.

[8] **Bernhardt, D., S. Krasa and M. Polborn (2008)**, "Political polarization and the electoral effects of media bias", *Journal of Public Economics*, 92(5-6): 1092-1104.

[9] **Besley, T. (2004)**, "Playing politicians: theory and evidence", *Journal of the European Economic Association*, 2(2-3), 193-215.

[10] **Besley, T. (2005)**, "Political selection", *Journal of Economic Perspectives*, 19(3), 43-60.

[11] **Besley, T. (2006)**, *Principled Agents? The Political Economy of Good Government*, Oxford University Press, Oxford (UK).

[12] **Besley, T. and S. Coate (1997)**, "An economic model of representative democracy", *Quarterly Journal of Economics*, 112(1), 85-114.

[13] **Besley, T. and A. Prat (2006)**, "Handcuffs for the grabbing hand? Media capture and government accountability", *American Economic Review*, 96(3), 720-736.

[14] **Brennan, D. and J. Buchanan (1980)**, *The Power to Tax: Analytical Foundations of a Fiscal Constitution*, Cambridge University Press, Cambridge (UK).

[15] **Caselli, F. and M. Morelli (2004)**, "Bad politicians", *Journal of Public Economics*, 88(3-4), 759-782.

[16] **Coate, S. and S. Morris (1995)**, "On the form of transfers to special interests", *Journal of Political Economy*, 103(6), 1210-1235.

[17] **Congleton, R. D. and Y. Zhang (2013)**, "Is it all about competence? The human capital of U.S. presidents and economic performance", *Constitutional Political Economy*, 24(2), 108-124.

[18] **Cooter, R. (2003)**, "Who gets on top in democracy: elections as filters", *Supreme Court Economic Review*, 10, 127-141.

[19] **Corneo, G. (2006)**, "Media capture in a democracy: the role of wealth concentration", *Journal of Public Economics*, 90(1-2), 37-58.

[20] **Dewan, T. and R. Hortala-Vallve (2013)**, "Electoral competition, control and learning", working paper, London School of Economics, London (UK).

[21] **Duggan, J. and C. Martinelli (2011)**, "A spatial theory of media slant and voter choice", *Review of Economic Studies*, 78(2), 640-666.

[22] **Epstein, D. and P. Zemsky (1995)**, "Money talks: deterring quality challengers in congressional elections", *American Political Science Review*, 89(2), 295-308.

[23] **Estache, A. and R. Foucart (2013)**, "Benchmarking politicians", CEPR Discussion Papers, 9467, London (UK).

[24] **Fudenberg, D. and J. Tirole (1991)**, "Game Theory", MIT Press, Cambridge (US).

[25] **Gehlbach, S. and K. Sonin (2014)**, "Government control of the media", *Journal of Public Economics*, 118(1), 163-171.

[26] **Goodliffe, J. (2005)**, "When do war chests deter?", *Journal of Theoretical Politics*, 17(2), 249-227.

[27] **Gordon, S. C., G. A. Huber and D. Landa (2007)**, "Challenger entry and voter learning", *American Political Science Review*, 101(2), 303-320.

[28] **Heckman, J. T. and B. E. Honoré (1990)**, "The empirical content of the Roy model", *Econometrica*, 58(5), 1121-1149.

[29] **Llody, J. and L. Toogood (2014)**, *Journalism and PR - News Media and Public Relations in the Digital Age*, I.B. Tauris, London (UK).

[30] **Manelli, A. M. (1996)**, "The convergence of equilibrium strategies of approximate signalling games", *Economic Theory*, 7(2), 323-335.

[31] **Messner, M. and M. Polborn (2004)**, "Paying politicians", *Journal of Public Economics*, 88(12), 2423-2445.

[32] **Muñoz-García, F. and A. Espínola-Arredondo (2011)**, "The intuitive and divinity criterion: interpretation and step-by-step examples", *Journal of Industrial Organization Education*, 5(1), 1-20.

[33] **Osborne, M. J. and A. Slivinski (1996)**, "A model of political competition with citizen-candidates", *Quarterly Journal of Economics*, 111(1), 65-96.

[34] **Petrova, M. (2008)**, "Inequality and media capture", *Journal of Public Economics*, 92(1-2), 183-212.

[35] **Poutvaara, P. and T. Takalo (2007)**, "Candidate quality", *International Tax and Public Finance*, 14(1), 7-27.

[36] **Prat, A. and D. Strömberg (2005)**, "Commercial television and voter information", CEPR Discussion Papers, 4989, London (UK).

[37] **Prat, A. and D. Strömberg (2005)**, "The political economy of mass media", CEPR Discussion Papers, 8246, London (UK).

[38] **Riley, J. G. (1979)**, "Noncooperative equilibrium and market signalling", *American Economic Review*, 69(2), 303-307.

[39] **Shaked, M. and J. G. Shanthikumar (2007)**, *Stochastic Orders*, Springer, New York (US).

[40] **Spence, M. (1973)**, "Job market signalling", *Quarterly Journal of Economics*, 87(3), 355-374.

[41] **van Damme, E. (1987)**, "Equilibria in non-cooperative games", in: Peters, H. and O. Vrieze (eds.), *Surveys in Game Theory and Related Topics*, Stichting Mathematisch Centrum, Amsterdam (Netherlands).

[42] **Wrighton, J. M. and P. Squire (1997)**, "Uncontested seats and electoral competition for the U.S. House of Representatives over time", *Journal of Politics*, 59(2), 452-468.

Chapter 3: Electoral Campaigning with Correlated Ignorance

Abstract

This chapter aims to understand the effects of the diffusion of political knowledge in an electorate on politicians' campaign structures in democratic elections. The term campaign structure emphasizes a distinction between valence and policy focused campaigns. For this purpose, a two-candidate probabilistic voting model of costly policy and valence campaigning with ignorant (purely valence-driven) voters is developed. The model shows that resources devoted to valence campaigning increase with the fraction of ignorant voters and proximity of candidate legacies to the ideal policy of non-ignorant pseudo-swing voter.[1] The latter implies that when the state of being ignorant is correlated with policy preferences, and when politicians represent opposite segments of the policy spectrum at the beginning of the campaign, an otherwise symmetric setting can lead to campaign divergence after the correlation is taken into account. That is, depending on the context, right or left-wing politicians might be more inclined to engage in valence campaigning, solely due to political awareness varying monotonically across the electorate.

1 Introduction

1.1 Motivation and Summary

Electoral campaigns are often more complex than envisaged by spatial models of policy announcement in the Downsian tradition. Besides manipulative techniques such as priming or framing, what Riker (1983) called heresthetics, they include at least two forms of campaign activities differing in their focus: policy and valence. Other than issue positioning, first one involves efforts of persuasion and commitment via detailed outlines, proposals and rhetorical engagements as emphasized

---

[1] Pseudo-swing voter refers to the decisive voter under costless policy announcement and no campaigning. Once campaigning is introduced, she is no longer swing voter but her position constitutes an important reference point.

by Hinich and Munger (1993). Valence is usually employed as a catch-all term encompassing non-policy factors that may influence voting decisions. Valence campaigning can include emphases of universally (or near-universally) desirable prospects, e.g. prosperity, justice, inclusive growth, moderate religiosity in some societies etc., or it can involve accentuation of candidates' positive character aspects and virtues, e.g. integrity, competence, leadership quality etc., what Stokes (1992) calls first and second elements of valence politics respectively.[2,3] It also comprises activities of image promotion, impressionistic advertising and even negative campaigning, also sometimes referred colloquially as mudslinging.[4,5]

There are several empirical studies on campaign structures. Lau and Pomper (2002) find that incumbents resort less to negative campaigning compared to challengers in U.S. senatorial elections between 1992-2002. With the same dataset, Brueckner and Lee (2013) show that centrist candidates tend to go more negative. In a recent study based on 2008 U.S. congressional elections, Gschwend et al. (2014) show that candidates who have a policy advantage conduct campaigns with a heavier policy focus and candidates with a valence advantage tend to do the opposite. Curini (2014), using a dataset covering 60 years in 37 countries, shows that the ideological distance of a candidate to its adjacent competitors has a negative impact on its tendency to emphasize issues related to character valence, particularly corruption and honesty. One factor that can influence the tone of a campaign, as well as lead different candidates to diverge in terms of campaign structures is informational status, or political awareness of an electorate. So far,

---

[2]Usage of the term valence is far from being uniform in the literature, mostly owing to a lack of consensus on what constitutes policy and non-policy issues. Besides indicating a differentiation between what Berelson et al. (1954) call "style issues" in contrast to "position issues", it is generally used in a fashion implying a directional uniformity regarding the desirability of a valence issue. However, it is also used in contexts where the issue at hand can generate positive, as well as negative support in different parts of the electorate.

[3]Related to the second footnote, political issues often have both spatial and valence aspects, e.g. immigration. For this chapter, valence dimension would be best thought as consisting of issues orthogonal to the policy dimension.

[4]There are several studies showing that looks of candidates are relevant in campaign designs and can matter for electoral outcomes. For instance, Atkinson et al. (2009) find that in U.S. Senate elections, parties tend to select "higher quality challenger faces" in more competitive districts. Todorov et al. (2005) conducted an experiment asking naive students to rank two candidates in terms of "facial competence" basing on a short exposure to their photographs. They found that students' responses can predict the outcomes of U.S. congressional elections with 70% accuracy.

[5]Negative campaigns are often conducted via ads that either directly attack one's opponent, or contrast him/her with the attacker. It can be seen as a form of valence campaigning involving significant risks of backlash as demonstrated by Dole & Hagan affair in 2008 U.S. Senate elections in North Carolina. It should also be kept in mind that some authors argue (e.g. West (1993)) that voters can define some campaigns as being negative simply due to their dislike towards them.

this remains unexplored empirically and received less attention theoretically but there are legitimate reasons to expect such effect. Since the seminal study of Converse (1964) interpreting results from American National Election Studies panel surveys covering the decade of 50's, pervasive voter cluelessness is an acknowledged phenomenon. Going beyond lack of information, Converse (1964) showed a lack of both temporal and ideological incoherence characterizing the answers of a significant majority regarding positional issues, which he interpreted as a major disconnect between mindsets of political elites and "men in the street". Converse (1964) also showed that this lack of political sophistication correlates negatively with levels of education and political knowledge.[6] Returning to lack of information, Delli Carpini and Keeter (1993) cite 90-91 Michigan NES survey, revealing that 55% of the voters could not correctly identify parties' positions regarding federal spending, and 53% of them did not know which party held the majority in the Senate. Delli Carpini and Keeter (1996) show that this lack of information is negatively correlated with socioeconomic variables such as income and education. A frequently cited TÜSİAD (2001) study reports strong positive correlations between various measures of political information and socioeconomic status for Turkish voters.

Why might voters' levels of political knowledge or degrees of awareness influence the way candidates conduct their campaigns? Clarke et al. (2009) provide some evidence from UK for leader effects, and more generally character valence mattering more for politically less sophisticated voters. Leiter (2013), analyzing individual level data from Germany, Britain and Netherlands, shows that politically sophisticated voters care more about proximity in positional issues, but both types (sophisticated and unsophisticated) of voters care equally about valence issues.[7] Likability heuristics, as suggested by Brady and Sniderman (1985), or usage of leader images as cues for policy, as put forth by Clarke et al. (2004) are among some of the potential explanations proposed for existence of valence voters and their seeming prevalence among the less informed. Popkin (1994), and later Lupia and McCubins (1998) have argued that voters can use a variety of low-information heuristics to approximate an informationally complete rational

---

[6]Converse's (1964) analysis revealed the existence of some voters who hold positions on a variety of issues, albeit inconsistent from a unidimensional liberal-conservative framework. This contributed to later alternative models of voting, such as multidimensional voting à la Plott (1967) or average "consideration ownership" model of Zaller (1992). Nevertheless, it also revealed that a significant portion of voters seem to vote without any policy or issue considerations (i.e. noise voters).

[7]Both studies measure political sophistication using an interaction term that consists of formal education and political knowledge.

decision. Yet, there is also evidence for less knowledgeable voters being moved by superficial differences in politicians or basing their votes on frivolous candidate traits. Lau and Redlawsk (2006) provide experimental evidence from mock elections on how subjects with less interest and knowledge on an issue can show support for (physically) more attractive versus less attractive candidates, despite the less attractive candidate representing a position closer to their preferences. Lenz and Lawson (2011), using individual-level voting data from 2006 US gubernatorial and senatorial elections, find that poorly informed voters who also watch TV disproportionately cast their votes on the basis of candidates' physical appearances. Negiz and Akyıldız (2012), using survey data from 2009 Turkish local elections, provide evidence for voters having low levels of formal education attaching greater importance to physical characteristics in candidates compared to relatively better educated voters. These patterns suggest that politicians running for office and their campaign teams have sufficient reason to structure their campaigns by taking not only the electorate's policy preferences, but also their audience's cognitive/informational limitations into account.

In this chapter, I build a model of electoral campaigning bringing these considerations together. The model is a variation on Baron's (1994) and later Grossman and Helpman's (1996) noise voters framework, with elements of costly policy and valence campaigning and correlated ignorance.[8] The electorate is divided into two types of voters. Informed or politically aware voters who can observe the messages generated by campaigns in the policy dimension, and ignorant voters who can't condition their voting decisions on policies.[9] All voters get affected by valence campaigning, which is assumed to operate by inducing an additively separable partisanship effect.[10] In addition, this informational status is assumed to depend monotonically on policy preferences in a probabilistic sense. I first specify a simple probabilistic voting model with two office-motivated politicians and a continuum of voters indexed according to their preferred policies, incorporating this idea of correlated informational segregation, on top of which I introduce

---

[8] I use the term "correlated" figuratively, referring to a particular form of stochastically monotonic association defined in assumption 1 later on, which also implies correlation literally.

[9] This sort of ignorant voter behaviour can be justified with an argument invoking incomplete information (being unable to observe policies) and Knightian uncertainty regarding initial positions of the politicians. Alternatively, one can go with a bounded rationality argument involving voter sophistication. In any case, voters are algorithmic players in my model as I focus on the way politicians structure their campaigns around policy and valence elements, by assuming that campaigns do work as intended for whatever reason. There is a separate literature of microfounded campaigns dealing with the question of why do campaigns work, which is briefly reviewed in the next subsection.

[10] This is a convenience assumption. Assuming that only ignorant voters are influenced by valence campaigning would only strengthen the conclusions of this chapter.

elements associated with campaigning. Politicians are endowed with limited resources, which they can allocate between policy and valence campaigning. While valence campaigning works as mentioned above, policy campaigning allows politicians to commit on policies differing from their pre-campaign positions, which can be interpreted as their reputations or legacies they inherited from their party affiliations.

The model is partly inspired by an observational curiosity characterizing a large part of Turkish electoral politics in the last decade which left many observers puzzled. AKP, the heir of "Islamic left", frequently associated with a conservative, rural and relatively uneducated voter base, was a newcomer back in 2002 Turkish elections. Unlike its main opponent CHP which is often associated with better educated and relatively liberal urban voters and whose campaign was heavily based on valence issues, AKP conducted a campaign with an almost-exclusive policy focus and positioned itself to a much more liberal stance than one would expect by looking at preferences of its voter base. The "puzzle" is somehow related to the characteristic image of Turkish rural voters being susceptible to valence-related communal appeals popularized in literature and cinema, as well as a common preconception that parties campaign in their own backyards. While the latter preconception was more or less valid for the two decades preceding 2002, AKP's campaign was hugely successful partly because it broke this cycle.[11] In fact, if the characteristic image contains a grain of truth and CHP and AKP's voter bases were indeed more policy and valence sensitive respectively in a statistical sense, then they would have incentives to behave in a manner compatible with the observation, as already being the boss of your own house gives all the reason for campaigning at each other's backyard. This chapter essentially provides a simple model capturing this idea using a correlated ignorance argument.

The chapter has two central results. First one is a reiteration of already known results with a slight twist and comes from the benchmark probabilistic voting model without campaigning. If whether voters are ignorant or not is independent of their policy preferences, then existence of ignorant voters is irrelevant and candidates converge to the policy dictated by the centripetal force generated by the entirety of the electorate. This is basically Wittman's (1989)

---

[11]The cycle was a partial byproduct of a military coup in 1982 which resulted in a continuous army tutelage for almost 20 years, as well as a heavy "political landscape engineering" by military cadres. See Zürcher (2004) chapter 15 for a brief overview of Turkish politics after 80's and Hale and Ozbudun (2009) for a detailed narrative of the period following AKP's emergence, as well as a comparison of its election campaigns with those of its opponents'.

argument, later called "miracle of aggregation" by Brennan (2012). On the other hand, no matter how large the fraction of ignorant voters is, as long as there is some systematic relation between policy preferences and informational status, ignorant voters' preferences are completely ignored. This is in accord with Baron (1994) and Grossman & Helpman's (1996) insights derived from models with finite numbers of voters. The twist is: when this relation is monotonic, i.e. when ignorance and preferences are correlated, the electoral outcome has a left or right bias, depending on the direction of the correlation.[12] Second, and the main result is obtained when campaigning is introduced. Assuming that candidates are positioned at the opposing segments of the policy spectrum at the beginning of the campaign, a monotonic association between policy preferences and informational status *ceteris paribus* introduces asymmetries into their campaign strategies. That is, assuming a positive (negative) correlation between ignorance and policy preferences, candidates competing in an otherwise symmetric environment would differ in their campaign focuses with the left (right) candidate conducting a more valence focused campaign and vice versa. At the heart of the result lies the role played by the effective proximity of politicians' initial positions (legacies) to pseudo-swing voter's position. The latter represents the most beneficial position in the policy dimension in terms of victory probabilities assuming no valence campaigning. A lower distance between the two incentivizes a candidate to follow a more valence oriented campaign due to decreasing marginal rate of substitution between different campaign types. There are several additional results such as the no-longer-irrelevance of the fraction of ignorant voters under costly campaigning, which can be read in the concluding section.

1.2 Previous Literature and Contribution

As mentioned in the previous section, one strand of campaigns literature is involved in microfounding them. It evolved as a substrand of the literature aiming to characterize conditions under which information transmission (from politicians to voters) occurs in elections, which in turn got its kickstart with Banks' (1990) seminal paper. Banks' (1990) model is not really a model of campaigning, but a simple signalling extension of the standard Downsian framework, where candidates have preferred positions unknown to voters and announcing a policy that differs from this position is costly for the winner for some reason. He shows that

---

[12]This result is reminiscent of Bartels (2008) who is a strong proponent of the view that a large impoverished part of the population is not politically represented due to lack of political knowledge.

information transmission, i.e. separating equilibria, occurs only if deviation costs are sufficiently high and politicians are extreme enough in their initial positions. Prat (2002) is among the first papers attempting to microfound electoral campaigns. He considers a median-dominated Downsian setting with two politicians who (besides policy) compete in quality, true value of which is neither known to voters, nor to a lobbying group who doesn't care about quality but has its own policy preference. Both the lobby and voters get noisy signals about the quality gap between candidates, with former's signal being private and latter's signal being public. The lobby picks a candidate to whom it offers contributions, which pays off if the candidate wins and which the candidate receives if he wins by announcing a position matching lobby's preference. Candidates can use contributions in advertisement, which generates no direct information or preference manipulation, but indirectly provides voters with the information that the candidate is being financed by the lobby. Since the lobby wants its candidate to win, and since it uses its private signal to forecast the subsequent public signal, this otherwise useless advertisement activity signals a higher candidate quality to voters. Comparing the value of this information to the welfare loss stemming from policy distortion towards lobby's preference allows Prat (2002) to analyze welfare effects of banning campaign contributions. Coate builds two models where campaigns play a directly informative role and are funded by a lobby as well. In Coate (2004a), he assumes that candidates differ in an unobservable quality dimension but unlike Prat (2002), he considers advertisement as conveying truthful information to the electorate, i.e. good quality candidates can advertise and reveal themselves to some voters with the fraction increasing in advertisement expenditure, but bad quality candidates cannot. By allowing candidates to bargain policy positions with the lobby in return for contributions, he can study welfare effects of introducing contribution limits, which turn out to be positive. In Coate (2004b), there are prospective candidates with fixed but privately known policy positions. Candidates are chosen by their parties at the beginning of elections according to their likelihood of winning. There are two lobbies representing left and right segments of the policy spectrum and campaign contributions can be used for truthfully conveying candidate positions in a manner similar to Coate (2004a). The dual-partisan nature of lobbying, as well as the lack of an electorate-wide desirable attribute such as quality reverses his previous normative conclusions, and he shows that limiting campaign contributions can be Pareto worsening. Gul and Pesendorfer (2012) propose an alternative model of directly informative campaigning. There are two candidates, one decisive voter

and a state of nature: the identity of the candidate holding the correct (beneficial for the voter) position. Information is symmetric and no one knows this state. Campaigning conveys truthful information to everyone in the form of a continuous stochastic process pointing out towards the correct candidate as long as the campaigner keeps incurring its cost, which is interpreted as candidate's fund raising ability. In addition, only the underdog (in terms of voter support, which depends on public beliefs) can campaign. They show that if an intervention equalizes the playing ground by evening out the cost gap between candidates' campaign technologies, then it is welfare improving due to strategic substitutability of campaign activities. From this brief summary, it should be apparent that one big advantage of microfounding campaigns as informative activities is that it allows for a satisfactory normative analysis of policy interventions by considering informational costs. Naturally, the model presented here will be lacking on that front. However, since the main goal of the chapter is to lay down the foundations of a framework where asymmetries in the way candidates structure their campaigns can be tracked down to campaign responsiveness of their audience, this is not too big of a loss. Relatedly, there are no contributors but politicians are endowed with exogenously given campaign resources in my setting.

Since Stokes' (1963) seminal critique of Downs (1957), there is an ongoing effort of coming up with theories of electoral competition taking valence considerations into account but models incorporating policy and valence dimensions into a unified setting appeared relatively recently. Although they usually differ in their focus and scope, they can be roughly thought as belonging to a common strand of unmicrofounded and persuasive campaigns.[13,14] This chapter can be loosely placed within this literature. A standard result from the spatial voting literature is the mean voter theorem, emphasizing the policy location of the swing voter when candidates compete in a probabilistic framework and voters have quadratic utilities. One of the first papers from the persuasive campaigns literature is by Schofield (2003), who shows that when candidate valence is generated by activist coalitions, and when the valence these coalitions generate can be influenced by politicians via policy accommodation, the mean value theorem no longer holds. Wiseman (2006) presents a model that is slightly closer to the one in this chapter. His setting features two candidates moving sequentially (an incumbent and a challenger), each offering platforms that consist of a costless pol-

---

[13]For the sake of brevity, this strand will henceforth be referred as persuasive campaigns literature.

[14]Interestingly and unusually, this literature did not appear before but arose simultaneously with the microfounded campaigns literature.

icy position and an amount of campaign spending which translates linearly into valence support in the voting stage. Politicians have intrinsic policy preferences and are constrained with limited resources but since campaign resources have no alternative use, they allocate all of their budget to obtain valence support, so valence is in effect exogenously determined. His focus is rather on positional deviations caused by the presence of valence campaigning and the influence exercised by valence budgets. His general result is that incumbent's policy stance depends on challenger's valence budget and if the challenger is relatively poor in terms of resources, then the incumbent can get away with a less moderate policy position. Continuing with the theme of valence-policy interaction, Carrillo and Castanheira (2008) consider a model of sequential campaigning where two candidates with intrinsic policy preferences (left and right) first announce one of three possible policies: L, M, R with M coinciding to median voter's position and left candidate can only announce L, M and vice versa. They then make a costly valence investment which is interpreted as a competence demonstration. They assume that a fraction of the electorate does not observe the valence dimension, and find out that if this fraction is too high or too low, convergence to median prevails. For an intermediate range of uninformedness, policy divergence remains a possibility. Ashworth and Bueno de Mesquita (2009), maintaining policy first, valence second timing, specify a probabilistic voting model with stochastic preferences where purely office motivated candidates are uncertain of voters' bliss points. They show that even though candidates have no intrinsic policy preferences, they announce polarized platforms in equilibrium in order to soften the subsequent valence competition. Zakharov (2009), again within a policy first, valence second framework, focuses on the effects of exogenous partisanship. His model has two parties, each having its own partisan base consisting of voters who either vote for the party with which they are affiliated, or abstain. Parties can engage in costly valence spending to swing non-partisan voters after announcing their policy positions.[15] He shows that an increase in partisan voters makes candidates engage more intensively in valence campaigning, as well as making them strategically diverge in policy positions to reduce costs associated with acquiring valence. Serra (2010) reverses the timing and assumes that candidates first invest in valence, then announce policies. She shows that in a non-probabilistic framework, valence spending is again associated with platform polarization. Moving on to a different theme, Meirowitz (2007) starts with a pure valence campaign

---

[15]Herrera, Levine and Martinelli (2008) consider a framework similar to Zakharov's (2008) with partisan voters but assume that campaigning can probabilistically stop voters from abstaining.

setting, modelling valence competition as an all pay auction with effort costs being analogous to the price paid by bidders, and electoral victory representing the auctioned good. By considering mixed as well as pure strategies, he shows that the candidate with an electoral advantage (which can be interpreted as an incumbency advantage) exerts less effort and the candidate with a cost advantage exerts more effort for valence campaigning. His results imply that while a policy aiming to increase funding costs could be opposed by both parties, a policy introducing a spending cap would always make the electorally disadvantaged party better off. Morton and Myerson (2012) integrate lobbying and persuasive campaigning within a probabilistic voting framework. They assume that candidates first announce policy platforms, then competitively raise campaign funds which they then spend to acquire valence. Interest groups' contributions depend not only on the policy commitments in the first stage, but also on probabilities of an election victory. While their setting generates a multiplicity of equilibria, they show that the contribution market can negate the natural uncertainty associated with probabilistic voting, ensuring the victory of one of the candidates. In all the models mentioned so far in this paragraph, policy competition occurs via costless announcements. Eyster and Kittsteiner (2007) specify a model without valence elements where two parties choose a policy platform at the beginning of the election, and then conduct costly policy campaigns in separate constituencies with different median voters in order to relocate from their initial platforms. They show that in equilibrium, parties position themselves asymmetrically to right and left and proceed to carving out their own "home turfs" within neighbourhoods of their announced positions. Unlike the models presented so far, campaign technology in this chapter contains both policy and valence elements. Relocating is costly as in Eyster and Kittsteiner (2007), albeit candidates' initial positions are exogenously given. Valence is costly as well, as in Zakharov (2009) or Ashworth and Bueno de Mesquita (2009), and candidates are endowed with campaign budgets as in Wiseman (2006), which they can allocate between these two types of campaign activities.

There are several more papers that should be mentioned here, as they share some common elements with the model presented in the next section. Baron (1994), Grossman and Helpman (1996) and Strömberg (2004) are among important papers which model uninformed voters as noise voters who don't respond to policy proposals. The first two assume that uninformed voters are "impressionable" in the sense that they can be swinged by campaign spending as in my model. The difference is, they assume that this valence-susceptibility is exclusive

to uninformed voters. Their goal is to provide insight on how the policy making process is captured by interest groups who provide politicians with means of capturing uninformed masses. Strömberg (2004) uses the same framework to study the effects of mass media on policy competition in elections. His model features no valence campaigning but rather a media industry which determines who gets the news, i.e. who becomes a noise voter and who responds to policy. He shows that if the media industry is characterized by large fixed costs and increasing returns, then large groups (of voters having the same policy preferences) are more likely to end up informed, which implies that equilibrium platforms are biased towards their preferences. Regarding policy campaigning, the main inspiration for the specification in this chapter came from Harrington and Hess (1996) who use a similar costly relocation model in the context of negative campaigning. Their model features no valence campaigning but candidates, besides relocating themselves by spending campaign resources, are also able to shift each other back towards initial positions by undertaking costly and persuasive extremism allegations. There are also two other papers which attempt to introduce correlation between policy preferences and ignorance like I do in this chapter. Making a detour to the lobbying theme, Bardhan and Mookherjee (2000) study incentives governing interest groups regarding local versus national capture in a multi-district probabilistic voting model with purely office motivated politicians and noise voters. Their model features an electorate divided into three classes: rich, middle class and poor with a fraction of rich (and informed) voters exogenously organized into a lobbying group. Interpreting voters who respond to policy competition as politically aware voters, they assume that the rich group have the highest number of politically aware voters and vice versa. They show that high inequality districts are more prone to interest group capture due to a lower fraction of politically aware voters, which results campaign contributions being more valuable at local level then at national level. Finally, Lind and Rohner (2013) builds a Wittmanian probabilistic voting model featuring noise voters and politicians with intrinsic policy preferences, which makes them subject to a utility cost when committing to a different platform.[16] Their motivation is to build a lobbying-free model that can explain the "rich bias" in redistributive policies across U.S. states. Their specification do not feature valence campaigning but intrinsic policy preferences combined with deviation costs have effects similar to policy campaigning with costly relocation employed in this chapter. Their conclusions regarding policy platforms are similar: introduction of correlation leads

---

[16]See Wittman (1983).

one candidate becoming more moderating, and the other becoming more extreme. Besides the difference in focus and lack of campaigning, the way they introduce this correlation is different. They apply a constant perturbation to the conditional probability function at a cutoff policy preference and apply variational calculus to study its effects. My model instead uses stochastic order relations, which also allows me to study the effects of strengthening the correlation.

Contributions of this chapter are twofold. First, it provides a reasonably simple way of introducing correlated ignorance in a probabilistic voting model, which allows to not only study the effects of a correlation between policy preferences and ignorance in isolation from the fraction of the latter, but also to assess the impact of strengthening this correlation. Although the electorate is assumed to be a continuum in this chapter, the method is equally applicable when the policy conflict is characterized by a finite or a countably infinite number of policy positions. Second, it offers a novel potential explanation for asymmetries in politicians' campaign structures without exclusively relying on candidate-specific characteristics and instead arguing that correlated political awareness, interacting with candidate legacies can *ceteris paribus* provide sufficiently strong incentives to candidates for differentiating their campaign tones. This explanation can possibly offer guidance for future empirical work.

Next section presents the model. It starts by specifying a simple probabilistic voting model with noise voters, then introduces correlation between ignorance and policy preferences. This is followed by the introduction of campaigning, which is embedded on top of this benchmark probabilistic voting model. One shortcoming of the campaign model with costly policy relocation is that it doesn't allow for cheap (free) announcement of policies. Before concluding, I attempt to address this in an alternative model where candidates first costlessly announce policies, then allocate their resources between policy and valence campaigning, where policy campaigns serve to reduce the post election ambiguity faced by policy aware (and risk averse) voters. Final section concludes.

2   Model

The main model consists of two modifications to a standard probabilistic voting framework with two politicians and a continuum of voters indexed according to their preferred policies who rank different policies using a utility difference and additive bias specification. First, electoral competition involves two dimensions: valence and policy. Engaging in both types of competition requires spending

resources, with which politicians are endowed in different amounts. Second, a fraction of the voters are assumed to be noise voters, in the sense that their voting decisions are based purely on the valence dimension. These voters are deemed ignorant because assuming that they care about policy, they should either be unable to observe the policy competition, or they should have cognitive limitations preventing them from assessing policy proposals. Being ignorant and having a specific policy preference are not *ex ante* independent events, but they are monotonically related in a probabilistic sense. I build the model step by step, starting with a probabilistic voting model with ignorant voters, no valence campaigning and costless policy commitment. This not only constitutes a benchmark for the full model but also provides a point for assessing the effects of correlated ignorance on campaign behaviour later on.

2.1 Probabilistic Voting with Correlated Ignorance

There are two purely office motivated politicians labelled A and B. They can costlessly announce and commit to policies $X_A, X_B \in [0, 1]$. It is assumed that they maximize the probability of electoral victory, although one can equivalently assume that they maximize expected vote shares.[17] There is a continuum of sincere voters with measure 1, who are indexed according to their policy preferences $x \sim [0, 1]$ distributed according to an absolutely continuous distribution F. For a given policy X, the reduced-form policy preference of a voter with bliss point x is captured by the following utility function.

$$U(X; x) = -\mathcal{D}(|X - x|), \tag{1}$$

where $\mathcal{D} : \mathbb{R}_+ \to \mathbb{R}_+$ represents voter's disutility from facing a policy different from her bliss point, which is assumed to depend symmetrically on the distance between actual and ideal policies. It is further assumed that $\mathcal{D}$ is strictly increasing, strictly convex and sufficiently smooth so that $\mathcal{D} \circ |\cdot|$ is twice continuously differentiable.[18] While strict convexity ensures the existence of a unique pure strategy equilibrium, symmetry allows to get a clean peak at the effects of correlated ignorance on campaigning by construction of a symmetric equilibrium.[19]

---

[17]This is a trivial consequence of both expected vote share and probability of victory maximizations yielding the same best responses. Patty (2002) shows the equivalence of two objectives in two candidate elections without abstention or coordinated voting.

[18]This specification, first used by Davis et al. (1970), is quite common in positive political theory literature.

[19]The conclusions derived from the probabilistic voting model in this section is robust under any smooth (twice continuously differentiable) voter utility satisfying single-crossing in policy-

Voters are either informed or ignorant. Informed voters are assumed to observe a perfectly informative private signal revealing policy announcements. In addition, they care about the identity of the politician announcing the policy in a specific way. Letting $I = 1$ denote the state of being informed, the pre-election utility of an informed voter from a candidate i victory through announcement of policy X is given by the following.

$$W(X, i; x, I = 1) = \pm\gamma_I^x + U(X; x). \tag{2}$$

In (2), γ_I^x represents a bias towards the candidate A (so $\pm = +$ if $i = A$). This is the widely used additive bias specification, first popularized by Enelow and Hinich (1982). Under *ex ante* politician uncertainty on the bias, it gives rise to the standard utility-difference probabilistic voting model *à la* Lindbeck and Weibull (1987). The bias consists of two components. An individual ideological component σ_I^x, which can *ex ante* be seen as an idiosyncratic preference shock that is independent across voters, and a common valence component δ, which can be seen as a systemic preference shock that affects all voters in the same way.

$$\gamma_I^x = \sigma_I^x + \delta, \tag{3}$$

where σ_I^x and δ are distributed uniformly over supports $[-\beta_I, \beta_I]$ and $[-\alpha, \alpha]$ respectively. Uniformly distributed biases is a widely made assumption in probabilistic voting models. Under more general voter utilities[20], it ensures the global concavity of politician payoffs. Here, it simplifies the exposition and the results would hold if one assumed a symmetric distribution for the idiosyncratic shock and log-concave distribution for the common shock. Their realizations are not observed by politicians at the time of policy announcements.

Ignorant voters neither observe policy announcements, nor receive any policy signal that can be conditioned upon. The pre-election utility of an ignorant voter from a candidate i victory is given by the following.

$$W(X, i; x, I = 0) = \pm\gamma_U^x. \tag{4}$$

This approach for modelling uninformed voting behaviour is quite standard and goes back to Baron (1994). It is later used by Grossman and Helpman (1996) in the context of lobbying, and by Strömberg (2004) in the context of mass

---

type pair when voters are indexed in some unidimensional type-space instead of being indexed according to their bliss points. See appendix.

[20]Even more general than single-crossing.

media's effect on electoral competition. It can be justified in two ways. First, it can be taken as an *ad hoc* bounded rationality specification. Alternatively, it can be assumed that upon not-observing the policy announcement, rational but ignorant voters hold common beliefs over symmetric policy announcements across politicians.[21] The bias γ_U^x is specified as before.

$$\gamma_U = \sigma_U^x + \delta, \tag{5}$$

where σ_U^x is distributed uniformly over support $[-\beta_U, \beta_U]$. Notice that the possibility of differing (non-policy) ideological heterogeneity between two types of voters ($\beta_I \neq \beta_U$) is left open. At worst one can assume $\beta_I = \beta_U$. If not, $\beta_I \neq \beta_U$ can be taken as an approximation to existence of socioeconomic groups governing informational status, e.g. urban dwellers versus rural inhabitants. It is assumed that informational status and policy preferences are dependent, and the dependence is summarized by the following conditional mass function.

$$\Gamma(x) = \mathbb{P}(I = 1 | x). \tag{6}$$

The timing of events is as follows. [1] Both politicians, knowing the distributions of policy preferences, σ_I^x, σ_U^x, δ and $\Gamma(x)$, simultaneously and noncooperatively announce their electoral platforms: $X_A, X_B \in [0, 1]$. [2] Nature chooses who is ignorant and who is not, idiosyncratic and common shocks realize. [3] Voters vote according to (2) and (4). [4] Winner implements the policy.

Given two policy proposals X_A, X_B, an informed voter with a policy preference x strictly prefers voting for A if the following condition holds.

$$\sigma_I^x - \mathcal{D}(|X_A - x|) + \delta > -\mathcal{D}(|X_B - x|)$$
$$\implies \sigma_I^x > \mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|) - \delta, \tag{7}$$

i.e. only if her bias towards candidate A is strong enough to compensate for the extra disutility she incurs by voting for A instead of B. Given any realization of the common shock δ, it follows that the probability she votes for the candidate A is given by the following.

---

[21]This makes sure that conditional on beliefs, expected policy-utility differences for uninformed voters from voting on A or B are zero. In fact, since politicians are identical besides their label, and since uninformed voters' policy preferences will not be taken into account by politicians under any belief system, beliefs that put all the mass on the policy to which politicians converge would constitute a Perfect Bayesian Equilibrium to a game with voters as Bayesian actors. Keeping this in mind, as typical in probabilistic voting literature, I take voter decisions axiomatically and focus on the game between politicians.

$$\mathbb{P}\left(\sigma_I^x > \mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|) - \delta\right)$$

$$= \int_{\mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|) - \delta}^{\beta_I} \frac{1}{2\beta_I} \mathrm{d}\sigma_I^x$$

$$= \frac{1}{2} + \frac{1}{2\beta_I} \left\{ \mathcal{D}(|X_B - x|) - \mathcal{D}(|X_A - x|) + \delta \right\}, \tag{8}$$

where the second line follows from the fact that σ_I^x is uniformly distributed. Similarly, from (4), for a given realization of δ, an ignorant voter with policy preference x votes for candidate A if the following condition is satisfied.

$$\sigma_U^x > -\delta. \tag{9}$$

It follows that the probability she votes for A is given by the following.

$$\mathbb{P}\left(\sigma_U^x > -\delta\right) = \frac{1}{2} + \frac{1}{2\beta_U}\delta. \tag{10}$$

Using the fact that population is of measure 1, for a given realization of δ and a given pair of electoral platforms (X_A, X_B), the expected measure of informed voters voting for candidate A is given by the following.

$$S_A^I(\delta) = \int_0^1 \Gamma(x) \mathbb{P}\left(\sigma_I^x > \mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|) - \delta\right) \mathrm{d}F(x). \tag{11}$$

Similarly, the expected measure of uninformed voters preferring candidate A is given below.

$$S_A^U(\delta) = \int_0^1 (1 - \Gamma(x)) \mathbb{P}\left(\sigma_U^x > -\delta\right) \mathrm{d}F(x). \tag{12}$$

To begin with, assume that informational status and policy preferences are independent. Then $\Gamma(x) = \hat{\Gamma}$ for some $\hat{\Gamma} \in [0, 1]$ for all x. So (11) becomes:

$$S_A^I(\delta) = \hat{\Gamma} \int_0^1 \mathbb{P}\left(\sigma_I^x > \mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|) - \delta\right) \mathrm{d}F(x)$$

$$= \frac{\hat{\Gamma}}{2} + \frac{\hat{\Gamma}}{2\beta_I} \left\{ \int_0^1 \left[\mathcal{D}(|X_B - x|) - \mathcal{D}(|X_A - x|)\right] \mathrm{d}F(x) + \delta \right\}. \tag{13}$$

Similarly, (12) becomes:

$$S_A^U(\delta) = (1 - \hat{\Gamma}) \int_0^1 \mathbb{P}\left(\sigma_U^x > \delta\right) \mathrm{d}F(x)$$

$$= \frac{(1 - \hat{\Gamma})}{2} + \frac{(1 - \hat{\Gamma})}{2\beta_U}\delta. \tag{14}$$

So the expected total measure of voters preferring candidate A is given by:

$$S_A(\delta) = S_A^I(\delta) + S_A^U(\delta)$$

$$= \frac{1}{2} + \left(\frac{\hat{\Gamma}}{2\beta_I} + \frac{(1 - \hat{\Gamma})}{2\beta_U} \right)\delta + \frac{\hat{\Gamma}}{2\beta_I} \int_0^1 \left[\mathcal{D}(|X_B - x|) - \mathcal{D}(|X_A - x|) \right] \mathrm{d}F(x). \tag{15}$$

Notice that from law of large numbers, (15) is also the actual share of votes that the candidate A would receive under the realized δ and given policy platforms. An electoral victory requires at least half of the votes, i.e.

$$S_A(\delta) \geq \frac{1}{2}$$

$$\iff \delta \geq \tilde{\Gamma}(\hat{\Gamma}, \beta_I, \beta_U) \int_0^1 \left[\mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|) \right] \mathrm{d}F(x), \tag{16}$$

where $\tilde{\Gamma}(\cdot)$ denotes the effective policy weight of informed voters taking into account relative voter heterogeneity, i.e.

$$\tilde{\Gamma}(\hat{\Gamma}, \beta_I, \beta_U) = \frac{\frac{\hat{\Gamma}}{2\beta_I}}{\frac{\hat{\Gamma}}{2\beta_I} + \frac{(1 - \hat{\Gamma})}{2\beta_U}} = \frac{1}{1 + \frac{\beta_I}{\beta_U}\frac{1 - \hat{\Gamma}}{\hat{\Gamma}}} \equiv \tilde{\Gamma}(\hat{\Gamma}, \frac{\beta_I}{\beta_U}). \tag{17}$$

If $\beta_I = \beta_U$, then $\tilde{\Gamma}(\cdot)$ is simply the share of informed voters in the electorate. The more ideologically heterogenous informed voters are relative to ignorant voters, the more noisy their policy preferences become in the eyes of politicians, which reduces their effective weight in policy proposals. From (16) and due to the fact that δ is distributed uniformly in $[-\alpha, \alpha]$, the probability of an electoral victory for candidate A is given by the following for a fixed policy announcement pair (X_A, X_B).

$$\mathbb{P}(S_A(\delta) > \frac{1}{2}) = \frac{1}{2} + \frac{1}{2\alpha} \left\{ \tilde{\Gamma}(\hat{\Gamma}, \frac{\beta_I}{\beta_U}) \int_0^1 \left[\mathcal{D}(|X_B - x|) - \mathcal{D}(|X_A - x|) \right] \mathrm{d}F(x) \right\}. \tag{18}$$

(18) is the objective function for candidate A, which he maximizes with respect to X_A given X_B. Analogously, the objective function for candidate B is given by the following.

$$\mathbb{P}(S_B(\delta) > \frac{1}{2}) = \frac{1}{2} + \frac{1}{2\alpha}\left\{\tilde{\Gamma}(\hat{\Gamma}, \frac{\beta_I}{\beta_U})\int_0^1 \left[\mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|)\right] \mathrm{d}F(x)\right\}. \tag{19}$$

A Nash equilibrium of the electoral game without campaigning will be a pair of policy platforms (X_A^*, X_B^*) such that:

$$X_A^* \in \arg\max_{X_A \in [0,1]} \mathbb{P}(S_A(\delta; X_A, X_B^*) > \frac{1}{2}),$$

$$X_B^* \in \arg\max_{X_B \in [0,1]} \mathbb{P}(S_B(\delta; X_A^*, X_B) > \frac{1}{2}). \tag{20}$$

Strict convexity of the voter disutility ensures that the electoral game has a unique pure-strategy Nash equilibrium.

Proposition 1 (Electoral game - Existence and uniqueness of pure-strategy Nash equilibrium): The electoral game with costless policy commitment and no valence campaigning has a unique pure-strategy Nash equilibrium.

Proof: Strategy spaces, being $[0,1]$, are convex and compact. From continuity of $\mathcal{D}(\cdot)$, $\mathbb{P}(S_k(\delta; X_k, X_j) > \frac{1}{2})$ is continuous in X_k for any X_j and vice versa for $k, j \in \{A, B\}$ and $k \neq j$. Since $\mathcal{D}(\cdot)$ is strictly increasing and strictly convex, $-\mathcal{D}(\cdot)$ is strictly decreasing and strictly concave. Furthermore; $|X - x|$ is convex in X for any x. It follows that $-\mathcal{D}(|X - x|)$ is strictly concave in X for any x. Integral of (strictly) concave functions is (strictly) concave, so $\mathbb{P}(S_k(\delta; X_k, X_j) > \frac{1}{2})$ is strictly concave in X_k for any X_j for $k \in \{A, B\}$ and $k \neq j$. Thus, from Debreu-Glicksberg-Fan theorem, the electoral game has a unique pure-strategy Nash equilibrium, characterized by a pair of policy platforms (X_A^*, X_B^*) satisfying (20). ∎

Focusing on the problem of candidate A (candidate B's problem and first-order condition is symmetric), the first-order sufficient condition for an interior maximum is given by the following.[22]

$$-\frac{\tilde{\Gamma}(\hat{\Gamma}, \frac{\beta_I}{\beta_U})}{2\alpha}\frac{\partial}{\partial X_A}\left(\int_0^1 \mathcal{D}(|X_A - x|)\mathrm{d}F(x)\right) = 0$$

$$\Longleftrightarrow -\frac{\partial}{\partial X_A}\left(\int_0^{X_A} \mathcal{D}(X_A - x)\mathrm{d}F(x) + \int_{X_A}^1 \mathcal{D}(x - X_A)\mathrm{d}F(x)\right) = 0$$

$$\Longleftrightarrow -\int_0^{X_A} \mathcal{D}'(X_A - x)\mathrm{d}F(x) + \int_{X_A}^1 \mathcal{D}'(x - X_A)\mathrm{d}F(x) = 0, \tag{21}$$

---

[22]Global concavity of the objective function ensures the sufficiency of the first-order condition.

where the third line follows due to $\mathcal{D}$ being twice continuously differentiable, F being absolutely continuous and $\left[\mathcal{D}'(|X_A - x|)\right]_{x=X_A} = 0$.[23] At the optimal platform, the marginal probability gain from offering a policy slightly more aligned with preferences of the voters to the right should be exactly compensated by the marginal probability loss from that policy being slightly more distant to bliss points of the voters to the left. Two facts stand out from (21). First, from the symmetry of first-order conditions, policy convergence occurs, i.e. both politicians commit to same platforms. This is a natural consequence of politicians having no intrinsic policy preferences or any constraints preventing them from freely relocating across the policy space. Second, neither existence of ignorant voters, nor any increase in the share of ignorant voters have any effect on equilibrium platforms. This is in stark contrast with Lind and Rohner (2013), who, assuming intrinsic policy preferences for politicians, show that a decrease in the fraction of ignorant voters leads to at least one candidate choosing a less polarized platform. Their result stems from the fact that when candidates have preferred policies, there is a tension between "centrifugal" (own policy) and "centripetal" (electorate's preferences) forces.[24] An increase in the fraction of ignorant voters favours centrifugal forces, allowing politicians to win elections by proposing platforms closer to their own bliss points. In the current model, the hegemony belongs to the centripetal force, on which the presence of ignorant voters have no effect, unless their policy preferences systematically differ from the preferences of informed voters. The lack of association between informational status and policy preferences implies that politicians would have no reason to assume that informed voters, who can be swinged by offering different policies, will have different preferences compared to the general population. From (21), one can immediately see that there will be a unique interior electoral platform that will satisfy the first-order condition.

Lemma 1 (Electoral game - Unique interior equilibrium policy): Assuming full support for F, in equilibrium, there exists a unique $X^* \in (0, 1)$ announced by both politicians.

Proof: Policy convergence follows from the equivalence of first-order conditions under the symmetric payoffs given in (20). $\mathcal{D}$ is strictly increasing so at $X_A = 0$, (21) is strictly positive and at $X_A = 1$, it is strictly negative. Since $\mathcal{D}$ is con-

---

[23]These ensure that the derivative can be passed under the integral. The last one implies that boundary variations from Leibniz rule get cancelled out. Under a more general voter utility, the same cancellation would occur from the envelope theorem.

[24]This terminology is first employed by Cox (1990).

tinuously differentiable, the left-hand side of (21) is continuous in X_A, so from intermediate value theorem, there is some $X^* \in (0, 1)$ such that (21) holds when $X_A = X^*$. $\mathcal{D}$ is strictly convex, so the left-hand side of (21) is strictly decreasing in X_A, which implies that this X^* is unique. ∎

There is also a special case under which the position of X^* can exactly be located.

Lemma 2 (Electoral game - Equilibrium policy with symmetric distribution): Assume that F has a symmetric density f. Then $X^* = \frac{1}{2}$.

Proof: Start by rewriting the first-order condition.

$$-\int_0^X \mathcal{D}'(X - x)\mathrm{d}F(x) + \int_X^1 \mathcal{D}'(x - X)\mathrm{d}F(x) = 0. \tag{22}$$

Symmetry implies $F(\frac{1}{2}) = \frac{1}{2}$ and $f(\frac{1}{2} + k) = f(\frac{1}{2} - k)$ for all $k \in \left(0, \frac{1}{2}\right)$. So let $X = \frac{1}{2}$ and pick some $k \in \left(0, \frac{1}{2}\right)$. Consider voters with bliss points $\frac{1}{2} - k$ and $\frac{1}{2} + k$. Their marginal disutilities are both $\mathcal{D}'(k)$ and they get exactly the same weight due to symmetry of f. So the marginal probability loss from the former is exactly compensated by the marginal probability gain from the latter. Varying k from 0 to $\frac{1}{2}$ completes the argument. Uniqueness follows from lemma 1. ∎

Suppose now that informational status and policy preferences are *ex ante* dependent in a probabilistic sense, i.e. (6) is no longer constant but varies with x. Politicians would then make use of the statistical information this fact provides when announcing their electoral platforms. So (11) becomes as follows.

$$\begin{aligned} S_A^I(\delta) &= \int_0^1 \Gamma(x)\mathbb{P}\left(\sigma_I^x > \mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|) - \delta\right)\mathrm{d}F(x) \\ &= \Gamma \int_0^1 \mathbb{P}\left(\sigma_I^x > \mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|) - \delta\right)\mathrm{d}F(x|I = 1) \\ &= \frac{\Gamma}{2} + \frac{\Gamma}{2\beta_I}\left\{\int_0^1 \left[\mathcal{D}(|X_B - x|) - \mathcal{D}(|X_A - x|)\right]\mathrm{d}F(x|I = 1) + \delta\right\}, \end{aligned} \tag{23}$$

where the second line follows from a straightforward application of the Bayes' rule and Γ is the unconditional probability of a voter receiving the perfectly informative signal, or equivalently, the total measure of informed voters as the population is of measure 1, i.e.

$$\Gamma = \mathbb{E}\left[\Gamma(x)\right] = \mathbb{P}(I = 1). \tag{24}$$

Since correlation or not, politicians have no means of capturing the support of

ignorant voters, the expression for the measure of ignorant voters who vote for A given in (14) remains the same with $\hat{\Gamma}$ replaced by $\Gamma = \mathbb{E}\left[\Gamma(x)\right]$. So given a pair of electoral platforms, probabilities of an electoral victory are given by the following.

$$\mathbb{P}(S_A(\delta) > \frac{1}{2}) = \frac{1}{2} + \frac{1}{2\alpha}\left\{\tilde{\Gamma}(\Gamma, \frac{\beta_I}{\beta_U})\int_0^1\left[\mathcal{D}(|X_B - x|) - \mathcal{D}(|X_A - x|)\right]\mathrm{d}F(x|I=1)\right\},$$

$$\mathbb{P}(S_B(\delta) > \frac{1}{2}) = \frac{1}{2} + \frac{1}{2\alpha}\left\{\tilde{\Gamma}(\Gamma, \frac{\beta_I}{\beta_U})\int_0^1\left[\mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|)\right]\mathrm{d}F(x|I=1)\right\}.$$

$$(25)$$

Both proposition (1) and lemma 1 are still valid, and politicians will announce some unique platform $X_A = X_B = X^* \in (0,1)$ as long as $f(x|I=1)$ does not put all the mass to 0 or 1. This policy should solve the following simplified first-order condition.

$$-\int_0^X \mathcal{D}'(X - x)\mathrm{d}F(x|I=1) + \int_X^1 \mathcal{D}'(x - X)\mathrm{d}F(x|I=1) = 0. \qquad (26)$$

So in a probabilistic voting model with costless policy commitment, presence of noise voters affect equilibrium platforms only if preferences of the latter differs systematically from those of the informed voters because, as can be seen more clearly from (26), politicians only care about policy preferences of informed voters. Hence, the closer the association is between policy preferences and informational status, the stronger will be the divergence between the equilibrium policy and preferences of ignorant voters. This also implies that unlike the previous case, there will be a divergence between the political outcome and a planner outcome assuming that the social planner treats all agents in an egalitarian manner. To see this, just notice that informational status is irrelevant for the planner. Hence, while the solution to planner's problem would satisfy (22), the political outcome would solve (26).

How will the equilibrium policy differ under correlated ignorance? To determine the direction of bias, one needs to posit some form of monotonicity governing the relation between informational status and policy preferences. In the context of a unidimensional policy space, there are a variety of reasons why informational status can be monotonically associated with policy preferences. One reason might be cognitive: a certain level of human capital accumulation may be required to map electoral discourses onto pledged policies and this can be related to policy preferences. For instance, the policy variable can be thought as representing a

redistributive policy instrument (e.g. tax level or public good with a balanced budget), preferences toward which can be related to gross incomes, which can in turn be associated with educational attainment. Another reason might be informational in the literary sense: A certain time-commitment to follow the media might be required to accurately learn about policy positions and the extent with which voters can access media resources may depend on factors such as wealth or geography, which can in turn induce a somehow monotonic association between informational status and policy preferences. For example, the policy variable can be representing some metric of religious intensity in education, which can be more preferable for rural conservative voters, who might have more limited access to informational resources. Without loss of generality, suppose that informational status and policy preferences have a probabilistically decreasing association in the following sense.

Assumption 1 (Likelihood ratio (LR) dominance): $F(x|I = 1)$ is LR-dominated by $F(x|I = 0)$, i.e. $\frac{f(x|I=0)}{f(x|I=1)}$ is increasing in x.

Assumption 1 says that ignorant voters are more likely to have higher policy preferences compared to ignorant voters.[25] From law of total probability, $F(x)$ is a weighted average of these two conditional distributions so it also implies that $F(x|I = 1)$ is LR-dominated by $F(x)$, i.e. $\frac{f(x)}{f(x|I=1)}$ is increasing in x. So compared to $F(x)$, $F(x|I = 1)$ signifies a left shift in the mass of its density, reflecting the assumption that compared to general population, informed voters are more likely to prefer lower policies. In fact, since informational status is a binary random variable, assumption 1 is equivalent to monotonicity of $\Gamma(x)$.

Lemma 3 (Monotonicity of $\Gamma(x)$ and LR-dominance): $F(x|I = 1)$ being LR-dominated by $F(x|I = 0)$ is equivalent to $\Gamma'(x) = \frac{\partial \mathbb{P}(I=1|x)}{\partial x} \leq 0$, assuming that $\Gamma(\cdot)$ is differentiable.

Proof: Appendix. ∎

At this point, it should be noted that assumption 1 is stronger than necessary, because $U = -\mathcal{D}$ is supermodular in (X, x). In the appendix, I show that the monotonicity described in proposition 3 below prevails under first-order stochastic dominance order as long as voter utilities are supermodular. The reason for imposing the stronger assumption is that it allows for a much more intuitive exposition. This is because one implication of assumption 1 is that not only $F(x|I = 1)$

---

[25]Such as higher taxes, higher provision of religious education, etc. The opposite assumption would yield the opposite conclusions.

is first-order stochastically dominated by $F(x)$, but that this dominance survives on any measurable set $B \subseteq [0, 1]$, i.e. $F(x|x \in B) \leq F(x|x \in B, I = 1)$ for all $x \in B$. In fact, to get the intuitive exposition in proposition 2, it would be enough to assume the slightly weaker condition that $F(x|I = 1)$ is *dual*-hazard rate dominated by $F(x|I = 0)$, that is, both hazard-rate and reverse-hazard rate dominated. This would provide just enough truncation-proofness (the requirement is that first-order stochastic dominance survives only on half truncations) and is implied by but does not imply LR-dominance.[26] While assumption 1 means that a voter with a high policy preference always has a lower chance of receiving the policy signal compared to a voter with a lower policy preference; both of the weaker assumptions would allow for the possibility of *ex ante* political-savviness among voters who prefer higher policies. It should also be mentioned that under assumption 1, lemma 3 implies that the state of being informed and policy preferences are negatively correlated in the literary sense.

Corollary 1 (Negative correlation between informational status and policy preferences): Under assumption 1, the random variables x and I are negatively correlated, i.e. $\text{cov}(x, I) \leq 0$.[27]

Proof: Appendix. ∎

To determine the direction of policy bias, I start by rewriting the first-order condition (22) in a modified but equivalent form.

$$-F(X)\mathbb{E}\left[\mathcal{D}'(X - x)|x \leq X\right] + (1 - F(X))\mathbb{E}\left[\mathcal{D}'(x - X)|x > X\right] = 0, \quad (27)$$

where $\mathbb{E}$ denotes the expectation operator. Doing the same for (26) gives the following.

$$
\begin{aligned}
& -F(X|I = 1)\mathbb{E}\left[\mathcal{D}'(X - x)|x \leq X, I = 1\right] \\
& + (1 - F(X|I = 1))\mathbb{E}\left[\mathcal{D}'(x - X)|x > X, I = 1\right] = 0.
\end{aligned}
\quad (28)
$$

Next proposition shows that the solution of (28) should be non-greater than the solution of (27).

---

[26] See chapter 1 in Shaked and Shantikumar (2007). Dual-hazard rate dominance can be thought as the two-sided version of the concept of one-sided conditional stochastic dominance used by Maskin and Riley (2000). In addition, if the voter disutility in (1) is quadratic, a simple negative correlation between I and x would be sufficient.

[27] Note that negative correlation does not imply assumption 1. In fact, it doesn't even imply first-order stochastic dominance. Correlation is a much weaker measure capturing only linear association. See Yi and Tongyu (2004).

Proposition 2 (Monotonicity of equilibrium electoral platforms): Let X^* denote the solution to (27) and X_c^* denote the solution to (28), then under assumption 1, $X_c^* \leq X^*$, with the inequality being strict if the LR-dominance is strict.

Proof: Suppose assumption 1 holds, i.e. $\frac{f(x|I=0)}{f(x|I=1)}$ is increasing in x. Then due to the following, $F(x|I=1)$ is LR-dominated by $F(x)$.

$$
\begin{aligned}
\frac{f(x)}{f(x|I=1)} &= \frac{\mathbb{P}(I=1)f(x|I=1) + \mathbb{P}(I=0)f(x|I=0)}{f(x|I=1)} \\
&= \mathbb{P}(I=1) + \mathbb{P}(I=0)\frac{f(x|I=0)}{f(x|I=1)}.
\end{aligned}
\tag{29}
$$

$F(x|I=1)$ being LR-dominated by $F(x)$ implies that it is also first-order stochastically dominated by it. So for any X, $F(X) \leq F(X|I=1)$ and thus $(1-F(X)) \geq (1-F(X|I=1))$. Furthermore, from truncation-proofness of the LR order, both $F(x|x \leq X, I=1)$ and $F(x|x > X, I=1)$ are LR-dominated (and thus first-order stochastically dominated) by $F(x|x \leq X)$ and $F(x|x > X)$ respectively. Since $\mathcal{D}'(\cdot)$ is strictly increasing due to strict convexity, it follows that $\mathcal{D}'(X-x)$ is strictly decreasing in x for $x \in [0, X)$ and $\mathcal{D}'(x-X)$ is strictly increasing in x for $x \in (X, 1]$. This implies that for any given X, following inequalities hold as well.

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{D}'(X-x)|x \leq X, I=1\right] &\geq \mathbb{E}\left[\mathcal{D}'(X-x)|x \leq X\right], \\
\mathbb{E}\left[\mathcal{D}'(x-X)|x > X, I=1\right] &\leq \mathbb{E}\left[\mathcal{D}'(x-X)|x > X\right].
\end{aligned}
$$

These imply that (28), and thus (26) evaluated at X^* does not hold with equality but rather with $\leq$, with the inequality being strict if one of the above inequalities is strict. Moreover, the left-hand side of (26) is strictly decreasing in X due to strict convexity of $\mathcal{D}$. It follows that $X_c^* \leq X^*$. ∎

The intuition behind proposition 2 is straightforward. At any optimal interior policy, marginal probability loss for electoral victory due to positioning further away from voters to the left should be exactly compensated by the marginal probability gain resulting from offering a policy slightly closer to bliss points of those to the right. The negative association between being informed and policy preferences not only implies that voters to the left are weighted more heavily in their contribution to these marginal probabilities compared to the case with no association, but also that voters who matter more in the margin, i.e. the "tail" voters who are furthest away from the proposed policy are more heavily

represented at the left portion of the distribution. As a result, politicians take this left shift in the center of mass of the centripetal force into account and commit to lower platforms. Below, I provide a simple example with linear disutilities.[28]

Example: Suppose $\mathcal{D}(\cdot)$ is linear, i.e.

$$\mathcal{D}(|X - x|) = |X - x|. \tag{30}$$

Under no association between informational status and policy preferences, politicians' simplified first-order conditions become the following.

$$-\frac{\partial}{\partial X}\left\{\int_0^1 |X - x|\, \mathrm{d}F(x)\right\} = 0$$

$$\iff -\frac{\partial}{\partial X}\left\{\int_0^X (X - x)\, \mathrm{d}F(x) + \int_X^1 (x - X)\, \mathrm{d}F(x)\right\} = 0. \tag{31}$$

Using the Leibniz rule, (31) yields the following.

$$-F(X) + \big[F(1) - F(X)\big] = 0. \tag{32}$$

This is the standard median voter result under probabilistic voting with Euclidean preferences.

$$F(X^*) = \frac{1}{2}. \tag{33}$$

Under dependence, the same condition becomes:

$$F(X_c^*|I = 1) = \frac{1}{2}. \tag{34}$$

So instead of the median of the entire distribution, the informed median holds all the strings. Furthermore, under assumption 1, $F(X^*|I = 1) \geq F(X^*) = 1/2$. Since $F(\cdot|I = 1)$ is increasing, it follows that $X_c^* \leq X^*$. $\square$

Proposition 2 shows that introducing a positive association between policy preferences and ignorance leads to a left bias in equilibrium policy proposals.[29]

---

[28]Note that with linear disutilities, $\mathcal{D} \circ |\cdot|$ is no longer differentiable at 0. Nevertheless, since the set of non-differentiable points is of measure zero, they can be integrated away so the first-order conditions and previous conclusions still hold with the exception of boundary variations as shown in the example.

[29]Left is used in a purely spatial sense in the context of the current model. If the policy variable were to represent tax rate, right would be more reality-accommodating.

A related consideration would be to assess the impact of the strength of this dependence. What is the effect of a stronger positive association between policy preferences and ignorance on equilibrium platforms? To shed light on this, suppose there is some parameter $\theta \in [0, \infty)$ that measures the strength of dependence by indexing conditional distributions in the following sense.

Assumption 2 (Monotone likelihood ratio (MLR) ordering): The family of absolutely continuous and smooth distributions $\mathcal{F} = \big\{ F(x|I = 1, \theta) : \theta \in [0, \infty) \big\}$ satisfies the following.

$$\frac{\mathrm{d}}{\mathrm{d}x} \left[\frac{f_\theta(x|I = 1, \theta)}{f(x|I = 1, \theta)} \right] \leq 0, \qquad \forall F(\cdot|I = 1, \theta) \in \mathcal{F}, \tag{35}$$

with $F(x|I = 1, 0) = F(x)$, where f_θ denotes the partial derivative of f.

Assumption 2 is a continuous version of the likelihood-ratio ordering.[30] It amounts to saying that for any $\theta' > \theta$ and $x' > x$, the following condition holds.

$$\frac{f(x'|I = 1, \theta)}{f(x'|I = 1, \theta')} \geq \frac{f(x|I = 1, \theta)}{f(x|I = 1, \theta')}$$
$$\iff \frac{f(x|I = 1, \theta')}{f(x'|I = 1, \theta')} \geq \frac{f(x|I = 1, \theta)}{f(x'|I = 1, \theta)}. \tag{36}$$

In short, distribution of preferences conditional on having received the policy signal puts relatively more weight on lower policies under a stronger dependence between informational status and policy preferences.

Lemma 4 (Stronger correlation): Assume that the population distribution of preferences $F(x)$ and the total measure (unconditional probability) of ignorant voters remain fixed. Let θ be as described in assumption 2 and suppose $\theta' > \theta$. Then $\mathrm{cov}_{\theta'}(x, I) \leq \mathrm{cov}_\theta(x, I) \leq 0$.

Proof: Appendix. ∎

From the result in proposition 2, one can guess that a stronger association between policy preferences and informational status in the sense of assumption 2 would lead politicians to propose lower policies in equilibrium. This is indeed the case.

Proposition 3 (MLR-monotonicity of equilibrium platforms): Let $X^*(\theta)$ denote the equilibrium policy proposal by candidates where θ denotes the strength of negative association between being informed and policy preferences of voters.

---

[30]See Milgrom (1981).

Then under assumption 2, $X^*(\theta)$ is decreasing, i.e. $\frac{\partial X^*(\theta)}{\partial \theta} \leq 0$.

Proof: Fix a θ and consider the simplified first-order condition to the policy proposal problem.

$$-\int_0^X \mathcal{D}'(X-x)\mathrm{d}F(x|I=1,\theta) + \int_X^1 \mathcal{D}'(x-X)\mathrm{d}F(x|I=1,\theta) = 0. \quad (37)$$

Let $Z(X,\theta)$ denote the left-hand side of (37). Since $\mathcal{D}$ is twice continuously differentiable and $F(\cdot|\cdot)$ is smooth, the implicit function theorem implies that (37) defines X as a function $X \equiv X^*(\theta)$ on an open neighbourhood around the given θ and that the following condition holds.

$$\frac{\mathrm{d}X^*(\theta)}{\mathrm{d}\theta} = -\frac{\partial Z(X,\theta)/\partial\theta}{\partial Z(X,\theta)/\partial X}. \quad (38)$$

From strict convexity of $\mathcal{D}$, we have $\partial Z(X,\theta)/\partial X < 0$. So $\mathrm{d}X^*(\theta)/\mathrm{d}\theta$ will have the same sign as $\partial Z(X,\theta)/\partial\theta$. Using the fact that $\mathrm{d}F(x|I=1,\theta) = f(x|I=1,\theta)\mathrm{d}x$; this is in turn given by the following.

$$\frac{\partial Z(X,\theta)}{\partial\theta} = -\int_0^X \mathcal{D}'(X-x)f_\theta(x|I=1,\theta)\mathrm{d}x$$
$$+ \int_X^1 \mathcal{D}'(x-X)f_\theta(x|I=1,\theta)\mathrm{d}x. \quad (39)$$

Again using the same fact, (39) can be rewritten as follows.

$$\frac{\partial Z(X,\theta)}{\partial\theta} = -\int_0^X \mathcal{D}'(X-x)\frac{f_\theta(x|I=1,\theta)}{f(x|I=1,\theta)}\mathrm{d}F(x|I=1,\theta)$$
$$+ \int_X^1 \mathcal{D}'(x-X)\frac{f_\theta(x|I=1,\theta)}{f(x|I=1,\theta)}\mathrm{d}F(x|I=1,\theta). \quad (40)$$

From (35) and (37), the first term in (40) must be non-smaller than the second term in absolute value, which implies $\partial Z(X,\theta)/\partial\theta \leq 0$, and thus $\mathrm{d}X^*(\theta)/\mathrm{d}\theta \leq 0$. ∎

This is intuitive, as a stronger positive dependence between ignorance and policy preferences implies that the subpopulation of policy-relevant voters is characterized by a relatively higher proportion of people having lower policy preferences. Also notice that proposition 2 can be seen as a special case of proposition 3.

Next section builds a simple model of electoral campaigns with ignorant voters and an embedded tradeoff between policy and valence campaigning on top of the framework proposed in this section to study the influence of this demographic covariation on campaign strategies.

2.2 Valence and Costly Policy Campaigning

I now assume that both politicians are endowed with campaign resources denoted by R_A, R_B. They can allocate these resources between two types of campaign activities: policy and valence. R_j can be interpreted as financial resources, as well as time or human resource endowments, or some measure reflecting a combination of them.

$$P_j + V_j \leq R_j, \qquad j \in \{A, B\}, \tag{41}$$

where P_j and V_j denote the amounts spent on policy and valence campaigning respectively by the politician j.

Valence campaigning takes the form of an additively separable spread-preserving mean shift on the common popularity shock, used extensively in the lobbying literature. Specifically, δ in (3) and (5) takes the following form.[31]

$$\delta = \tilde{\delta} + \mu_A g(V_A) - \mu_B g(V_B), \tag{42}$$

where $\tilde{\delta}$ is distributed uniformly in $[-\alpha, \alpha]$, $g(\cdot)$ is a strictly increasing and strictly concave valence production function with $g(0) = 0$, and μ_j is a parameter measuring individual valence productivities intending to capture any kind of *potential* valence advantage that a candidate might possess. Concavity incorporates the idea of decreasing returns to valence spending, perhaps due to increasing difficulty of coming up with original and/or non-conflicting campaign messages, or due to valence-related messages losing their effectivity after a while. Valence campaigning can be thought as any kind of non-policy related political advertising with the goal of acquiring common support. It is in essence a form of persuasive advertising that can involve emphasizing one's character virtues such as leadership, integrity, competence and empathy as identified by Kinder (1986), associating one's image with emotional collective mobilizers such as patriotism, nationalistic values, etc., attacking approximate-measure zero minorities in highly xenophobic electorates, or even commissioning researchers to dig into opponent's past in order to find material for denigrating his integrity (negative campaigning), which can make a candidate more popular relative to his opponent in the valence dimension

---

[31]Snyder (1989) and Baron (1994) consider instead ratio of expenditures as the variable swinging uninformed voters. (42) can be thought as valence advertising operating as in Grossman and Shapiro (1984), where the probability of a voter voting for a candidate increases in the number of valence-related advertisement messages she receives, with $g(\cdot)$ reflecting the number of valence messages generated for a given amount of valence spending.

if successful.

Regarding electoral platforms, the assumption of costless policy announcements is dropped. Instead, commitment to electoral platforms requires spending resources on policy campaigning, which is assumed to take the form of costly relocation along the policy space given initial positions. This form of policy campaigning is first employed by Harrington and Hess (1996), who used it in the context of positive and negative campaigns trade-off in electoral competitions. If one thinks of the standard spatial policy competition model as a political adaptation of the horizontal differentiation framework from industrial organization, this specification can be considered as the political analogue of costly horizontal differentiation due to marketing costs or redesign costs as in Chang (1992).[32] More specifically, it is assumed that both politicians are endowed with initial locations L_A, L_B on the policy space $[0, 1]$. These can be thought as representing the electorate's premise on candidates' default policy positions based on their party affiliations or past engagements, i.e. their "legacies". In that case, policy campaigning should be interpreted as an effort to convince voters by shifting their perceptions using a combination of rhetorical and factual advertising. Alternatively, they can be interpreted as a commonly known (by informed voters) pair of policy positions, the implementation of which is familiar to candidates due to their past experiences or backgrounds. Under such interpretation, policy campaigning can be thought as detailed outlines, or policy implementation plans credibly showing that the candidate has "what it takes" to implement his electoral promise.[33] Whatever the interpretation is, it is most certain that policy campaigning is a non-trivial task which requires politicians to go beyond the announce and forget strategy in actual electoral competitions. Furthermore, it is usually the case that political competitors do not come in *tabula rasa* and it is reasonable to assume that this fact has an effect on their policy campaign strategies. At worst, the form of policy campaigning employed here should be thought as a reduced-form approximation to whatever mechanism nudging politicians to engage in such activity in the first place. One thing to keep in mind is that policy messages sent as a result of policy campaigning are factual, so post election credibility is not an issue.[34] Returning to the specification itself, the mechanism

---

[32]When combined with valence campaigning, the entire electoral campaigning framework can be seen as the political economy analogue of an industrial organization model of integrated horizontal and vertical differentiation.

[33]At the end of this section, an alternative specification of policy campaigning that might be more suitable for this interpretation is considered.

[34]Addressing such credibility issues is likely to play a major role in existence of policy campaigns and the way they are devised. See Banks (1990) and Chappell (1994).

by which candidates relocate themselves is simple. For each unit of campaign resources allocated to policy campaigning, candidate j relocates (to right or left) within a distance of ρ_j from his initial position. So for a given policy campaign expenditure P_j, his policy position becomes the following.

$$X_j = L_j \pm \rho_j P_j. \tag{43}$$

In (43), ρ_j represents the productivity of candidate j's policy campaigning technology. It can be argued that this technology should be characterized by decreasing returns rather than being linear, reflecting the increasing difficulty that the candidate might face in convincing the electorate when proposing a policy too distant from his legacy. There are two reasons underlying the choice of linearity. First, it simplifies the exposition.[35] Second, as it will be apparent later on, what matters for resource allocation is relative returns between policy and valence campaigning. Since valence campaigning is already characterized by a decreasing returns technology, marginal rate of substitution between campaign types will be decreasing. Furthermore, convexity of voter disutilities already captures the idea of decreasing returns to policy campaigning, because it becomes less effective (in terms of increasing the chance of victory) as the politician draws closer to the bliss point of the swing voter. Let $x_s = X^*$, where X^* denotes the solution to (22). This is the bliss-point of the old swing voter (henceforth referred as pseudo-swing voter) under no valence campaigning and costless policy commitment setting described in the previous section. I assume $L_A < x_s < L_B$. Furthermore, following inequalities are assumed to hold.

$$L_A + \rho_A R_A < L_B,$$
$$L_B - \rho_B R_B > L_A. \tag{44}$$

So a left-wing politician cannot campaign his way to the right of his right-wing opponent and vice versa.[36] Positions of L_A and L_B relative to x_s, along with inequalities in (44) imply that candidates will never find it optimal to campaign

---

[35]It is possible to introduce decreasing returns to both types of campaign activities at the expenditure level by using a resource constraint specification $E_p(P_j) + E_v(V_j) \leq R_j$ with convex, increasing E_p, E_v and a linear policy relocation technology as in (43), along with a linear valence generation process $\delta = \tilde{\delta} + \mu_A V_A - \mu_B V_B$. After appropriate substitutions, this formulation would have a reduced-form representation very close to the current model without any qualitative impact on conclusions. This is because $g(R_j - P_j)$ is concave in P_j in the current specification, and $V_j = E_v^{-1}(R_j - E_p(P_j))$ is concave in P_j in the alternative formulation.

[36]Henceforth, candidates A and B will be referred to as left and right-wing respectively.

their way to a more extreme position than their initial policy positions.[37] So one can denote post policy campaigning positions as follows.

$$X_A \equiv X_A(P_A; L_A) = L_A + \rho_A P_A, \qquad (45)$$

$$X_B \equiv X_B(P_B; L_B) = L_B - \rho_B P_B \qquad (46)$$

Timing of the electoral campaigning game is similar to the timing in the previous section, with politicians simultaneously choosing their campaign allocations (P_j, V_j) at the first stage.[38] Informational structure also remains the same. I assume that voters cannot observe campaign expenditures and informed voters get a perfectly informative signal on policy positions (X_A, X_B) following policy campaigning. Ignorant voters vote purely on the basis of valence dimension as earlier, which is now influenced by valence campaigning.[39] One possible objection to the application of this behavioural framework to a campaign setting can underline selective reception by ignorant voters. The fact that they get influenced by valence advertising but at the same time remain in the dark regarding policy might appear odd, especially if the contents of both types of campaigns are delivered through the same medium (e.g. mass media). It should be kept in mind that underlining and backing up a specific policy proposal would usually demand a lengthier and more elaborated campaign, often requiring a sequence of presentations and repeated communication, whereas valence issues can be handled within a much easier package to consume such as simple catch phrases.[40] This difference in the nature of policy and valence campaigning can imply that even though they use the same medium to reach their audiences, their transmission rates can greatly vary, making selective reception plausible. Another argument

---

[37]This is due to such positions constituting strictly dominated strategies.

[38]Changing the timing between policy and valence campaigning has no effect on the results due to the particular specification of probabilistic voting model employed. If one used a stochastic preference framework where politicians are uncertain about voters' policy preferences instead of the stochastic partisanship specification employed in the current model, then the timing would matter. See Ashworth and Bueno de Mesquita (2009).

[39]One might argue that if uninformed voters are rational, and legacies, as well as other parameters such as resource endowments are known to them, then they can solve politicians' problem and deduce their policy proposals from the valence shock they receive, as they observe $g(V_A) - g(V_B)$. A way to get around this conceptual difficulty would be to assume that ignorant voters have neither any information, nor any priors on these parameters, facing Knightian uncertainty. Another way would be to reverse engineer appropriate priors that would imply beliefs that put all the mass on identical policy positions. Any *post hoc* justification attempt might be considered a stretch, in which case one can simply interpret them as boundedly rational voters.

[40]A nice example of this is the slogan "A Tall Man", used in Turkish presidential elections in 2014, referring to the height of the Islamist candidate Recep Tayyip Erdoğan, possibly to emphasize a connection between his physical attributes and leadership skills.

can be built around the nature of ignorance. If one takes ignorance as a much more fundamental and cognitive phenomenon, then the distinction reduces to one between sophisticated voters qualified to judge the content of policy campaigns and unsophisticated voters who lack the means to do so.[41]

Turning the attention back to the model, given campaign allocations (P_A, V_A) and (P_B, V_B), a policy aware voter with policy preference x votes for the candidate A if the following inequality is satisfied.

$$\sigma_I^x > \mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|) - \tilde{\delta} + \mu_B g(V_B) - \mu_A g(V_A), \qquad (47)$$

where X_A and X_B are functions of P_A and P_B respectively, as defined in (45) and (46). Using the fact that $\tilde{\delta}$ is uniformly distributed in $[-\alpha, \alpha]$, for a given realization of $\tilde{\delta}$, the probability that a policy aware voter with preference x votes for A is given by the following.

$$\mathbb{P}(\sigma_I^x > \cdot) = \frac{1}{2} + \frac{1}{2\beta_I} \left\{ \mathcal{D}(|X_B - x|) - \mathcal{D}(|X_A - x|) + \tilde{\delta} + \mu_A g(V_A) - \mu_B g(V_B) \right\}$$
$$(48)$$

Similar calculations show that the probability of an ignorant voter with preference x voting for A is given by the following.

$$\mathbb{P}\left(\sigma_U^x > \cdot\right) = \frac{1}{2} + \frac{1}{2\beta_U} \left\{ \tilde{\delta} + \mu_A g(V_A) - \mu_B g(V_B) \right\}. \qquad (49)$$

To establish a baseline and to obtain some useful comparative statics, assume first that informational status and policy preferences are independent and let $\mathbb{P}(I = 1|x) = \hat{\Gamma}$ for all $x \in [0, 1]$. Let $S_A(\tilde{\delta}) = S_A^I(\tilde{\delta}) + S_A^U(\tilde{\delta})$ denote the total measure of voters voting for candidate A for a given realization of $\tilde{\delta}$. Candidate A wins the election if he gets at least half of the votes, i.e. $S_A(\tilde{\delta}) > 1/2$. Calculations analogous to those made in the previous section show that the probability of an electoral victory for candidate A is given by the following.

$$\mathbb{P}(S_A(\tilde{\delta}) > \frac{1}{2}) = \frac{1}{2} + \frac{1}{2\alpha} \left\{ \tilde{\Gamma} \int_0^1 \left[\mathcal{D}(|X_B - x|) - \mathcal{D}(|X_A - x|) \right] \mathrm{d}F(x) \right.$$
$$\left. + \mu_A g(V_A) - \mu_B g(V_B) \right\}, \qquad (50)$$

---

[41]This is the interpretation used by Bardhan and Mookherjee (2000) in a similar model with lobbying. It is also a common theme in Turkish literature and cinema attempting to portray politicians campaigning in rural areas.

where $\tilde{\Gamma} \equiv \tilde{\Gamma}(\hat{\Gamma}, \tilde{\beta})$ is as defined in (17), with $\tilde{\beta} = {}^{\beta_I}\!/_{\beta_U}$ representing the relative ideological (non-policy) heterogeneity of informed voters. Victory probability for candidate B is the mirror-image of this:

$$\mathbb{P}(S_B(\tilde{\delta}) > \frac{1}{2}) = \frac{1}{2} + \frac{1}{2\alpha}\left\{\tilde{\Gamma}\int_0^1 \left[\mathcal{D}(|X_A - x|) - \mathcal{D}(|X_B - x|)\right]\mathrm{d}F(x) \right.$$
$$\left. + \mu_A g(V_B) - \mu_B g(V_A)\right\}. \tag{51}$$

A Nash equilibrium to the campaign game will be a pair of campaign allocations $((P_A^*, V_A^*); (P_B^*, V_B^*))$ constituting mutual best-responses.

$$(P_A^*, V_A^*) \in \arg\max_{\substack{(P_A, V_A) \in \mathbb{R}_+^2 \\ \text{s.t. } P_A + V_A \leq R_A}} \mathbb{P}(S_A(\delta; P_A, V_A, P_B^*, V_B^*) > \frac{1}{2}), \tag{52}$$

$$(P_B^*, V_B^*) \in \arg\max_{\substack{(P_B, V_B) \in \mathbb{R}_+^2 \\ \text{s.t. } P_B + V_B \leq R_B}} \mathbb{P}(S_A(\delta; P_A^*, V_A^*, P_B, V_B) > \frac{1}{2}). \tag{53}$$

Strict convexity of voter disutilities, along with strict concavity of valence campaigning technologies ensures that the campaign game has a unique pure-strategy Nash equilibrium.

Proposition 4 (Campaign Game - Existence and uniqueness of pure-strategy Nash equilibrium): The campaign game with costly policy relocation and valence campaigning has a unique pure-strategy Nash equilibrium.

Proof: Strategy spaces, being $\left\{(P_j, V_j) \in \mathbb{R}_+^2 : P_j + V_j \leq R_j\right\}$, are convex and compact. From continuity of $\mathcal{D}(\cdot)$ and $g(\cdot)$, $\mathbb{P}(S_k(\delta; P_k, N_k, P_j, N_j) > \frac{1}{2})$ is continuous in (P_j, N_j) for any (P_k, N_k) and vice versa for $k, j \in \{A, B\}$ and $k \neq j$. Since $\mathcal{D}(\cdot)$ is strictly increasing and strictly convex, $-\mathcal{D}(\cdot)$ is strictly decreasing and strictly concave. Furthermore; $|B + \rho P - x|$ is convex in P for any x. It follows that $-\mathcal{D}(|B + \rho P - x|)$ is strictly concave in P for any x. Integral of (strictly) concave functions is (strictly) concave, and $g(V_A)$ is strictly concave as well. So $\mathbb{P}(S_k(\delta; P_k, N_k, P_j, N_j) > \frac{1}{2})$ is strictly concave in (P_k, N_k) for any (P_j, N_j) for $k \in \{A, B\}$ and $k \neq j$. Thus, from Debreu-Glicksberg-Fan theorem, the electoral game has a unique pure-strategy Nash equilibrium, characterized by a pair of campaign allocations $((P_A^*, V_A^*); (P_B^*, V_B^*))$ satisfying (52). $\blacksquare$

Since problems of candidates are symmetric, I focus on the problem of candidate A for the subsequent discussion. At any optimal campaign allocation, all resources should be spent, i.e. constraints in (52) should bind. If one of them was

slack, then that politician could unambiguously increase his probability of victory by increasing V_j, as $g(\cdot)$ is strictly increasing. Substituting for $V_A = R_A - P_A$, the necessary and sufficient first-order condition for an equilibrium strategy becomes the following after simplifying.[42]

$$
\left\{ -\int_0^{L_A + \rho_A P_A} \mathcal{D}'(L_A + \rho_A P_A - x)\mathrm{d}F(x) \right.
$$
$$
\left. + \int_{L_A + \rho_A P_A}^1 \mathcal{D}'(x - (L_A + \rho_A P_A))\mathrm{d}F(x) \right\} \leq \frac{\mu_A}{\tilde{\Gamma}\rho_A} g'(R_A - P_A) \quad (=\ \text{if } P_A > 0).
$$
$$
\tag{54}
$$

An immediate fact visible from (54) is the no-longer-neutrality of the fraction of informed voters, even if there is no statistical dependence between informational status and policy preferences. This is because the trade-off between policy and valence campaigning implies a competition between these two activities for scarce campaign resources. Under such competition, fraction of informed voters have an effect on campaign structures because it directly affects relative productivities of different campaign activities.

There are two types of corner solutions that can arise from (54). Analysis of these reveals under what conditions candidates might focus exclusively on policy or valence campaigning.

1. Pure Valence Campaigning ($\mathbf{P_A = 0}$, $\mathbf{V_A = R_A}$):

Candidate A will conduct a valence-only campaign if the following inequality is satisfied.

$$
-\int_0^{L_A} \mathcal{D}'(L_A - x)\mathrm{d}F(x) + \int_{L_A}^1 \mathcal{D}'(x - L_A)\mathrm{d}F(x) \leq \frac{\mu_A}{\tilde{\Gamma}\rho_A} g'(R_A). \tag{55}
$$

Notice that the left-hand side of (55) is strictly positive, as it is assumed that $L_A < x_s$. One can see that several factors will play a role in determining whether candidate A's campaign will have a pure-valence focus or not. First, relative curvatures of voter disutility and valence technology will matter, since these have a direct impact on relative marginal gains from both activities in the corner. Second, relative average productivity of valence campaigning with respect to policy campaigning (μ_A/ρ_A) will be important. If one thinks of high μ_A as a *potential* valence advantage and low ρ_A as a comparative disadvantage when designing policy campaigns, then a candidate whose valence related consolidation skills greatly

---

[42]Inequality (54) is the optimality condition when $P_A \in [0, R_A)$. See (56).

outmatching his policy explaining skills might go for a purely valence focused campaign. Third, both a high fraction of ignorant voters, and a strong relative heterogeneity of informed voters can tip the scales towards a valence only campaign. While the former reduces marginal returns to policy campaigning by decreasing the size of its target audience, the latter has the same effect by decreasing the effective size of the policy aware electorate. Finally, proximity of s candidate's initial position to the preferred policy of the pseudo-swing voter is also an influencing factor. The closer L_A is to x_s, the smaller the left-hand side of (55) will be, reflecting the effect of decreasing marginal returns to policy campaigning kicking in when the candidate is already close to the centripetal core in the policy dimension.

2. Pure Policy Campaigning ($\mathbf{P_A = R_A}$, $\mathbf{V_A = 0}$):

Candidate A's electoral campaign will have an exclusive policy focus if the following inequality is satisfied.

$$\left\{ -\int_0^{L_A + \rho_A R_A} \mathcal{D}'(L_A + \rho_A R_A - x)\mathrm{d}F(x) \right.$$

$$\left. + \int_{L_A + \rho_A R_A}^1 \mathcal{D}'(x - (L_A + \rho_A R_A))\mathrm{d}F(x) \right\} \geq \frac{\mu_A}{\tilde{\Gamma}\rho_A}g'(0). \qquad (56)$$

Same factors that might encourage a candidate to conduct a valence only campaign can also lead him to conduct an exclusively policy based campaign if they come in opposite forms. For instance, a low potential valence advantage, a high fraction of informed voters, or a combination of high policy campaigning efficiency and a relatively homogenous policy aware electorate might lead to low relative returns for valence campaigning in terms of marginal gains in the probability of an election victory even if the candidate is not spending any resources on it, and this can nudge the politician towards an exclusively policy focused campaign. Another important factor is the effective size of campaign resources in terms of policy campaigning ($\rho_A R_A$). If this is already large in the sense that $L_A + \rho_A R_A > x_s$, then the candidate would never spend all his resources in policy campaigning because he can relocate to x_s by spending a lower amount of resources, and any resources spent on policy campaigning over and above this level would have a negative gross marginal return, excluding the opportunity cost in terms of valence loss. So a necessary condition to conduct a pure policy campaign is to either have limited campaign resources, or a relatively extremist legacy (further away from x_s).

Henceforth, I assume that politicians have interior solutions for their campaign problems. For candidate A, this implies that the first-order condition should be satisfied with equality at the optimal resource allocation pair (P_A^*, V_A^*):

$$\left\{ -\int_0^{L_A + \rho_A P_A} \mathcal{D}'(L_A + \rho_A P_A - x)\mathrm{d}F(x) \right.$$
$$\left. + \int_{L_A + \rho_A P_A}^1 \mathcal{D}'(x - (L_A + \rho_A P_A))\mathrm{d}F(x) \right\} = \frac{\mu_A}{\tilde{\Gamma}\rho_A} g'(R_A - P_A). \quad (57)$$

First thing to notice from (57) is that even if candidates have sufficient resources, they would never relocate to the preferred policy of the old swing voter (pseudo-swing voter).

Lemma 5 (Tragedy of the old swing voter): In any interior optimum, candidates' final policy positions should be in between their initial positions and x_s, i.e.

$$X(P_A^*) = L_A + \rho_A P_A^* < x_s < X(P_B^*) = L_B - \rho_B P_B^*. \quad (58)$$

Proof: The right-hand side of (57) is strictly positive for any $P_A < R_A$, and for any $L_A + \rho_A P_A \geq x_s$, the left-hand side of (57) is non-positive. ∎

This result might appear slightly odd at first, as x_s refers to the position of the swing voter when the valence dimension is not taken into account. If, say, a politician has sufficient funds for relocating to x_s, why doesn't he get the best of both worlds by positioning himself in accordance with the centripetal force arising from electorate's policy tastes and spend the rest of his resources for valence campaigning? In fact, there is a reason why the voter preferring x_s is called the pseudo-swing voter, and that reason is precisely the fact that x_s is no longer swing, but an inframarginal voter. Holding opponent's valence fixed, any resources spent to valence campaigning introduce an electoral bias in favour of the politician, which implies that the *ex ante* decisive voter shifts to a position closer to the legacy of the politician. As a corollary, presence of valence campaigning introduces policy divergence, and this divergence is not necessarily a byproduct of candidate characteristics (except initial positions) as long as both candidates' solutions are interior. Presence of legacies, opportunity to campaign for valence, along with the policy-valence trade-off act as a centrifugal force, pulling politicians back to their initial policy positions. This is because the marginal cost of a slightly higher (or lower if its candidate B) policy no longer consists solely of

the marginal probability loss of victory due to increasing disutilities of voters to the left, but also reflects the opportunity cost of a lower valence advantage. If one takes the view that valence campaigning (at least partially) operates through a partisanship effect that has no material implications for *ex post* well-beings of voters, then the outcome in (58) constitutes a wasteful divergence from the egalitarian planner outcome.

Next proposition presents the way campaign structures respond to changes in politician and demographic characteristics.

Proposition 5 (Comparative statics for the campaign game): Comparative statics for candidate A's campaign strategies are given by the following matrix.

$$
\operatorname{sgn}\left\{
\begin{bmatrix}
\begin{bmatrix}
\frac{\partial P_A}{\partial L_A} & \frac{\partial P_A}{\partial \hat{\Gamma}} & \frac{\partial P_A}{\partial \tilde{\beta}} \\
\frac{\partial P_A}{\partial \rho_A} & \frac{\partial P_A}{\partial \mu_A} & \frac{\partial P_A}{\partial R_A} \\
\frac{\partial V_A}{\partial L_A} & \frac{\partial V_A}{\partial \hat{\Gamma}} & \frac{\partial V_A}{\partial \tilde{\beta}} \\
\frac{\partial V_A}{\partial \rho_A} & \frac{\partial V_A}{\partial \mu_A} & \frac{\partial V_A}{\partial R_A}
\end{bmatrix}
\end{bmatrix}
\right\}
=
\begin{bmatrix}
- & + & - \\
? & - & + \\
+ & - & + \\
? & + & +
\end{bmatrix}.
\tag{59}
$$

Proof: Appendix. ∎

Candidate B will have the same comparative statics, except for opposite signs regarding the effect of L_B. This is because an increase in L_B takes candidate B farther away from x_s rather than taking him closer to it as an increase in L_A does for candidate A. First thing to notice is that for almost all derivatives, those involving V_j have opposite signs compared to their P_j counterparts. This is a trivial consequence of the fact that $V_j = R_j - P_j$. Second, both types of campaign allocations are increasing in available resources R_j. If all of the resource increase was allocated to one type of campaign activity, its' marginal benefit would exceed that of the other. In analogy with consumption theory, both campaign types are normal goods for politicians.[43] Comparative statics effects of both $\hat{\Gamma}$ and $\tilde{\beta}$ are intuitive. Policy campaigning increases in the former as the audience responding to it gets larger, whereas it decreases in the latter because the more heterogenous policy aware voters become, the harder it becomes to satisfy them as a group via policy proposals. Resources allocated to policy campaigning decrease in the

---

[43]Which type increases more depends on parameters governing productivities of these two activities, as well as relative curvatures of valence production function and voter disutility. A high μ_A, a low ρ_A or $\tilde{\Gamma}$, as well as a more concave $-\mathcal{D}$ relative to g makes the share of valence campaigning going up more likely. Of course if the politician is at a corner solution, all of the resource increase is absorbed by one type of campaign activity.

valence productivity parameter μ_A, since an increase in the latter imply higher opportunity costs in terms of foregone valence advantage per unit of policy campaign spending. On the other hand, an increase in policy campaign productivity has an ambiguous effect on the campaign composition. The ambiguity stems from the fact that while an increase in ρ_A increases the relative appeal of policy campaigning with respect to valence advertising, it also implies that any policy position can be reached by allocating a smaller amount of resources. Finally, an increase in L_A increases policy campaigning. This has to do with the proximity effect mentioned previously.[44] The less extreme a candidate's legacy is, the closer he is to the bliss point of the pseudo-swing voter. This implies that the marginal gain in probability of victory from converging to a platform closer to x_s falls. If one considers negative campaigning as a form of valence campaigning, then Brueckner and Lee (2013) provide some empirical support for the idea that proximity to center has an increasing effect on the valence campaigning.[45]

Within this framework, any divergence between candidates regarding campaign structures can be traced back to differences in candidate characteristics, campaign technologies and resources. However, such divergence can also arise under an otherwise symmetric setting purely due to a statistical dependence between informational status and policy preferences. This is because such dependence, under costly policy campaigning and politician legacies, gives rise to an asymmetry in the respective audiences of candidates. To show this, I start by constructing a symmetric equilibrium.

Lemma 6 (A special symmetric campaigning equilibrium): Suppose $\mu_A = \mu_B = \mu$, $\rho_A = \rho_B = \rho$, $R_A = R_B = R$. Assume a symmetric $F(\cdot)$ and suppose that $L_A + L_B = 1$. Then the strategy profile in the unique Nash equilibrium consists of $P_A = P_B = P^*$ and $V_A = V_B = V^*$.

Proof: Under the stated assumptions, the first-order necessary and sufficient condition for candidate A becomes the following.

$$
\left\{ -\int_0^{L_A + \rho P_A} \mathcal{D}'(L_A + \rho P_A - x)\mathrm{d}F(x) \right.
$$
$$
\left. + \int_{L_A + \rho P_A}^1 \mathcal{D}'(x - (L_A + \rho P_A))\mathrm{d}F(x) \right\} = \frac{\mu}{\tilde{\Gamma}\rho}g'(R - P_A). \qquad (60)
$$

---

[44]Note that proximity to pseudo-swing voter should not be taken as the absolute distance between candidate's initial position and x_s. It is rather the effective distance taking the mass of voters in between two policies into account, i.e. $|F(x_s) - F(L_j)|$.

[45]They find that in US Senatorial elections between 1992 to 2002, centrist candidates tended to invest more heavily in negative campaigning, controlling for other factors.

Same condition for candidate B is given as follows.

$$\left\{ \int_0^{L_B - \rho P_B} \mathcal{D}'(L_B - \rho P_B - x)\mathrm{d}F(x) \right.$$
$$\left. - \int_{L_B - \rho P_B}^1 \mathcal{D}'(x - (L_B - \rho P_B))\mathrm{d}F(x) \right\} = \frac{\mu}{\tilde{\Gamma}\rho}g'(R - P_B). \qquad (61)$$

For any interior P in accordance with lemma 5, the following equality should hold.

$$\int_0^{L_A + \rho P} \mathcal{D}'(L_A + \rho P - x)\mathrm{d}F(x) = \int_{L_B - \rho P}^1 \mathcal{D}'(x - (L_B - \rho P))\mathrm{d}F(x). \qquad (62)$$

To see this, fix some $k > 0$ and take two preferences $L_A + \rho P - k$ and $L_B - \rho P + k$. Then their marginal disutilities are equal, being $\mathcal{D}'(k)$. Moreover, symmetry of the distribution, along with symmetrically extreme (equidistant to $x_s = \frac{1}{2}$) legacies ensure that these two preferences are supported by the same mass of voters, as the following equality shows.

$$f(L_A + \rho P - k) = f(L_B - \rho P + k) = f(1 - (L_A + \rho P - k)), \qquad (63)$$

where the third equality follows from $L_A + L_B = 1$. Varying k in $0 < k < L_A + \rho P = 1 - (L_B - \rho P)$, equality in (62) follows. By an analogous argument, one can show the following equality.

$$\int_{L_A + \rho P}^1 \mathcal{D}'(x - (L_A + \rho P))\mathrm{d}F(x) = \int_0^{L_B - \rho P} \mathcal{D}'(L_B - \rho P - x)\mathrm{d}F(x). \qquad (64)$$

Since P was arbitrary, (63) and (64) imply that equations describing first-order conditions in (60) and (61) are identical. Since $\mathcal{D}$ is strictly convex and g is strictly concave, there is a unique P^* which solves both of them, so in equilibrium, $P_A = P_B = P^*$ and $V_A = V_B = V^*$. ∎

One interesting implication of lemma 6 is that no candidate ends up with a valence advantage in the symmetric equilibrium, as valence effects of their campaign efforts cancel each other out. Next, I introduce some positive dependence between ignorance and policy preferences by holding the total measure of ignorant voters constant, and consider the asymmetric impact of this on candidates' campaign strategies.[46]

---

[46]This exercise, and in particular proposition 7 can be informally interpreted as "partially differentiating" with respect to the second moment, holding the first moment constant.

Proposition 6 (Asymmetric impact of correlated ignorance on campaign allocations): Starting from the symmetric setting in lemma 6, introduce some positive dependence between ignorance and policy preferences in the assumption 1 sense while keeping the mass of ignorant voters constant. Then letting (P_j^c, V_j^c) denote the campaign allocations under such dependence, $P_A^c \leq P_A = P^* = P_B \leq P_B^c$, and consequently, $V_B^c \leq V_B = V^* = V_A \leq V_A^c$.

Proof: Under the symmetric setting given in lemma 6, candidate A's first-order condition can be rewritten as follows.

$$-F(L_A + \rho P^*)\mathbb{E}\left[\mathcal{D}'(L_A + \rho P^* - x)|x \leq L_A + \rho P^*\right]$$
$$+(1 - F(L_A + \rho P^*))\mathbb{E}\left[\mathcal{D}'(x - (L_A + \rho P^*))|x \geq L_A + \rho P^*\right] \qquad (65)$$
$$= \frac{\mu}{\tilde{\Gamma}\rho}g'(R - P^*).$$

As before, assumption 1 implies the following inequalities.

$$F(L_A + \rho P^*|I = 1) \geq F(L_A + \rho P^*), \qquad (66)$$
$$F(x|x \leq L_A + \rho P^*, I = 1) \geq F(x|x \leq L_A + \rho P^*), \qquad (67)$$
$$F(x|x \geq L_A + \rho P^*, I = 1) \geq F(x|x \geq L_A + \rho P^*). \qquad (68)$$

The last two, along with the facts that $\mathcal{D}'(L_A + \rho P^* - x)$ is decreasing in x for x in $0 \leq x < L_A + \rho P^*$ and $D'(x - (L_A + \rho P^*))$ is increasing in x for x in $L_A + \rho P^* < x \leq 1$ imply the following inequalities.

$$\mathbb{E}\left[\mathcal{D}'(L_A + \rho P^* - x)|x \leq L_A + \rho P^*, \ I = 1\right]$$
$$\geq \mathbb{E}\left[\mathcal{D}'(L_A + \rho P^* - x)|x \leq L_A + \rho P^*\right]; \qquad (69)$$
$$\mathbb{E}\left[\mathcal{D}'(x - (L_A + \rho P^*))|x \geq L_A + \rho P^*, I = 1\right]$$
$$\leq \mathbb{E}\left[\mathcal{D}'(x - (L_A + \rho P^*))|x \geq L_A + \rho P^*\right]. \qquad (70)$$

So when evaluated at P^*, candidate A's first-order condition under correlated ignorance does not hold, but rather reflects the fact that marginal benefit of policy campaigning is below that of valence campaigning. This is shown in the inequality below.

$$-F(L_A + \rho P^*|I = 1)\mathbb{E}\left[\mathcal{D}'(L_A + \rho P^* - x)|x \leq L_A + \rho P^*, I = 1\right]$$
$$+(1 - F(L_A + \rho P^*|I = 1))\mathbb{E}\left[\mathcal{D}'(x - (L_A + \rho P^*))|x \geq L_A + \rho P^*, I = 1\right] \qquad (71)$$
$$\leq \frac{\mu}{\tilde{\Gamma}\rho}g'(R - P^*).$$

Left-hand side of (71) is strictly decreasing in P due to strict convexity of $\mathcal{D}$, and right-hand side is strictly increasing in it due to concavity of g.[47] It follows that $P_A^c \leq P^*$, where P_A^c is the allocation at which (71) holds with equality. This directly implies that $V_A^c \geq V^*$. By similar steps, one can deduce that the first-order condition of candidate B under correlated ignorance becomes the following when evaluated at P^*.

$$F(L_B - \rho P^*|I = 1)\mathbb{E}\left[\mathcal{D}'(L_B - \rho P^* - x)|x \leq L_B - \rho P^*, I = 1\right]$$
$$-(1 - F(L_B - \rho P^*|I = 1))\mathbb{E}\left[\mathcal{D}'(x - (L_B - \rho P^*))|x \geq L_B - \rho P^*, I = 1\right] \quad (72)$$
$$\geq \frac{\mu}{\tilde{\Gamma}\rho}g'(R - P^*).$$

It follows that $P_A^c \leq P_A = P^* = P_B \leq P_B^c$, and consequently $V_B^c \leq V_B = V^* = V_A \leq V_A^c$. ∎

For both politicians, the rationale behind spending resources on policy campaigning is to moderate their initial positions assuming that the population is relatively more concentrated at the center. At any allocation, by spending slightly more on policy campaigning, politicians gain the support of some voters at the center, lose the support of some tail voters, and also fail to obtain the support of some voters in the entire population due to foregone opportunity of valence campaigning. The symmetry of the environment before taking the dependence between policy preferences and informational status into account implies that the marginal calculus reflecting these gains and losses are the same for both politicians. Consequently, they each adopt the same campaign structure and end up as center-left and center-right candidates when the voting day arrives. Once the statistical dependence is taken into account, this picture changes. The shift in the mass of distribution strengthens the centrifugal force (valence) for the left candidate and the centripetal force (policy) for the right candidate. Technically, this is due to decreasing returns to moderation. Keeping in mind that only the policy preferences of informed voters are taken into account for policy proposals, a left-shift in the mass of the preference distribution due to dependence implies that while candidate A's legacy becomes more moderate, candidate B's legacy becomes less moderate. Another way of expressing this would be to point out that candidate A's initial proximity to pseudo-swing voter's bliss point increases and candidate B's initial proximity to it decreases. Both reflect the simple fact that while

---

[47]To see this more clearly, just differentiate (57) after conditioning on $I = 1$ with respect to P_A, recognizing the fact that variations at the boundary of the integral does not matter due to $\mathcal{D}'(0) = 0$

marginal losses (gains) due to diverging (converging to) from left are weighted more heavily, marginal gains (losses) due to converging to (diverging from) right are weighted less. As a result, relative marginal returns to policy campaigning are much higher (and stay higher up to a significant level of resource allocation) for candidate B, which manifests as diverging campaign structures. Note that proposition 6 and the above discussion assume an interior solution under statistical dependence. In fact, if such dependence is strong, it is possible that one, or even both candidates end up in corner solutions. Naturally, if such corner solutions were to occur, they would imply a pure valence campaigning candidate A and/or a pure policy campaigning candidate B.[48]

The result in proposition 6 might appear counter-intuitive at first, as the politician whose support base favours ignorant voters (candidate B) invests more in policy campaigning, whereas the candidate with a relatively more policy aware support base conducts a more valence oriented campaign. Intuitively, politicians would like to design their campaign structures in order to establish a balance between policy support and valence support. Politicians whose natural base consists mostly of ignorant voters have to engage in a much more policy intensive campaign in order to obtain a satisfying level of policy support, as they require the support of voters at the opposite end of the policy spectrum.[49] Conversely, politicians whose legacies are already more or less aligned with the policy-relevant, informed segment of the electorate can get away with a much higher level of valence activity, aiming to obtain support indiscriminately. This kind of reasoning offers a potential explanation for a pattern which emerged during the Turkish national elections in 2002. The current incumbent party AKP, a newcomer back in 2002, conducted a much more policy oriented campaign consisting of a variety of well-defined projects and detailed feasibility plans. This was in stark contrast to its main opponent CHP, whose campaign was almost exclusively dedicated to valence issues with numerous invocations of nationalistic themes and devotion of a significant amount of resources to negative campaigning.[50] While most analysts argued that this was indicative of AKP's campaign team being better organized and more professional, many observers remained puzzled due to two reasons. First, AKP is the heir of a political tradition that is frequently labelled as "Islamic left", yet its campaign was quite moderate, with privatization, lib-

---

[48]If the association between ignorance and policy preferences was instead negative, then the opposite conclusion would prevail.

[49]Natural base refers to voters at the vicinity of legacies in the context of the current model.

[50]Due to Islamist heritage of AKP, negative campaigning efforts mainly revolved around themes of Islamic fundamentalism. This pattern repeated itself during 2007 national elections. See Dinç (2008).

eralization and other market-based themes playing prominent roles. Second, its natural base consists of relatively undereducated voters compared to the population average, and according to many, this fact appeared hard to reconcile with the almost-policy-exclusive nature of AKP's election campaign. The folk wisdom of associating less educated people with relative policy unawareness is no doubt a simplification, but a simplification possibly containing a grain of truth. If it contains that grain of truth, then as the model shows, there is no reason for puzzlement. It is perfectly reasonable for parties to set their target audiences outside their natural bases. Mapping the model to reality, AKP can be thought as the candidate B, trying to accommodate the policy tastes of relatively more educated, urban voters by moderation and CHP as the candidate A targeting the electorate by valence campaigning.[51]

Returning back to the model, one does not need the symmetry in lemma 6 to show the asymmetric impact of correlated ignorance on campaign compositions. In fact, with an argument analogous to proposition 3, one can show that a stronger positive association between ignorance and policy preferences might imply a stronger campaign divergence under any parameter configuration.

Proposition 7 (Asymmetric campaign monotonicity): Let $\mathcal{F}$ be a family of absolutely continuous and smooth conditional distributions $F(x|I = 1, \theta)$ continuously indexed according to a parameter θ measuring the strength of the positive association between ignorance and policy preferences as in assumption 2. Furthermore, assume that this family of conditional distributions is associated with a family of joint probability distributions of (I, x) that possess the same marginal distribution for I. Then assuming interior solutions, following describes the effect of strengthening the positive association on campaign allocations.

$$
\frac{\partial P_A(\theta)}{\partial \theta} \leq 0 \iff \frac{\partial V_A(\theta)}{\partial \theta} \geq 0,
$$
$$
\frac{\partial P_B(\theta)}{\partial \theta} \geq 0 \iff \frac{\partial V_B(\theta)}{\partial \theta} \leq 0.
$$

Proof: Similar to proposition 3. Details are in the appendix. ∎

Proposition 7 is just an extension of proposition 6. It states that holding everything else constant, a stronger monotonic dependence between informational status and policy preferences makes the candidate whose is initially more aligned

---

[51]The former is even more apparent when one reflects upon AKP's strategy of allying itself with liberal columnists to market itself to urban upper middle class.

with ignorant voters devote less resources to valence campaigning and vice versa.

Corollary 2 (Increasing campaign divergence): If $P_B(\theta) \geq P_A(\theta)$, then $\mathrm{d}(P_B(\theta)-P_A(\theta))/\mathrm{d}\theta \geq 0$.

Note that proposition 7 does not imply an increasing campaign divergence *per se*, because differences in campaign technologies, candidate characteristics and legacies could imply $P_A(\theta) \geq P_B(\theta)$ to begin with. However, if candidates were more or less symmetric before taking the dependence in account, or candidate B was conducting a more policy oriented campaign under the given dependence strength despite asymmetry, then an increase in θ would imply a further campaign divergence.

Next subsection explores same ideas under an alternative policy campaigning specification that allows for costless policy announcements.

2.2.1 Ambiguity Reducing Policy Campaigning

The assumption that ignorant voters are unable to condition their votes on anything policy-related is maintained and notation remains unchanged. Again, candidates have certain amounts of campaign resources which they must allocate between policy and valence campaigning and the latter generates universal support in an additively separable fashion via a valence-production function $g(\cdot)$. The novelty is twofold. First, politicians can costlessly announce a policy, but given the announcement, there is some black-box commitment issue which introduces a random element to the *ex post* policy to be implemented in the eyes of policy aware voters. Upon seeing the policy announcement X_j and after observing the campaign process, these voters believe that the *ex post* policy to be implemented is a random variable $\tilde{X}_j$ described as follows.[52]

$$\tilde{X}_j = X_j + \epsilon, \tag{73}$$

where $\epsilon \sim [-k, k]$ is a zero-mean random variables with variance ν and $0 < k \leq \min\{L_A, 1 - L_B\}$.[53] The variance of these beliefs are assumed to depend on two factors: distance of announced policy platforms to legacies: L_A, L_B, and the amount spent on policy campaigning: P_A, P_B.

$$\nu \equiv \nu(D_j, P_j), \tag{74}$$

---

[52]For simplicity, I assume that beliefs are common across policy aware voters.

[53]The restriction on the support of ϵ makes sure that no positive probability is assigned to policies outside $[0, 1]$.

where $D_j = |X_j - L_j|$. It is further assumed that ν is twice continuously differentiable with $\nu'_D > 0$, $\nu'_P < 0$, $\nu''_{DP} \le 0$ and that ν is convex. These assumptions imply that the uncertainty surrounding the post election policy decreases in policy campaigning, but increases as the politician positions himself further away from his legacy.[54] The expression in (73) states that politically aware voters find the policy announcement credible enough to believe that it will be implemented "on average" but they don't rule out the possibility of a post election deviation. Furthermore, this possibility is evaluated as being more likely when the policy announcement is further away from politician's legacy. Nevertheless, politicians can engage in policy campaigning to address this commitment issue to some extent, partially convincing the voters. This *ad hoc* way of specifying a connection between beliefs, platforms and campaigning can be thought as an approximation to a reduced form representation of some underlying model with *ex post* institutional elements and politician preferences (and priors on these preferences) determining the scope of *ex ante* signalling and a rule governing Bayesian updating. While microfounding this connection would be an interesting (and a non-trivial) exercise, it is beyond the scope and outside the focus of this subsection.[55] For my purposes, it is sufficient in the sense that it provides a cheap and tractable way of analysing the impact of correlated ignorance on campaign allocations in a setting that captures the near-consensual view that a major purpose of electoral campaigns is to establish credibility, and the obvious implication that less resources being devoted to this activity would hinder the likelihood of achieving this goal. The mechanism depicted in (73) and (74) is in fact a combination of two specifications popular in positive political theory literature. Austen-Smith (1987) and Hinich and Munger (1989) are among first papers assuming that contributors or candidates themselves can spend resources to achieve a reduction in voter uncertainty in the form of variance a decrement. On the other hand, Bernhardt and Ingerman (1985) construct a model where variance of voter beliefs are increasing in the distance between candidate's current and past positions. The mechanism can be justified along the lines of two stories. First, one can imagine that policy campaigning is a credible demonstrator of learning, decreasing the

---

[54] The last one, $\nu''_{DP} \le 0$ says that if more resources to policy campaigning are allocated, then the increase in uncertainty due to announcing a policy further away from the initial legacy is not as dramatic.

[55] This is an active area of research initiated by Banks (1990) who devised a mechanism of campaigning as costly signalling to generate information transmission in order to address commitment issues. Later research focused on establishing information transmission via repeated elections (e.g. Aragonès et al. (2007)), via costly affiliation with screening political parties (e.g. Snyder and Ting (2002)), and on establishing the conditions under which cheap talk might be informative (e.g. Kartik and Van Weelden (2014)).

likelihood of a mistake by well-intending politicians when implementing the announced policy. Second, one can assume that policy campaigning takes the form of "ideological" campaigning. Establishing ideological positions might require repeated engagements in polemical discussions, allocation of significant amounts of time and resources for delivering speeches, appearing in media etc., and the nature of this process might make it quite hard to lie consistently without raising red flags. If there is some well recognized underlying mapping associating ideologies with policies, then the environment can be represented by a specification akin to that described above.[56]

Policy disutility is assumed to be quadratic in its distance from the bliss point. This allows for an easy characterization of the disutility associated with post election ambiguity. For any *ex post* policy implementation X_j, it is given by the following for a voter with bliss point x.

$$U(X_j; x) = -\mathcal{D}(|X_j - x|) = -(X_j - x)^2. \tag{75}$$

Given a policy proposal X_j by a candidate having a legacy L_j and a policy campaigning effort P_j, an informed voter will hold the belief that *ex post* policy to be implemented if j is victorious is a random variable $\tilde{X}_j$ with mean X_j and variance $\nu(D_j, P_j)$.[57] It follows that *ex ante*, expected policy disutility from a candidate j victory for an informed voter with bliss point x is as follows.

$$\mathbb{E}\left[U(\tilde{X}_j; x)|B_j, P_j\right] = -\mathbb{E}\left[(\tilde{X}_j - x)^2|B_j, P_j\right]$$
$$= -(X_j - x)^2 - \nu(D_j, P_j). \tag{76}$$

Given the assumption of sincere voting, her bias towards candidate A, and valence expenditures V_A, V_B, she votes for A if the following inequality is satisfied.[58]

$$\sigma_I^x - (X_A - x)^2 - \nu(D_A, P_A) + \tilde{\delta} + g(V_A) - g(V_B)$$
$$> -(X_B - x)^2 - \nu(D_B, P_B). \tag{77}$$

---

[56]The term "ideology" in the second story is used differently from when it was used before as describing idiosyncratic biases parametrized by σ. It rather refers to "an internally consistent set of propositions that makes both proscriptive and prescriptive demands on human behaviour" (Hinich and Munger (1993)), having well-defined policy implications. See Hinich and Munger (1996) for a detailed framework considering the place of ideology in formal models of political choice.

[57]Higher order moments of the belief distribution will be irrelevant, since disutilities are quadratic.

[58]For simplicity, I assume that no candidate has inherent valence or policy campaigning advantage.

Ignorant voters vote as before, purely on the basis of valence and idiosyncratic biases. The same condition for an ignorant voter with preference x is given below.

$$\sigma_U^x + \tilde{\delta} + g(V_A) > g(V_B). \tag{78}$$

As before, let $S_A(\tilde{\delta}) = S_A^I(\tilde{\delta}) + S_A^U(\tilde{\delta})$ denote the total measure of voters voting for candidate A for a given realization of $\tilde{\delta}$. Candidate A wins the election if $S_A(\tilde{\delta}) > 1/2$. Calculations analogous to those made before, as well as some straightforward manipulations yield the following objective function for candidate A.

$$\mathbb{P}(S_A(\tilde{\delta}) > \frac{1}{2}) = \frac{1}{2} + \frac{\tilde{\tilde{\Gamma}}}{2\alpha} \left\{ \frac{\mathbb{E}[\Gamma(x)]}{2\beta_I} \left(X_B^2 - X_A^2 + \nu(D_B, P_B) - \nu(D_A, P_A) \right) \right.$$
$$\left. + \frac{\mathbb{E}[x\Gamma(x)]}{\beta_I} (X_A - X_B) \right\} + \frac{1}{2\alpha} \left(g(V_A) - g(V_B) \right), \tag{79}$$

where $\tilde{\tilde{\Gamma}}$ is given by the following.

$$\tilde{\tilde{\Gamma}} = \left(\frac{\mathbb{E}[\Gamma(x)]}{2\beta_I} + \frac{1 - \mathbb{E}[\Gamma(x)]}{2\beta_U} \right)^{-1}. \tag{80}$$

Candidate B's objective function is the mirror-image of (79), with indices A and B switched. A pure-strategy Nash equilibrium will be a pair of mutually best-responding campaign allocations and policy proposals.

$$(X_A^*, P_A^*, V_A^*) \in \arg \max_{\substack{X_A \in [0,1], (P_A, V_A) \in \mathbb{R}_+^2 \\ \text{s.t. } P_A + V_A \leq R_A}} \mathbb{P}(S_A(\delta; X_A, P_A, V_A, X_B^*, P_B^*, V_B^*) > \frac{1}{2}),$$

$$(X_B^*, P_B^*, V_B^*) \in \arg \max_{\substack{X_B \in [0,1], (P_B, V_B) \in \mathbb{R}_+^2 \\ \text{s.t. } P_B + V_B \leq R_B}} \mathbb{P}(S_A(\delta; X_A^*, P_A^*, V_A^*, X_B, P_B, V_B) > \frac{1}{2}). \tag{81}$$

The problem in (81) will have a solution because ν and voter disutility are convex and g is concave. If these are strictly convex (voter disutility already is) and strictly concave, then the solution will be unique. The timing assumption is that candidates announce policies and conduct campaigns simultaneously. One could have made the more natural assumption that candidates announce policy platforms before engaging in campaign activities, as a result of which informed voters observe a noisy signal inducing the specified beliefs. Results would not

change due to additively separable effect of campaigning on politicians' objectives. In what follows, I focus only on the effect of a negative correlation (and a stronger negative correlation) between being informed and policy preferences as in the previous subsection.

Assuming $X_A^* > L_A^*$, simplified first-order necessary (and sufficient) conditions for candidate A is given below.[59]

$$X_A = -\frac{1}{2}\frac{\partial \nu(X_A - L_A, R_A - V_A)}{\partial D_A} + \frac{\mathbb{E}[x\Gamma(x)]}{\mathbb{E}[\Gamma(x)]}, \tag{82}$$

$$g'(V_A) = -\tilde{\Gamma}\frac{\partial \nu(X_A - L_A, R_A - V_A)}{\partial P_A}, \tag{83}$$

where $\tilde{\Gamma}$ is as defined in (17). For $X_B^* < L_B^*$, same condition for candidate B becomes the following.

$$X_B = \frac{1}{2}\frac{\partial \nu(L_B - X_B, R_A - V_A)}{\partial D_B} + \frac{\mathbb{E}[x\Gamma(x)]}{\mathbb{E}[\Gamma(x)]}, \tag{84}$$

$$g'(V_B) = -\tilde{\Gamma}\frac{\partial \nu(L_B - X_B, R_B - V_B)}{\partial P_B}. \tag{85}$$

An immediate implication of these first order conditions is policy divergence. This is consistent with Bernhardt and Ingerman (1985), and with the results in the previous section. It is also clear that as long as the environment is symmetric (and informational status is independent so that $\frac{\mathbb{E}[x\Gamma(x)]}{\mathbb{E}[\Gamma(x)]} = \mathbb{E}[x]$), both candidates choose identical campaign allocations. The important thing is, (82) and (84) allow one to easily deduce the impact of a stronger negative correlation between x and I on campaign strategies, as well as platform proposals. To see this, rewrite them as follows.

$$X_A = -\frac{1}{2}\frac{\partial \nu(X_A - L_A, R_A - V_A)}{\partial D_A} + \mathbb{E}[x] + \frac{\mathrm{cov}(x, \Gamma(x))}{\mathbb{E}[\Gamma(x)]}, \tag{86}$$

$$X_B = \frac{1}{2}\frac{\partial \nu(L_B - X_B, R_A - V_A)}{\partial D_B} + \mathbb{E}[x] + \frac{\mathrm{cov}(x, \Gamma(x))}{\mathbb{E}[\Gamma(x)]}. \tag{87}$$

If there was no ambiguity associated with policy proposals, then (86) and (87) together represent the familiar mean-value theorem under quadratic disutilities. Since there is ambiguity, an extra term pulling the swing voter towards the direction indicated by the correlation appears. The strength of this latter force is also what leads to differences in campaign tones.

---

[59]It is not hard to show that if $L_A < \frac{\mathbb{E}[x\Gamma(x)]}{\mathbb{E}[\Gamma(x)]} < L_B$, then $X_A^* > L_A$ and $X_B^* < L_B$. So assume that this condition holds.

Proposition 8 (Stronger correlation, policy proposals and campaign allocations): Holding total number of ignorant voters constant, a stronger negative correlation between x and I leads to both candidates proposing lower policy platforms, candidate A devoting more resources to valence campaigning and candidate B devoting more resources to policy campaigning.

Proof: Appendix. ∎

Proposition 8 is quite intuitive. As the positive correlation between ignorance and policy preferences gets stronger, there is less and less informed people who happen to have high policy preferences. So the average preferred policy of informed voters (which represents the centripetal core in a quadratic framework) shifts left. Since these are the voters whose preferences matter for policy proposals, both candidates accommodate this shift by announcing lower policies. Relative to initial positions, this implies less moderation by candidate A, and more moderation by candidate B. Consequently, B has to campaign harder in order to disperse the ambiguity surrounding his *ex post* policy implementation, whereas A can allocate more resources to valence campaigning as he ends up announcing a policy closer to his initial position. Once again, asymmetric returns to different campaign activities lead the candidate with a relatively more ignorant voter base to conduct a more policy focused campaign.

3 Conclusion

This chapter began by constructing a unidimensional probabilistic voting model featuring two purely office motivated candidates, a continuum of sincere voters with a fraction being ignorant, and statistical dependence between policy preferences and ignorance. This was motivated by the possible existence of factors that may coinfluence political awareness and policy preferences, e.g. education, which can be important especially in places with relatively low social mobility, or geography, which might matter particulary in countries with large regional disparities. Building on this framework, it then developed a model of electoral campaigning where candidates allocate their limited resources between two distinct types of campaign activities: policy promotion or valence campaigning. The former was assumed to allow candidates to announce and commit to policies unsupported by their legacies, the latter was assumed to generate universal electoral support via a partisanship effect. As an addendum, it considered an alternative formulation where candidates are free to announce any policy, and policy campaigning serves the purpose of partially addressing a commitment problem arising from

an announcement-reputation mismatch via ambiguity dispersal. Inspired by a persistent observation characterizing the last decade of Turkish campaign politics, the motivation was to assess the potential impact of an audience effect on campaign strategies, and to see whether a systematic relationship between policy preferences and policy awareness within an electorate can *ceteris paribus* lead to campaign divergence among different candidates.

Several conclusions were obtained. In the probabilistic voting model with no campaigning, politicians only take policy preferences of the informed/politically aware segment of the electorate into account when proposing platforms, and fraction of ignorant voters never matters by itself for policy proposals. These have two implications. First, if there is no systematic relation or statistical dependence between policy preferences and informational status, then the electoral outcome coincides with an egalitarian planner outcome where preferences of the entire electorate is taken into account, no matter how prevalent ignorance is. Second, if there is such dependence, then no matter how large as a group they are, policy preferences of ignorant voters are completely ignored and a tyranny of well-informed prevails.[60] As a corollary to the latter, a monotonic association between these two in a statistical sense implies a policy bias towards right or left depending on the context and the direction of association.[61] When policy commitment requires costly campaigning, which also provides an opportunity to capture the support of ignorant voters, neutrality of the size of ignorant population disappears. Even if the likelihood of lacking political awareness has no connection with policy preferences, a higher fraction of ignorant voters incentivizes both politicians to devote more resources to valence campaigning. This limits their ability to moderate their platforms and lead to policy divergence. In Lind and Rohner (2013), a similar conclusion appears due to intrinsic policy preferences of candidates. Here, it arises as a consequence of the interaction between legacies and policy-valence campaign trade-off giving rise to a similar centrifugal force. Differences in terms of responsiveness to policy campaigning among an electorate has also a more interesting effect in a campaign framework. In particular, when voter ignorance is correlated with policy preferences and candidates are positioned at opposing segments of the policy spectrum at the beginning of the campaign, an alternative channel for campaign divergence appears, leading otherwise symmetric candidates to follow asymmetric campaign strategies. For demonstration purposes, this chapter assumed positive correlation between igno-

---

[60]Of course this still requires a non-zero measure of informed voters.

[61]Lind and Rohner (2013) provide state-level evidence of a strong correlation between income and information leading to less redistributive platforms in U.S. House and Senate elections.

rance and preferences. Introduction of such positive correlation has a moderating effect on the left candidate's initial position and an extremizing effect on the right candidate's legacy. Consequently, the right candidate conducts a more policy oriented campaign in order to remoderate his position, whereas the left candidate can afford to conduct a campaign with a heavier valence focus. From a normative perspective, the model does not have much to say on efficiency aspects of campaigning under correlated ignorance, as all policies are efficient in a Pareto sense and potential issues about post election implications of campaign outcomes are swept aside by assuming credibility of campaign messages. Yet, it is still possible to unambiguously say that correlated ignorance reduces political representation of the less informed, and that possibility of valence campaigning worsens this underrepresentation.

There are several promising paths to follow from where this chapter has left off. First, the chapter has adopted a very simplistic view of political awareness by assuming that some people are completely in the dark regarding policies. One can bring the model a step closer to reality, for instance, by assuming varying signal strengths coupled with explicit belief formation. However, such an analytical detour would yield additional insights only if one had a solid framework picturing how the post election stage would play out. The latter would not only form a basis for Bayesian beliefs, but would also reveal potential sources of inefficiencies introduced by campaigns relying on partisanship effects, for instance, by limiting the screening ability of relatively better informed voters. Relatedly, channels through which campaigns operate were treated within a black-box approach and their effectiveness was taken as given. One can attempt to reconcile the correlated ignorance framework with a micro-founded campaigns framework. For instance, policy campaigning can be costly due to necessity of spending resources in order to transmit one's platform to the electorate, and it can also be informative due to party affiliations, which can make it costly for candidates to renege on their platforms post election, especially if parties have long horizon utilities.[62] This could also make valence cheap talk informative (as signalling a lack of commitment), which can provide rational incentives to uninformed voters at one end of the policy spectrum to vote for valence campaigning candidates at the opposing end of the spectrum. In a setting with correlated ignorance, this can imply cheap talking/ambiguous right versus policy campaigning/hard committing left or vice versa.[63] Second, one implication of the model presented in

---

[62]On this latter point, see Ashworth and Bueno de Mesquita (2008)

[63]This line of reasoning has its roots in the relatively small "strategic ambiguity" literature. See for instance Aragonès and Neeman (2000).

this chapter was that the candidate whose initial position is aligned with policy aware voters has a *ceteris paribus* competitive advantage. A legitimate question is then, why does any candidate, or any political party aligns itself with policy interests of ignorant voters in the first place? One possible answer is normative considerations by party founders. Another one can be constructed from a combination of changing demographics and institutional rigidity. A more interesting one, where correlated ignorance would also play an important role, considers costly party formation. Within the spirit of the current model, forming a party representing interests of informed voters can be costly, requiring a large coalition and time and resilience to establish credibility. Relative to this, one can argue that a right combination of idiosyncratic elements, e.g. charisma, oratorical skills etc. might be sufficient to form a loyalty base among informationally disadvantaged voters. Under correlated ignorance, this entry costs argument would also have implications of asymmetric political fractionalization. Finally, this chapter treated valence campaigning as a technology allowing politicians to generate expected universal support. One can instead assume that politicians have to direct such campaign efforts towards specific subgroups, with possible negative repercussions in the form of loss of support from other groups. Such subgroups can be thought as representing preexisting ethnolinguistic divisions or sociocultural fault lines. If the partition implied by such divisions roughly overlaps with the ignorant/informed segregation, correlated ignorance can generate asymmetries with respect to conciliatory/divisive campaign discourses. These issues are left for future research.

Appendix

Probabilistic Voting with Correlated Ignorance and Single-Crossing Voter Utilities:

Suppose that voters are indexed in some unidimensional type space $[\underline{\rho}, \overline{\rho}]$. Assume that their policy preferences are represented with the following value function.

$$U(X; \rho). \tag{88}$$

Further, suppose that $U(X; \rho)$ is submodular in $(X; \rho)$ and lower semi-continuous in $X \in [0, 1]$ for each ρ. One can imagine ρ representing gross incomes and X representing a linear income tax. From Topkis (1978), we know that $X^*(\rho)$ which

maximizes (88) should be decreasing in ρ due to submodularity. With (88), the objective function of candidate i becomes the following.

$$\mathbb{P}(S_i(\delta) > \frac{1}{2}) = \frac{1}{2} + \frac{1}{2\alpha}\left\{\tilde{\Gamma}(\Gamma, \frac{\beta_I}{\beta_U}) \int_{\underline{\rho}}^{\overline{\rho}}\left[U(X_i; \rho) - U(X_j; \rho)\right] \mathrm{d}F(\rho|I=1, \theta)\right\},$$

where θ orders $F(\rho|I = 1, \cdot)$ in terms of increasing first-order stochastic dominance, i.e. for any $\theta' > \theta$;

$$F(\rho|I=1, \theta') \leq F(\rho|I=1, \theta) \quad \forall \rho \in [\underline{\rho}, \overline{\rho}]. \tag{89}$$

In other words, an increase in θ corresponds to a distribution where informed people are more likely to have higher ρ's and thus lower policy preferences. Maximizing (89) will be equivalent to maximizing the following objective function.

$$\Phi(X_i, X_j; \theta) = \int_{\underline{\rho}}^{\overline{\rho}} U(X_i; \rho) \mathrm{d}F(\rho|I=1, \theta). \tag{90}$$

The game will have a pure-strategy equilibrium from Tarski (1955), as $\Phi(X_i, X_j; \theta)$ does not depend on X_j, so it is a submodular game, and lower semi-continuity is preserved under integration (Yannelis (1990)). Moreover, because $U(X; \rho)$ is submodular in $(X; \rho)$, and because $F(\rho|I=1, \cdot)$ is increasing in the first-order stochastic order, $\Phi(X_i, X_j; \theta)$ is submodular in (X_i, θ) as shown by theorem 14 in Van Zandt and Vives (2007). So the equilibrium policy platform $X_i^*(\theta) = X_j^*(\theta)$ will be decreasing in θ. $\qquad \square$

Proof of Lemma 3 (Monotonicity of $\Gamma(x)$ and LR-dominance):
Assume that $F(x|I=1)$ is LR-dominated by $F(x|I=0)$, i.e.

$$\frac{f(x|I=0)}{f(x|I=1)} \tag{91}$$

is increasing in x. Using Bayes' rule, (91) can be rewritten as follows.

$$\frac{f(x|I=0)}{f(x|I=1)} = \frac{\frac{(1-\Gamma(x))f(x)}{1-\hat{\Gamma}}}{\frac{\Gamma(x)f(x)}{\hat{\Gamma}}} = \frac{1-\Gamma(x)}{\Gamma(x)}\frac{\hat{\Gamma}}{1-\hat{\Gamma}}, \tag{92}$$

where $\hat{\Gamma}$ is the unconditional probability of being informed. So (91) is increasing if and only if

$$\frac{1-\Gamma(x)}{\Gamma(x)} \tag{93}$$

is increasing. If $\Gamma(x)$ is differentiable, this requires:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1-\Gamma(x)}{\Gamma(x)}\right) = -\frac{\Gamma'(x)}{(1-\Gamma(x))^2} \geq 0, \tag{94}$$

which is true if and only if $\Gamma'(x) \leq 0$. More generally, LR-dominance is equivalent to $\Gamma(x)$ being non-increasing, i.e. for any $x' > x$, $\Gamma(x') \leq \Gamma(x)$. ∎

Proof of Corollary 1 (Negative correlation between informational status and policy preferences):

Covariance between informational status and policy preferences can be expressed as follows.

$$\mathrm{cov}(x, I) = \int_0^1 \sum_{I \in \{0,1\}} \left(x - \mathbb{E}\left[x\right]\right)(I - \hat{\Gamma})\mathbb{P}(I|x)f(x)\mathrm{d}x, \tag{95}$$

where $P(I = 1|x) = \Gamma(x)$ is the conditional probability of being informed given policy preference x and $\hat{\Gamma} = \mathbb{E}[\Gamma(x)] = \mathbb{E}[I]$, which is also its unconditional probability. Simplifying the summation for a given x, one gets the following.

$$\sum_{I \in \{0,1\}} \left(x - \mathbb{E}\left[x\right]\right)(I - \hat{\Gamma})\mathbb{P}(I|x) = (x - \mathbb{E}(x))\left\{(1 - \hat{\Gamma})\Gamma(x) - (1 - \Gamma(x))\hat{\Gamma}\right\}$$

$$= (\Gamma(x) - \hat{\Gamma})(x - \mathbb{E}[x]). \tag{96}$$

It follows that:

$$\mathrm{cov}(x, I) = \int_0^1 (\Gamma(x) - \hat{\Gamma})(x - \mathbb{E}[x])f(x)\mathrm{d}x = \mathrm{cov}(x, \Gamma(x)). \tag{97}$$

Assumption 1 ($F(x|I = 1)$ is LR-dominated by $F(x|I = 0)$), was shown to be equivalent to $\Gamma(x)$ being monotone decreasing by lemma 3. As shown by Schmidt (2014), this implies $\mathrm{cov}(x, \Gamma(x)) \leq 0$. ∎

Proof of Lemma 4 (Stronger correlation):

Assumption 2 implies the following first-order stochastic dominance relations for any $\theta' > \theta$.

$$F(x|I = 1, \theta) \leq F(x|I = 1, \theta') \quad \forall x \in [0, 1], \tag{98}$$

$$F(x|I = 0, \theta) \geq F(x|I = 0, \theta') \quad \forall x \in [0, 1]. \tag{99}$$

Using Bayes' rule and assuming that total measure of ignorant voters remains fixed, covariance between x and I under θ and θ' respectively are given as follows

after opening up the summation in (95).

$$\text{cov}_\theta(x, I) = \hat{\Gamma}(1 - \hat{\Gamma}) \left\{ \int_0^1 (x - \mathbb{E}[x]) dF(x|I = 1, \theta) \right.$$

$$\left. - \int_0^1 (x - \mathbb{E}[x]) dF(x|I = 0, \theta) \right\}, \qquad (100)$$

$$\text{cov}_{\theta'}(x, I) = \hat{\Gamma}(1 - \hat{\Gamma}) \left\{ \int_0^1 (x - \mathbb{E}[x]) dF(x|I = 1, \theta') \right.$$

$$\left. - \int_0^1 (x - \mathbb{E}[x]) dF(x|I = 0, \theta') \right\}. \qquad (101)$$

Regarding the terms in the brackets, from (98) and (99), first term is bigger and second term is smaller in (100), as $(x - \mathbb{E}[x])$ is increasing in x. So $\text{cov}_{\theta'}(x, I) \leq \text{cov}_\theta(x, I) \leq 0$ follows. ∎

Proof of Proposition 5 (Comparative statics for the campaign game):
Start by rewriting the first-order sufficient condition for an interior optimum.

$$\left\{ -\int_0^{L_A + \rho_A P_A} \mathcal{D}'(L_A + \rho_A P_A - x) dF(x) \right.$$

$$\left. + \int_{L_A + \rho_A P_A}^1 \mathcal{D}'(x - (L_A + \rho_A P_A)) dF(x) \right\} - \frac{\mu_A}{\tilde{\Gamma} \rho_A} g'(R_A - P_A) = 0. \qquad (102)$$

Required conditions for the implicit function theorem are satisfied due to twice continuous differentiability of $\mathcal{D}$, g, strict convexity of $\mathcal{D}$ and strict concavity of g. Rewrite (102) as follows for sake of brevity.

$$\mathcal{Z}(P_A; L_A, \rho_A) - \frac{\mu_A}{\tilde{\Gamma} \rho_A} g'(R_A - P_A) = 0. \qquad (103)$$

1. $\left(\frac{\partial P_A}{\partial L_A} < 0 \right)$. Using the implicit function theorem, one gets the following (on an open neighbourhood centered at the vector that solves (103)).

$$\frac{\partial P_A}{\partial L_A} = -\frac{\overbrace{\dfrac{\partial \mathcal{Z}(P_A; L_A, \rho_A)}{\partial L_A}}^{-}}{\underbrace{\underbrace{\dfrac{\partial \mathcal{Z}(P_A; L_A, \rho_A)}{\partial P_A}}_{-} + \underbrace{\dfrac{\mu_A}{\tilde{\Gamma} \rho_A} g''(R_A - P_A)}_{-}}_{-}} < 0, \qquad (104)$$

where signs of the numerator and the first term of the denominator follow from the chain rule, strict convexity of $\mathcal{D}$, Leibniz rule and the fact that $\mathcal{D}'(0) = 0$. Sign of the second term of the denominator follows from strict concavity of g.

2. $\left(\frac{\partial P_A}{\partial \hat{\Gamma}} > 0\right)$. Using the implicit function theorem yields the following.

$$
\frac{\partial P_A}{\partial \hat{\Gamma}} = -\frac{\overbrace{(\tilde{\Gamma})^{-2}\dfrac{\mu_A}{\rho_A}\dfrac{\partial \tilde{\Gamma}}{\partial \hat{\Gamma}}g'(R_A - P_A)}^{+}}{\underbrace{\underbrace{\dfrac{\partial \mathcal{Z}(P_A; L_A, \rho_A)}{\partial P_A}}_{-} + \underbrace{\dfrac{\mu_A}{\tilde{\Gamma}\rho_A}g''(R_A - P_A)}_{-}}_{-}} > 0, \qquad (105)
$$

where the sign of the numerator follows from $g'(E - P_A) > 0$ and $\partial\tilde{\Gamma}/\partial\hat{r} > 0$.

3. $\left(\frac{\partial P_A}{\partial \tilde{\beta}} < 0\right)$. Using the implicit function theorem gives the following.

$$
\frac{\partial P_A}{\partial \tilde{\beta}} = -\frac{\overbrace{(\tilde{\Gamma})^{-2}\dfrac{\mu_A}{\rho_A}\dfrac{\partial \tilde{\Gamma}}{\partial \tilde{\beta}}g'(R_A - P_A)}^{-}}{\underbrace{\underbrace{\dfrac{\partial \mathcal{Z}(P_A; L_A, \rho_A)}{\partial P_A}}_{-} + \underbrace{\dfrac{\mu_A}{\tilde{\Gamma}\rho_A}g''(R_A - P_A)}_{-}}_{-}} < 0, \qquad (106)
$$

where the sign of the numerator follows from $\partial\tilde{\Gamma}/\partial\tilde{\beta} < 0$.

4. $\left(\frac{\partial P_A}{\partial \rho_A} \gtrless 0\right)$. Another application of the implicit function theorem gives the following.

$$
\frac{\partial P_A}{\partial \rho_A} = -\frac{\overbrace{\underbrace{\dfrac{\partial \mathcal{Z}(P_A; L_A, \rho_A)}{\partial \rho_A}}_{-} + \overbrace{(\rho_A)^{-2}\dfrac{\mu_A}{\tilde{\Gamma}}g'(R_A - P_A)}^{+}}^{?}}{\underbrace{\underbrace{\dfrac{\partial \mathcal{Z}(P_A; L_A, \rho_A)}{\partial P_A}}_{-} + \underbrace{\dfrac{\mu_A}{\tilde{\Gamma}\rho_A}g''(R_A - P_A)}_{-}}_{-}} \gtrless 0, \qquad (107)
$$

where the sign of the first term in the numerator follows from the chain rule, strict convexity of $\mathcal{D}$, Leibniz rule and the fact that $\mathcal{D}'(0) = 0$.

5. $\left(\frac{\partial P_A}{\partial \mu_A} < 0\right)$. Applying the implicit function theorem results in the following.

$$\frac{\partial P_A}{\partial \mu_A} = -\frac{\overbrace{-(\tilde{\Gamma}\rho_A)^{-1}g'(R_A - P_A)}^{-}}{\underbrace{\frac{\partial \mathcal{Z}(P_A; L_A, \rho_A)}{\partial P_A}}_{-} + \underbrace{\frac{\mu_A}{\tilde{\Gamma}\rho_A}g''(R_A - P_A)}_{-}} < 0. \qquad (108)$$

6. $\left(\frac{\partial P_A}{\partial R_A} > 0\right)$. Applying the implicit function theorem once more yields the following.

$$\frac{\partial P_A}{\partial R_A} = -\frac{\overbrace{-\frac{\mu_A}{\tilde{\Gamma}\rho_A}g''(R_A - P_A)}^{+}}{\underbrace{\underbrace{\frac{\partial \mathcal{Z}(P_A; L_A, \rho_A)}{\partial P_A}}_{-} + \underbrace{\frac{\mu_A}{\tilde{\Gamma}\rho_A}g''(R_A - P_A)}_{-}}_{-}} > 0. \qquad (109)$$

7. $\left(\frac{\partial V_A}{\partial R_A} > 0\right)$. Substituting for V_A in (103) and applying the implicit function theorem a final time yields the following.

$$\frac{\partial V_A}{\partial R_A} = -\frac{\overbrace{\frac{\partial \mathcal{Z}(R_A - V_A; L_A, \rho_A)}{\partial P_A}}^{-}}{\underbrace{\underbrace{-\frac{\partial \mathcal{Z}(R_A - V_A; L_A, \rho_A)}{\partial P_A}}_{-} - \underbrace{\frac{\mu_A}{\tilde{\Gamma}\rho_A}g''(V_A)}_{-}}_{+}} > 0, \qquad (110)$$

where the minus sign in front of the first term in the denominator follows from the chain rule. ∎

Proof of Proposition 7 (Asymmetric campaign monotonicity): As a reminder, assumption 2 is given below.

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{f_\theta(x|I = 1, \theta)}{f(x|I = 1, \theta)}\right] \leq 0, \qquad \forall F(\cdot|I = 1, \theta) \in \mathcal{F}. \qquad (111)$$

In addition, same marginal distributions for I implies that the fraction of ignorant voters is the same under any θ, so that $\tilde{\Gamma}$ is independent of θ. For any given θ,

assuming an interior solution, candidate B's first-order condition becomes the following.

$$\left\{ \int_0^{L_B - \rho_B P_B} \mathcal{D}'(L_B - \rho_B P_B - x) \mathrm{d}F(x|I=1,\theta) \right.$$
$$\left. - \int_{L_B - \rho_B P_B}^1 \mathcal{D}'(x - (L_B - \rho_B P_B)) \mathrm{d}F(x|I=1,\theta) \right\} - \frac{\mu_B}{\tilde{\Gamma}\rho_B} g'(R_B - P_B) = 0.$$
(112)

Let $Z_B(P_B, \theta)$ denote the left-hand side of (112). Since $\mathcal{D}$ is twice continuously differentiable and $F(\cdot|\cdot)$ is smooth, the implicit function theorem implies that (112) defines P_B as a function $P_B \equiv P_B(\theta)$ on an open neighbourhood around the given θ and that the following condition holds.

$$\frac{\mathrm{d}P_B(\theta)}{\mathrm{d}\theta} = -\frac{\partial Z_B(P_B,\theta)/\partial\theta}{\partial Z_B(P_B,\theta)/\partial P_B}.$$
(113)

Strict convexity of $\mathcal{D}$ and strict concavity of g implies $\partial Z_B(P_B,\theta)/\partial P_B < 0$. So $\mathrm{d}P_B(\theta)/\mathrm{d}\theta$ will have the same sign as $\partial Z_B(P_B,\theta)/\partial\theta$, which is in turn given as follows.

$$\left\{ \int_0^{L_B - \rho_B P_B} \mathcal{D}'(L_B - \rho_B P_B - x) \frac{f_\theta(x|I=1,\theta)}{f(x|I=1,\theta)} \mathrm{d}F(x|I=1,\theta) \right.$$
$$\left. - \int_{L_B - \rho_B P_B}^1 \mathcal{D}'(x - (L_B - \rho_B P_B)) \frac{f_\theta(x|I=1,\theta)}{f(x|I=1,\theta)} \mathrm{d}F(x|I=1,\theta) \right\}.$$
(114)

(114) is non-negative due to (111) and (112). It follows that $\frac{\partial P_B(\theta)}{\partial\theta} \geq 0$ and thus $\frac{\partial V_B(\theta)}{\partial\theta} \leq 0$. An analogous argument establishes $\frac{\partial P_A(\theta)}{\partial\theta} \leq 0$ and $\frac{\partial V_A(\theta)}{\partial\theta} \geq 0$. $\blacksquare$

Proof of Proposition 8 (Stronger correlation, policy proposals and campaign allocations):

First-order conditions for candidate A are given as follows.

$$X_A + \frac{1}{2}\nu_D'(X_A - L_A, R_A - V_A) - \mathbb{E}[x] - \frac{\mathrm{cov}\left(x, \Gamma(x)\right)}{\mathbb{E}[\Gamma(x)]} = 0, \qquad (115)$$

$$g'(V_A) + \tilde{\Gamma}\nu_P'(X_A - L_A, R_A - V_A) = 0, \qquad (116)$$

where $\nu_D'(\cdot)$ and $\nu_P'(\cdot)$ are short-hand notations for partial derivatives with respect to first and second arguments (distance and positive campaigning) respectively. Holding the total number of ignorant voters fixed while strengthening the correlation implies that $\tilde{\Gamma}$ and $\mathbb{E}[\Gamma(x)]$ stay constant while $\mathrm{cov}(x, \Gamma(x))$ becomes more

negative. Keeping these in mind, rewrite (115) and (116) as follows.

$$X_A + \frac{1}{2}\nu'_D - \mathbb{E}[x] + c = 0, \tag{117}$$

$$g'(V_A) + \tilde{\Gamma}\nu'_P = 0. \tag{118}$$

Reducing $\mathrm{cov}(x, \Gamma(x))$ (making it more negative) and increasing c will have same effects on X, up to a proportion. What is important is that differentiating with respect to c will represent the directional impact of a stronger negative correlation between being informed and policy preferences while holding $\mathbb{E}[\Gamma(x)]$ constant. Holding all parameters besides c constant and totally differentiating the above system yields the following.

$$\underbrace{\begin{bmatrix} 1 + \frac{1}{2}\nu''_{DD} & -\frac{1}{2}v''_{DP} \\ \tilde{\Gamma}\nu''_{DP} & g''(V_A) - \tilde{\Gamma}v''_{PP} \end{bmatrix}}_{\Omega_A} \begin{bmatrix} \mathrm{d}X_A \\ \mathrm{d}V_A \end{bmatrix} = \begin{bmatrix} -\mathrm{d}c \\ 0 \end{bmatrix}. \tag{119}$$

Calculating the determinant of Ω_A, one can see that it will be non-positive (I assume that it is strictly negative so that Ω_A is invertible).

$$\det(\Omega_A) = \underbrace{g''(V_A)(1 + \frac{1}{2}\nu''_{DD}) - \tilde{\Gamma}\nu''_{PP}}_{-} - \frac{1}{2}\tilde{\Gamma}\underbrace{(\nu''_{DD}\nu''_{PP} - (\nu''_{DP})^2)}_{+} < 0, \tag{120}$$

where the sign of the first term follows from strict concavity of g and convexity of ν, which implies $\nu''_{DD} \geq 0$ and $\nu''_{PP} \geq 0$. Sign of the last term follows from the diagonal condition for the convexity of ν. Using Cramer's rule yields:

$$\frac{\partial X_A}{\partial c} = \frac{\det\left(\begin{bmatrix} -1 & -\frac{1}{2}v''_{DP} \\ 0 & g''(V_A) - \tilde{\Gamma}v''_{PP} \end{bmatrix}\right)}{\det(\Omega_A)} < 0, \tag{121}$$

$$\frac{\partial V_A}{\partial c} = \frac{\det\left(\begin{bmatrix} 1 + \frac{1}{2}\nu''_{DD} & -1 \\ \tilde{\Gamma}\nu''_{DP} & 0 \end{bmatrix}\right)}{\det(\Omega_A)} \geq 0, \tag{122}$$

where the sign in (122) follows from the assumption that $\nu''_{DP} \leq 0$. Similarly, first-order conditions for candidate B can be rewritten as follows.

$$X_B - \frac{1}{2}\nu'_D(L_B - X_B, R_B - V_B) - \mathbb{E}[x] + c = 0, \tag{123}$$

$$g'(V_B) + \tilde{\Gamma}\nu'_P(L_B - X_B, R_B - V_B) = 0. \tag{124}$$

201

Totally differentiating in the same way and using Cramer's rule yields the following.

$$\frac{\partial X_B}{\partial c} = \frac{\det\left(\begin{bmatrix} -1 & \frac{1}{2}v''_{DP}+ \\ 0 & g''(V_B) - \tilde{\Gamma}v''_{PP} \end{bmatrix} \right)}{\det(\Omega_B)} < 0, \tag{125}$$

$$\frac{\partial V_B}{\partial c} = \frac{\det\left(\begin{bmatrix} 1 + \frac{1}{2}\nu''_{DD} & -1 \\ -\tilde{\Gamma}\nu''_{DP} & 0 \end{bmatrix} \right)}{\det(\Omega_B)} \leq 0, \tag{126}$$

where $\det(\Omega_B) < 0$. ∎

References

[1] **Ashworth, S. and E. B. de Mesquita (2008)**, "Informative party labels with electoral and institutional variation", *Journal of Theoretical Politics*, 20(3): 251-273.

[2] **Ashworth, S. and E. B. de Mesquita (2009)**, "Elections with platform and valence competition", *Games and Economic Behaviour*, 67(1): 191-216.

[3] **Aragonès, E., T. Palfrey and A. Postlewaite (2007)**, "Political reputations and campaign promises", *Journal of the European Economic Association*, 5(4): 846-884.

[4] **Aragonès, E. and Z. Neeman (2000)**, "Strategic ambiguity in electoral competition", *Journal of Theoretical Politics*, 12(2): 183-204.

[5] **Austen-Smith, D. (1987)**, "Interest groups, campaign contributions, and probabilstic voting", *Public Choice*, 54(2): 123-139.

[6] **Banks, J. S. (1990)**, "A model of electoral competition with incomplete information", *Journal of Economic Theory*, 50(2): 309-325.

[7] **Bardhan, P. K. and D. Mookherjee (2000)**, "Capture and governance at local and national levels", *American Economic Review*, 90(2): 135-139.

[8] **Baron, D. P. (1994)**, "Electoral competition with informed and uninformed voters", *American Political Science Review*, 88(1): 33-47.

[9] **Bartels, L. M. (2008)**, *Unequal Democracy: The Political Economy of the New Gilded Age*, Princeton University Press, New Jersey (US).

[10] **Berelson, B. R., P. F. Lazarsfeld and W. N. McPhee (1954)**, *Voting: A Study of Opinion Formation in a Presidential Campaign*, University of Chicago Press, Chicago (US).

[11] **Bernhardt, M. D. and D. E. Ingberman (1985)**, "Candidate reputations and the 'incumbency effect' ", *Journal of Public Economics*, 27(1): 47-67.

[12] **Brady, H. E. and P. M. Sniderman (1985)**, "Attitude attribution: a group basis for political reasoning", *American Political Science Review*, 79(4): 1061-1078.

[13] **Brennan, J. (2012)**, *The Ethics of Voting*, Princeton University Press, New Jersey (US).

[14] **Brueckner, J. K. and K. Lee (2013)**, "Negative campaigning in a probabilistic voting model", CESifo Working Paper Series, 4233, Munich (Germany).

[15] **Carrillo, J. and M. Castanheira (2008)**, "Information and strategic political polarisation", *Economic Journal*, 118(07), 845-874.

[16] **Chang, M. (1992)**, "Intertemporal product choice and its effect on collusive firm behaviour", *International Economic Review*, 33(4), 773-793.

[17] **Chappell, H. W., Jr. (1994)**, "Campaign advertising and political ambiguity", *Public Choice*, 79(3-4), 281-303.

[18] **Clarke, H. D., D. Sanders, M. C. Stewart and P. F. Whiteley (2009)**, *Performance Politics and the British Voter*, Cambridge University Press, Cambridge (UK).

[19] **Coate, S. (2004a)**, "Pareto-improving campaign finance policy", *American Economic Review*, 94(3), 628-655.

[20] **Coate, S. (2004b)**, "Political competition with campaign contributions and informative advertising", *Journal of the European Economic Association*, 2(5), 772-804.

[21] **Converse, P. E. (1964)**, "The nature of belief systems in mass publics", in: Apter, D. (ed.), *Ideology and Discontent*, Free Press, New York (US).

[22] **Cox, G. W. (1990)**, "Centripetal and centrifugal incentives in electoral systems", *American Journal of Political Science*, 34(4), 903-935.

[23] **Curini, L. (2014)**, "The conditional ideological inducement to campaign on character valence issues in multiparty systems: the case of corruption", *Comparative Political Studies*, 47(6), 1-25.

[24] **Davis, O. A., M. Hinich and P. C. Ordeshook (1970)**, "An expository development of a mathematical model of the electoral process", *American Political Science Review*, 64(2), 426-448.

[25] **Delli Carpini, M. X. and S. Keeter (1993)**, "Measuring political knowledge: putting first things first", *American Journal of Political Science*, 37(4), 1179-1206.

[26] **Delli Carpini, M. X. and S. Keeter (1996)**, *What Americans Know About Politics and Why It Matters*, Yale University Press, Connecticut (US).

[27] **Dinç, G. (2008)**, " 'Durmak yok, yola devam' - A close look at AKP's election success", Friedrich-Ebert-Stiftung Fokus Türkei, Nr. 6, Berlin (Germany).

[28] **Downs, A. (1957)**, *An Economic Theory of Democracy*, Harper, New York (US).

[29] **Enelow, J. M. and M. J. Hinich (1982)**, "Ideology, issues and the spatial theory of elections", *American Political Science Review*, 76(3), 493-501.

[30] **Eyster, E. and T. Kittsteiner**, "Party platforms in electoral competition with heterogenous constituencies", *Theoretical Economics*, 2(1), 41-70.

[31] **Grossman, G. M. and E. Helpman**, "Electoral competition and special interest politics", *Review of Economic Studies*, 63(2), 265-286.

[32] **Grossman, G. M. and C. Shapiro**, "Informative advertising with differentiated products", *Review of Economic Studies*, 51(1), 63-81.

[33] **Gul, F. and W. Pesendorfer**, "The war of information", *Review of Economic Studies*, 79(2), 707-734.

[34] **Hale, W. and E. Ozbudun (2009)**, *Islamism, Democracy and Liberalism in Turkey: The Case of the AKP*, Routledge, Oxford (UK).

[35] **Harrington, J. E., Jr. and G. D. Hess (1996)**, "A spatial theory of positive and negative campaigning", *Games and Economic Behaviour*, 17(2), 209-229.

[36] **Herrera, H., D. K. Levine and M. César (2008)**, "Policy platforms, campaign spending and voter participation", *Journal of Public Economics*, 92(3-4), 501-513.

[37] **Hinich, M. J. and M. C. Munger (1989)**, "Political investment, voter perceptions and candidate strategy: an equilibrium spatial analysis", in: Ordeshook, P. C. (ed.), *Models of Strategic Choice in Politics*, University of Michigan Press, Ann Arbor (US).

[38] **Hinich, M. J. and M. C. Munger (1993)**, "Political ideology, communication and community", in: Barnett, W. A., M. J. Hinich and N. J. Schofield (eds.), *Political Economy: Institutions, Competition and Representation*, Cambridge University Press, Cambridge (UK).

[39] **Hinich, M. J. and M. C. Munger (1996)**, *Ideology and the Theory of Political Choice*, University of Michigan Press, Ann Arbor (US).

[40] **Kartik, N. and R. Van Weelden (2014)**, "Informative cheap talk in elections", working paper, University of Chicago, Illinois (US).

[41] **Kinder, D. (1986)**, "Presidential character revisited", in: Lau, R. and D. Sears (eds.), *Political Cognition*, Lawrence Erlbaum, New Jersey (US).

[42] **Lau, R. and G. Pomper (2002)**, "Effectiveness of negative campaigning in U.S. senate elections", *American Journal of Political Science*, 46(1), 47-66.

[43] **Lau, R. and D. P. Redlawsk (2006)**, *How Voters Decide: Information Processing During Election Campaigns*, Cambridge University Press, New York (US).

[44] **Leiter, D. (2013)**, "The conditionality of voter decision-making: ideological proximity and valence in Western Europe", working paper, University of Missouri-Kansas City, Kansas (US).

[45] **Lenz, G. S. and C. Lawson (2011)**, "Looking the part: television leads informed citizens to vote based on candidates' appeerence", *American Journal of Political Science*, 55(3), 574-589.

[46] **Lind, J. T. and D. Rohner (2013)**, "Knowledge is power: a theory of information, income and welfare spending", Cahiers de Recherches Economiques du Département d'Econométrie et d'Economie Politique, Université de Lausanne, Lausanne (Switzerland).

[47] **Lindbeck, A. and J. Weibull (1987)**, "Balanced-budget redistribution as the outcome of political competition", *Public Choice*, 52(3), 273-297.

[48] **Lupia, A. and M. D. McCubins (1998)**, *The Democratic Dilemma: Can Citizens Learn What They Need to Know?*, Cambridge University Press, Cambridge (UK).

[49] **Maskin, E. and J. Riley (2000)**, "Asymmetric auctions", *Review of Economic Studies*, 67(3), 413-438.

[50] **Meirowitz, A. (2007)**, "Electoral contests, incumbency advantages and campaign finance", *Journal of Politics*, 70(3), 681-699.

[51] **Milgrom, P. R. (1981)**, "Good news and bad news: representation theorems and applications", *Bell Journal of Economics*, 12(2), 380-391.

[52] **Morton, R. and R. Myerson (2012)**, "Decisiveness of contributors' perceptions in elections", *Economic Theory*, 49(3), 571-590.

[53] **Negiz, N. and F. Akyıldız (2012)**, "The impact of image of the candidate on the voter's choice in local elections: an example of Uşak", *Journal of Süleyman Demirel University Institute of Social Sciences*, 15, 171-199.

[54] **Patty, J. W.**, "Equivalence of objectives in two candidate elections", *Public Choice*, 112(1-2), 151-166.

[55] **Popkin, S. L. (1994)**, *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns* (2nd ed.), University of Chicago Press, Illinois (US).

[56] **Plott, C. R. (1967)**, "A notion of equilibrium and its possibility under majority rule", *American Economic Review*, 57(4), 787-806.

[57] **Prat, A. (2002)**, "Campaign advertising and voter welfare", *Review of Economic Studies*, 69(4), 999-1017.

[58] **Riker, W. H. (1983)**, "Political theory and the art of heresthetics", in: Hinifter, A. W. (ed.), *Political Science: The State of the Discipline*, American Political Science Association, Washington DC (US).

[59] **Schmidt, K. D. (2014)**, "On inequalities for moments and the covariance of monotone functions", *Insurance: Mathematics and Economics*, 55(C), 91-95.

[60] **Schofield, N. (2003)**, "Valence competition in the spatial stochastic model", *Journal of Theoretical Politics*, 15(4), 371-383.

[61] **Serra, G. (2010)**, "Polarization of what? A model of elections with endogenous valence", *Journal of Politics*, 72(2), 426-437.

[62] **Shaked, M. and J. G. Shanthikumar (2007)**, *Stochastic Orders*, Springer, New York (US).

[63] **Snyder, J. M., Jr. (1989)**, "Election goals and the allocation of campaign resources", *Econometrica*, 57(3), 637-660.

[64] **Snyder, J. M., Jr. and M. M. Ting (2002)**, "An information rationale for political parties", *American Journal of Political Science*, 46(1), 90-110.

[65] **Stokes, D. E. (1963)**, "Spatial models of party competition", *American Political Science Review*, 57(2), 368-377.

[66] **Strömberg, D. (2004)**, "Mass media competition, political competition, and public policy", *Review of Economic Studies*, 71(1), 265-284.

[67] **Tarski, A. (1955)**, "A lattice-theoretical fixpoint theorem and its applications", *Pacific Journal of Mathematics*, 5(2), 285-309.

[68] **Todorov, A., A. N. Mandisodza, A. Goren and C. C. Hall (2005)**, "Inferences of competence from faces predict election outcomes", *Science*, 308(5728), 1623-1626.

[69] **TÜSİAD (2001)**, *Seçim sistemi ve siyasi partiler araştırması - Cilt 1: Bulgu ve Sonuçlar*, TÜSİAD Yayınları, Istanbul (Turkey).

[70] **Van Zandt, T. and X. Vives (2007)**, "Monotone equilibria in Bayesian games of strategic complementarities", *Journal of Economic Theory*, 134(1), 339-360.

[71] **West, D. M. (1993)**, *Air Wars: Television Advertising in Election Campaigns, 1952-1992*, CQ Press, Washington DC (US).

[72] **Wiseman, A. E. (2006)**, "A theory of partisan support and entry deterrence in electoral competition", *Journal of Theoretical Politics*, 18(2), 123-158.

[73] **Wittman, D. (1983)**, "Candidate motivation: a synthesis of alternative theories", *American Political Science Review*, 77(1), 142-157.

[74] **Wittman, D. (1989)**, "Why democracies produce efficient results", *Journal of Political Economy*, 97(6), 1394-1424.

[75] **Yannelis, N. C. (1990)**, "On the upper and lower semicontinuity of the Aumann integral", *Journal of Mathematical Economics*, 19(4), 373-389.

[76] **Yi, Z. and L. Tongyu (2004)**, "A further study on correlation order", *Applied Mathematics-A Journal of Chinese Universities*, 19(4), 429-434.

[77] **Zakharov, A. V. (2009)**, "A model of candidate location with endogenous valence", *Public Choice*, 138(3-4), 347-366.

[78] **Zaller, J. (1992)**, *The Nature and Origins of Mass Opinion*, Cambridge University Press, Cambridge (UK).

[79] **Zürcher, E. J. (2004)**, *Turkey: A Modern History* (3rd ed.), I.B. Tauris, London (UK).