unibz
Fakultät für Wirtschaftswissenschaften
Facoltà di Economia
Faculty of Economics and Management

BEMPS –
Bozen Economics & Management
Paper Series

NO 30 / 2015

Should Different People Have Different Governments?

Federico Boffa, Amedeo Piolatto, Giacomo Ponzetto

Should Different People Have Different Governments?[1]

Federico Boffa
Free University of Bolzano

Amedeo Piolatto, IEB,
Universitat de Barcelona

Giacomo A. M. Ponzetto
CREI, Universitat Pompeu Fabra, and Barcelona GSE

April 2015

[1]We are grateful for their helpful comments to Peter Backus, Christopher Berry, Paula Bustos, Vasco Carvalho, Steve Cicala, Ruben Enikolopov, Gino Gancia, Edward Glaeser, Matthew Kahn, Yuriy Kaniovskyi, Gianmarco León, Ana Nuevo-Chiquero, Luigi Pascali, Maria Petrova, Diego Puga, Thijs van Rens, Kurt Schmidheiny, Hannes Schwandt, Paolo Surico, Jaume Ventura, Yanos Zylberberg; and to seminar participants at CEMFI, Chicago Harris, Collegio Carlo Alberto, CREI, EIEF, Erasmus University Rotterdam, Florence, IESE, IMT Lucca, LUISS, Nottingham, Toulouse School of Economics, UPF, Vanderbilt, Yale, the Royal Economic Society 2013 Annual Conference, the 8th Meeting of the Urban Economic Association, the VATT Empirical Political Economics and Political Science Workshop, the 12th Annual Inernational Industrial Organization Conference, and the 2014 NBER Summer Institute Urban Economics Workshop. Piolatto acknowledges financial support from the Spanish Ministry of Science and Innovation (grant ECO-2012-37131) and the Generalitat de Catalunya (grant 2014SGR420). Boffa acknowledges financial support from the Free University of Bolzano (grants WW82). Ponzetto acknowledges financial support from the Spanish Ministry of Science and Innovation (grants ECO-2011-25624 and Juan de la Cierva JCI- 2010-08414), the Spanish Ministry of Economy and Competitiveness (Severo Ochoa Programme for Centres of Excellence in R&D SEV-2011-0075 and Ramón y Cajal grant RYC-2013-13838), the Barcelona GSE Research Network and the Generalitat de Catalunya (2009 SGR 1157), and the BBVA Foundation through its rst call for Grants to Researchers, Innovators and Cultural Creators. E-mail: fboffa@unibz.it, piolatto@ub.edu, gponzetto@crei.cat.

Abstract

The classic theory of fiscal federalism suggests that different people should have different governments. Yet, separate local governments with homogeneous constituents often end up doing poorly. This paper explains why and answers three questions: when regions are heterogeneous, what determines if power should be centralized or decentralized? How many levels of government should there be? How should state borders be drawn? We develop a model of political agency in which voters differ in their ability to monitor rent-seeking politicians. We find that rent extraction is a decreasing but convex function of the share of informed voters, because voter information improves monitoring but also reduces the appeal of holding office. As a consequence, information heterogeneity makes centralization appealing as a way of reducing rent extraction. Conversely, taste heterogeneity prompts decentralization as a way of matching local preferences. We also explain why the proliferation of government tiers harms efficiency. We find economies of scope in accountability: a single government in charge of many policies has better incentives than many special-purpose governments splitting its budget. Thus, a federal system is desirable only if information varies enough across regions. Our model implies that optimal borders should cluster by tastes but also ensure diversity of information. Quantitatively, our findings suggest excessive government fragmentation in the United States.

Keywords: Federalism, Government Accountability, Imperfect Information, Interregional Heterogeneity, Elections

JEL codes: D72, D82, H73, H77

1 Introduction

In the run-up to Scotland's 2014 independence referendum, the Scottish Government published a guide setting out its case for independence. Alex Salmond, the premier, argued that Scotland ought to become independent because its people are different from those of other parts of the British Isles and thus need a different government of their own. "After Scotland becomes independent ... the people of Scotland are in charge. It will no longer be possible for governments to be elected and pursue policies against the wishes of the Scottish people" (Salmond 2013, pp. x-xi).

The Scottish leader's argument finds support in the standard economic theory of fiscal federalism. Its core result is the Decentralization Theorem: absent policy spillovers, decentralization is more efficient than centralization if regions are not identical. This proposition, introduced by Oates (1972), has proved a remarkably general paradigm (Lockwood 2006). Local governments can tailor their choices to the particular conditions of each jurisdiction and thus provide higher social welfare than a single policy adopted by a common government. With no economies of scale, each group with distinct preferences should have an independent government (Tiebout 1956; Bewley 1981). Increasing returns and externalities promote political integration, but heterogeneity raises the downsides of large jurisdictions (Alesina and Spolaore 1997, 2003). Political-economy frictions provide rigorous microfoundations for the inability of a central government to match local preferences (Lockwood 2002; Besley and Coate 2003; Harstad 2007).

Nevertheless, a majority of Scottish voters rejected independence in the referendum. Their choice may well have been wise because, in reality, governments targeted to a specific slice of the population have not always delivered the benefits promised by theory. On the contrary, carving out an overly homogeneous constituency—particularly for a lower-status group—has often led to bad policies, mismanagement and distress (Glaeser and Shleifer 2005). Detroit under the twenty-year leadership of Coleman Young provides one cautionary tale. As the city became more homogeneously African-American, its voters supported Young's re-election by widening margins. Without the check of a diverse electorate, however, the mayor's policies and his confrontational politics contributed to Detroit's economic decline and enduring social problems. The first years of Slovakia's independence present some parallels at a national level. The country emerged in 1993 from a union whose federal leadership had long been dominated by the Czech. The politician who negotiated independence, Vladimir Mečiar, enjoyed widespread popular support and went on to lead Slovakia for most of the decade. However, he exploited his popularity to concentrate power in the hands of his party, and his administration was marked by inefficiency, clientelism, cronyism

and pervasive corruption. The Slovak economy almost collapsed in 1998. Governance has improved markedly since Mečiar was kicked out of office, but corruption remains a major concern.

In this paper, we develop a model of political agency that explains why separate governments for different groups can be a failure. Our model answers three key questions: when regions are heterogeneous, what determines if power should be centralized or decentralized? How many levels of government should there be? How should state borders be drawn? The core idea that underpins our theory is that heterogeneous regions do not differ only in preferences, the focus of the classic theory of fiscal federalism and of Salmond's plea for Scottish independence. Different groups also have different abilities to monitor elected officials and hold government accountable. Accordingly, political accountability varies substantially within the United States: in the most corrupt states, such as Louisiana and Mississippi, official corruption is five times as prevalent as in the least corrupt ones, such as Oregon and Washington (Glaeser and Saks 2006).

We study public goods provided by self-interested politicians whose goal is to extract wasteful rents. In order to keep extracting rents, however, they need to win re-election, so their rent-seeking is constrained by career concerns. Politicians have heterogeneous skills and elections work as a screening mechanism: voters re-elect skilled incumbents and dismiss unskilled ones. This electoral discipline generates incentives for the incumbent to perform: if he provides public goods instead of extracting rents, voters' inference of his ability improves and so do his chances of re-election.

Our model has several key features. First, we introduce heterogeneous accountability due to differences in voters' information. Some voters correctly observe and understand policy outcomes, while others do not and remain unable to infer the incumbent's ability. Second, we allow for a probabilistic component in voting. Voters do not aim only at selecting competent politicians, but they also have idiosyncratic political preferences independent of competence. Finally, we study a dynamic model with a recursive incentive structure. The expectation of future electoral discipline affects politicians' current trade-off between rent extraction and re-election.

In such a dynamic setting, a permanent increase in voter information has two effects on electoral discipline. First, it improves incentives for politicians to refrain from extracting rents in order to increase their likelihood of re-election. On the other hand, as equilibrium rent extraction declines, so does the value of holding office. The reduced appeal of re-election moderates the decline in rent extraction. In our model we find that the first effect always dominates, but the second effect entails that voter information reduces rent extraction at a declining rate. An improvement in monitoring is highly beneficial when the starting level is

low, so politicians react sharply because the value of office is high. Further improvements yield lower benefits.

Our key theoretical insight follows directly from the concave impact of an informed population on the quality of government. When different regions have different shares of informed voters, centralization reduces aggregate rent extraction. Political integration creates a single electorates with the average share of informed voters. Rent extraction decreases a lot in less-informed regions, while it only increases a little in better-informed ones.

However, we also find that the distribution of the efficiency gains from centralization is problematic. A centralized government is more accountable than the average decentralized government, but it is disproportionately accountable to the most informed regions. If politicians enjoy full discretion over the geographic distribution of public goods, they favor informed regions and neglect uninformed ones. The ensuing misallocation is so costly that centralization lowers social welfare despite reducing rents. As a consequence, our model implies that centralization must be accompanied by a uniformity constraint that requires at least some public goods to be provided identically in all regions.

Heterogeneous information thus drives an endogenous trade off at the core of our theory. Centralization improves accountability, but it foregoes the ability to match public goods to idiosyncratic preferences in different regions. Section 3 analyzes this trade off and answers our motivating question: should government be decentralized when regions are different? The answer depends on what type of heterogeneity is starkest across regions. Taste heterogeneity leads to decentralization; information heterogeneity leads instead to centralization.

Empirical evidence lends support to our results. Absent a uniformity constraint, politicians allocate public funds across regions in response to voter information rather than actual needs (Strömberg 2004). On the contrary, when the central government provides public goods uniformly centralization benefits mainly the uninformed: decentralizing reforms to public education in Argentina and Italy had regressive effects and worsened inequality (Galiani, Gertler and Schargrodsky 2008; Durante, Labartino and Perotti 2014).

Our prediction that centralization improves government accountability is consistent with American history. Two former state governors—Don Siegelman of Alabama and Rod Blagojevich of Illinois—are currently in prison for corruption. Corruption has been considered a distinctive plague of city and state governments (Steffens 1904; Wilson 1966). The patronage and political manipulation that had characterized state and local welfare programs were eradicated by federal intervention during the New Deal (Wallis 2000, 2006; Wallis, Fishback, and Kantor 2006). Although cross-country studies of decentralization and corruption have not reached robust conclusions (Treisman 2007; Fan, Lin, and Treisman 2009), world history also offers several examples of accountability gains from centralization: in transition

economies (Blanchard and Shleifer 2001), in pre-colonial Africa (Gennaioli and Rainer 2007a, b) and in early modern Europe (Besley and Persson 2011; Dincecco 2011).

European history also provides some direct evidence supporting our theoretical conclusion that heterogeneous accountability prompts centralization (Ziblatt 2006). Germany and Italy were both unified as nation-states in the late nineteenth century. Germany, which had relatively homogeneous institutional quality, was organized as a federal country. Italy, whose pre-unitary institutions were highly heterogeneous, became a centralized nation-state instead. Both regional differences in accountability and the degree of centralization are still higher in Italy than Germany today.

In Section 4 we study how many levels of government there should be. The standard logic of fiscal federalism suggests there should be many because every policy should be matched to the right geographic unit. In our framework, however, we find that multiplying government tiers is costly because there are economies of scope in accountability. When politicians are responsible for providing a larger set of public goods and control a proportionately larger budget, their incentives improve and they devote a lower share of the budget to rents. Thus, a unitary government is optimal if information is homogeneous. A federal system can be optimal only if differences in information are large enough. Then the federal government provides large accountability benefits to poorly informed regions, but it is also crucial that their local governments should retain control of policies for which preferences are very heterogeneous.

Our model thus provides a theoretical explanation for the empirical evidence that government quality declines as the number of government tiers rises. In the United States, Berry (2009) documents that the proliferation of overlapping special-purpose local governments in charge of specific policies has been a fiasco: special-purpose districts are inefficient and prone to capture by special interests. In Europe, too, multiple sub-national levels of governments have proven a source of inefficiencies, and their reduction and simplification is now on the agenda. Cross-country evidence shows a robust positive correlation between corruption and the number of levels of government (Fan, Lin, and Treisman 2009).

Section 5 considers what should determine the boundaries of governments when people are not naturally sorted into regions that are internally homogeneous. We find that optimal borders have two characteristics: they should cluster by tastes, but ensure maximum diversity of information. The second goal can trump the first when geographic constraints create a tension between the two. A disadvantaged uninformed group should not be left as a local minority; it should join better informed voters with similar preferences in a larger polity. Thus, our model suggests it would be detrimental to break up California. Social welfare needs educated San Francisco liberals to share a state government with working-class left-

wingers in the Central Valley.

In Section 6 we broaden our focus beyond utilitarian welfare maximization. Can centralization be beneficial for better informed regions? Our framework provides two routes to a positive answer. First, if there are interregional spillovers we find that the screening of politicians is better at the central than the local level. Second, we show that unanimous support for centralization can be obtained by imposing a partial uniformity constraint on the central government. Then the uninformed enjoy greater accountability in the provision of uniform public goods, and the uninformed greater influence over the provision of discretionary ones.

Finally, Section 7 develops a quantitative application of our model to state borders in the United States. Our theory suggests that excessive fragmentation may be harming the efficiency of American state governments. We calibrate the model using presidential vote shares as a proxy for preferences and the share of college graduates as a proxy for voter information. We find that several mergers of contiguous states would be welfare-enhancing, given that their residents differ considerably more in their human capital than in their political preferences.

This paper furthers the literature on fiscal federalism and the geographic structure of government. Starting with Tiebout's (1956) and Oates's (1972) seminal contributions, prior work focused exclusively on differences in preferences. We show that this is only one half of the story. Once we consider also differences in voter information across regions, we find that the two types of heterogeneity have opposite implications on the optimal government structure.

Differences in preferences promote decentralization if the central government cannot tailor policies to local preferences (Oates 1972; Alesina and Spolaore 1997, 2003; Alesina, Angeloni, and Etro 2005). Assuming that accountability is homogeneous across regions, the literature has endogenized the failure of preference-matching under centralization through frictions in political bargaining and the formation of minimum winning coalitions (Lockwood 2002; Besley and Coate 2003; Harstad 2007; Hindriks and Lockwood 2009). We provide a complementary microfoundation through heterogenous voter information.

More important, we show that differences in information promote centralization because they entail larger accountability gains from political integration. Our findings suggest that heterogeneous information is the key determinant of accountability gains from centralization. In our framework, political integration unambiguously alleviates the moral-hazard problem of political agency. On the contrary, previous studies have highlighted the potential for accountability gains from decentralization (Lockwood 2006). Decentralization can help voters monitor their local governments thanks to yardstick competition (Besley and Case 1995; Belleflamme and Hindriks 2005; Besley and Smart 2007), while centralization entails a

common-agency problem that makes politicians less accountable to voters in any single region (Seabright 1996; Persson and Tabellini 2000; Tommasi and Weinschelbaum 2007). Absent differences in voter information, potential sources of accountability gains from centralization have proved ambiguous. The common-agency problem might be counterbalanced by economies of scale in the exogenous "ego rents" from holding office (Seabright 1996; Persson and Tabellini 2000). Centralization might decrease, or conversely increase, the government's susceptibility to capture by special interest groups (Bardhan and Mookherjee 2000, 2006a, b; Blanchard and Shleifer 2001; Lockwood 2008). Hindriks and Lockwood (2009) highlight conflicting forces in a model of signaling: centralization unambiguously reduces voters' ability to screen and dismiss corrupt politicians; yet, it might also incentivize them to reduce their first-term rents in order to gain re-election and extract large rents in a second term.

Furthermore, we provide the first theory of economies of scope in government accountability. Prior studies have considered each policy instrument in isolation, typically assessing if it would be best centralized or decentralized (Oates 1999). Joanis (2014) microfounds this classic focus on the two extremes, showing that accountability declines if both central and local governments are simultaneously responsible for the same policy. In a dynamic setting, however, incentives for policy experimentation may be optimized if a policy choice is made by local governments first and then transferred to the central government (Kotsogiannis and Schwager 2006; Callander and Harstad 2015). We complement and extend this line of research by studying the pros and cons of having multiple levels of government when each is in charge of providing distinct public goods.

2 Political Agency and Public-Good Provision

In this section, we present the model of political agency that underpins our analysis of optimal political integration. Imperfectly informed voters face the problem of selecting and incentivizing self-interested rent-seeking politicians. We model electoral discipline in a framework of political career concerns (Persson and Tabellini 2000; Alesina and Tabellini 2008). Voters try to retain competent politicians and dismiss incompetent ones. In solving this screening problem, they endogenously create incentives for politicians to provide public goods. The incumbent moderates rent extraction because higher public-good provision raises voters' inference of his ability and thereby increases his chances of re-election.

2.1 Preferences and Technology

The economy is populated by a continuum of infinitely lived agents, whose preferences are separable over time and additive in utility from private consumption and public goods. Individual i in period t derives instantaneous utility

$$u_t^i = \tilde{u}_t^i + \sum_{p=1}^{P} \alpha_p^i \log g_{p,t}, \tag{1}$$

where $\tilde{u}_t^i$ is exogenous utility from private consumption, and $g_{p,t}$ the provision of public good p. We treat $\tilde{u}_t^i$ as an exogenous mean-zero shock and focus exclusively on public goods. The relative importance of each good for individual i is described by the ideal shares $\alpha_p^i \geq 0$ such that $\sum_{p=1}^{P} \alpha_p^i = 1$.

Each public good p is produced by the government with technology

$$g_{p,t} = e^{\eta_{p,t}} x_{p,t}. \tag{2}$$

The production technology has constant returns to scale: $x_{p,t}$ measures per-capita investment in providing public good p. We rule out economies of scale in public-good provision, which would provide an immediate technological rationale for centralization.

Productivity $\eta_{p,t}$ represents the stochastic competence of the incumbent politician in providing public good p. It follows a first-order moving-average process

$$\eta_{p,t} = \varepsilon_{p,t} + \varepsilon_{p,t-1}. \tag{3}$$

The shocks $\varepsilon_{p,t}$ are independent and identically distributed across goods, over time and across politicians. They have support $[\check{\varepsilon}, \hat{\varepsilon}]$, mean zero and variance σ^2. Our preferred interpretation is that parties are composed of overlapping generations of politicians. The period-t government consists of older party leaders with competence $\varepsilon_{p,t-1}$ and young party members with competence $\varepsilon_{p,t}$. At $t+1$, former party leaders retire, rising young politicians take over the leadership, and a new cohort joins the party.

Politicians are self-interested rent seekers. Their objective is to maximize the present value of the rents they can extract while in office, discounted by the discount factor $\delta \in (0, 1]$. Each period, the government allocates a fixed government budget b. The incumbent chooses the amount $x_{p,t}$ of expenditure on each public good. He extracts as rent the reminder

$$r_t = b - \sum_{p=1}^{P} x_{p,t}, \tag{4}$$

which represents public resources devoted to socially unproductive projects.[1]

2.2 Elections and Information

The incumbent faces reelection at the end of each period. If ousted he will never return to power. Politicians lack the ability to make credible policy commitments, so the election is not based on campaign promises, but on retrospective evaluation of the incumbent's track record. Voters do not observe directly the incumbent's competence nor his actions. Their inference is entirely based on an imperfect signal of public-good provision. The textbook model of career concerns assumes that voters observe policy outcomes with additive noise. We assume instead that voter information is binary. An informed voter observes perfectly the vector $\mathbf{g}_t$ of realized public goods. An uninformed voter receives no informative signal of $\mathbf{g}_t$, or proves completely incapable of understanding information about $\mathbf{g}_t$.[2]

The electorate consists of a continuum of atomistic voters, which can be partitioned into J groups. Group j comprises a fraction λ_j of voters. They have identical preferences described by the vector α^j of their ideal shares. The fraction of members of group j who are informed about public-good provision is a random variable Θ_t^j that is independent and identically distributed over time. Our model is robust to an arbitrary correlation of information across voters.[3] The expected share of informed voters θ_j provides our measure of voter information.

We allow for an intensive margin of political support, following the probabilistic voting approach (Lindbeck and Weibull 1987). Each voter's preferences consist of two independent elements. First, agents have preferences over the provision of public goods they expect from either politician (the incumbent I or the challenger C) in the following period. These

[1] Rent extraction could identically be interpreted as slacking (Seabright 1996; Persson and Tabellini 2000). Politicians enjoy an "ego rent" b from holding office. However, they incur a cost $x_{p,t}$ from exerting effort to provide public goods. Then r_t then captures politicians' failure to work diligently in their constituents' interest.

[2] Uninformed voters may not realize that public goods affect their utility. Such ignorance is particularly natural for public goods that yield long-run benefits. Voters may also understand the benefits of public goods, but fail to understand how they depend on the incumbent's actions and competence (Strömberg 2004).

[3] Most simply, information could be uncorrelated across voters. Each voter in group j has probability θ_j of being informed. Then in every period a share θ_j of group members are informed. This assumption is consistent with imperfect sharing of information within a group (Ponzetto 2011; Ponzetto and Troiano 2014). First, agents privately acquire information. Some fail to observe $\mathbf{g}_t$. Second, agents communicate with a finite number of neighbors. Some remain uninformed because none of their neighbors observed $\mathbf{g}_t$. If instead information sharing is perfect, information is perfectly correlated within each group. With probability θ_j the entire group is informed ($\Theta_t^j = 1$), and with probability $1 - \theta_j$ the entire group is uninformed ($\Theta_t^j = 0$).

preferences are summarized by the difference

$$\Delta^i \equiv \sum_{p=1}^{P} \alpha_p^i \mathbb{E}_i \left(\log g_{p,t+1}^I - \log g_{p,t+1}^C \right), \tag{5}$$

where $\mathbb{E}_i$ denotes the rational expectation given voter i's information. Second, voters have preferences for candidates' characteristics other than their competence: e.g., oratorical skill, personal likability, or party ideology. These preferences can be decomposed into an aggregate shock Ψ_t and an idiosyncratic shock ψ_t^i that is independent and identically distributed across voters.

Voting is costless and all voters cast a ballot for their preferred candidate. Thus, voter i votes for the incumbent if and only if $\Delta^i \geq \Psi_t + \psi_t^i$. As in Baron (1994) and Grossman and Helpman (1996), informed voters consider policy outcomes when deciding how to vote. Conversely, the behavior of uninformed voters is independent of public-good provision. They choose which candidate to support purely on the basis of preferences unrelated to competence.[4]

The distribution of the shocks Ψ_t and ψ_t^i is symmetric around zero, so voters do not favor systematically incumbents or challengers. For analytical tractability, we assume that the two shocks are uniformly distributed: $\Psi_t \sim U\left[-1/\left(2\phi\right), 1/\left(2\phi\right)\right]$ and $\psi_t^i \sim U\left[-\bar{\psi}, \bar{\psi}\right]$. The support of preference shocks is wide enough and the support of competence innovations $\varepsilon_{p,t}$ narrow enough that

$$\frac{1}{2\phi} - \bar{\psi} \leq \check{\varepsilon} < \hat{\varepsilon} \leq \bar{\psi} - \frac{1}{2\phi} \text{ and } -\frac{1}{2\phi} \leq \check{\varepsilon} < \hat{\varepsilon} \leq \frac{1}{2\phi}. \tag{6}$$

The first set of inequalities ensures that every voter's ballot is imperfectly predictable, irrespective of $\mathbf{g}_t$. The second set ensures that the outcome of the election is never entirely predictable either. The most capable incumbent has a non-zero chance of being dismissed and the least capable a non-zero chance of being re-elected.

The timeline within each period t is the following.

1. The incumbent politician's past competence shocks $\boldsymbol{\varepsilon}_{t-1}$ become common knowledge.

2. The incumbent chooses investments $\mathbf{x}_t$ and rent r_t.

[4] Unlike Baron (1994) and Grossman and Helpman (1996), we do not assume that only uninformed voters are impressionable. Informed voters, too, are also swayed by politician characteristics other than competence. The standard assumption of sincere voting is not inconsistent with strategic rationality because a continuum of voters entails strategic insignificance: no voter can ever be pivotal.

3. The competence shocks $\boldsymbol{\varepsilon}_t$ are realized and the provision of public goods $\mathbf{g}_t$ is determined.

4. Voter information is realized: a share Θ_t^j of members of group j perfectly observe $\mathbf{g}_t$. The rest remain completely uninformed. No voter has any direct observation of $\boldsymbol{\varepsilon}_t$, $\mathbf{x}_t$, or r_t.

5. An election is held, pitting the incumbent against a single challenger, randomly drawn from the same pool of potential office-holders.

2.3 Political Career Concerns

Voters rationally expect every politician to choose the stationary investment $\bar{\mathbf{x}}$. The equilibrium allocation is time-invariant because the environment is stationary. It does not vary with the incumbent's observed skills $\boldsymbol{\varepsilon}_{t-1}$ because performance is separable in effort and ability. It cannot vary with the competence innovations $\boldsymbol{\varepsilon}_t$ because they are unknown to the politicians themselves when they make policy choices.[5] Thus, the outcome of the election affects expected public-good provision only through differences in politicians' skills:

$$\Delta^i = \sum_{p=1}^{P} \alpha_p^i \mathbb{E}_i \left(\eta_{p,t+1} - \eta_{p,t+1}^C \right) = \sum_{p=1}^{P} \alpha_p^i \mathbb{E}_i \left(\varepsilon_{p,t} - \varepsilon_{p,t}^C \right) = \sum_{p=1}^{P} \alpha_p^i \mathbb{E}_i \varepsilon_{p,t}. \tag{7}$$

No information exists about future competence innovations (either the incumbent's $\boldsymbol{\varepsilon}_{t+1}$ or the challenger's $\boldsymbol{\varepsilon}_{t+1}^C$), nor about the challenger's current ability ($\boldsymbol{\varepsilon}_t^C$). Thus, their expectation is nil for all voters. Uninformed voters are also incapable of assessing the incumbent's ability, so they retain the unconditional expectation $\mathbb{E}\varepsilon_{p,t} = 0$. Informed voters, instead, can infer the incumbent's ability from their knowledge of public-good provision:

$$\mathbb{E} \left(\varepsilon_{p,t} | g_{p,t} \right) = \log g_{p,t} - \log \bar{x}_p - \varepsilon_{p,t-1}. \tag{8}$$

In a rational-expectation equilibrium their inference is perfectly accurate ($x_{p,t} = \bar{x}_p$ entails $\mathbb{E} \left(\varepsilon_{p,t} | g_{p,t} \right) = \varepsilon_{p,t}$).

[5] The agent's lack of private information is the defining technical feature of career-concern models (Holmström 1999). The assumption is natural when politicians are selected on the basis of differences in their ability rather than in their benevolence (Besley 2006). Moreover, in this setting the results of the career-concern model mirror those of a more complicated signaling model (Banks and Sundaram 1993, 1998). Politicians with different types are incentivized to undertake costly hidden actions so that voters should infer high competence. In equilibrium, voters correctly infer higher competence from better realized policy outcomes, and successfully screen more competent politicians.

From the politician's perspective, the probability of re-election as a function of his policy choices is

$$\pi\left(\mathbf{x}_t\right) = \frac{1}{2} + \phi \sum_{j=1}^{J} \theta_j \lambda_j \sum_{p=1}^{P} \alpha_p^j \left(\log x_{p,t} - \log \bar{x}_p\right), \tag{9}$$

as we derive in the Appendix. The incumbent faces a trade off. Investing in public goods reduces his rents but increases his chances of re-election by raising informed voters' inference of his ability. A politician who values re-election R chooses to extract rents

$$r = b - \phi R \sum_{j=1}^{J} \theta_j \lambda_j. \tag{10}$$

In a dynamic equilibrium, the value of re-election R is the expected present value of the future rents from holding office. In a rational-expectation equilibrium voters cannot be fooled ($\bar{x}_p = x_{p,t}$). Then in every election the incumbent wins with probability $\pi = 1/2$. Voter preferences are not exogenously biased in favor of incumbents or against them (the distribution of Ψ_t and ψ_t^i is symmetric around zero). An endogenous incumbency advantage does not arise because politicians' ability evolves as a first-order moving-average process. The impact of each competence shock lasts for two periods only, so past screening of incumbents does not translate in a forward-looking electoral advantage as it does with longer-lasting competence shocks (Banks and Sundaram 1993, 1998; Ashworth and Bueno de Mesquita 2008).[6] As a consequence, a politician who rationally anticipates extracting rent r whenever in office has an expected net present value of re-election

$$R = \delta \sum_{t=0}^{\infty} \left(\frac{\delta}{2}\right)^t r = \frac{2\delta}{2-\delta} r. \tag{11}$$

2.4 Government Accountability from Voter Information

Let $\rho \equiv r/b \in [0,1]$ denote the fraction of the budget allocated to rents in the stationary rational-expectation equilibrium.

[6] If the period-t incumbent was re-elected at $t-1$ the expectation of current productivity $\boldsymbol{\eta}_t$ is above average. Senior party leaders have proved their competence and won re-election. However, their known ability $\boldsymbol{\varepsilon}_{t-1}$ is orthogonal to future performance $\boldsymbol{\eta}_{t+1}$ because they are about to retire. A new cohort leades the party into the period-t election. Their skills $\boldsymbol{\varepsilon}_t$ can be inferred from policy outcomes $\mathbf{g}_t$, but not from the past re-election of their retiring colleagues.

Lemma 1 *In equilibrium, ruling politicians extract rents*

$$\rho = \left(1 + \frac{2\delta}{2-\delta}\phi \sum_{j=1}^{J} \theta_j \lambda_j\right)^{-1}$$

and have expected ability

$$\mathbb{E}\hat{\eta}_{p,t} = \phi\theta\sigma^2 \sum_{j=1}^{J} \alpha_p^j \theta_j \lambda_j.$$

Rent extraction is a decreasing and convex function of voter information ($\partial\rho/\partial\theta_j < 0$ and $\partial^2\rho/\partial\theta_j^2 > 0$). An increase in voter information θ_j increases the ability of ruling politicians $\hat{\eta}_{p,t}$ in the sense of first-order stochastic dominance.

Better information improves government accountability because it enables voters to monitor politicians more closely. It alleviates both the moral-hazard problem of politicians' incentives and the adverse-selection problem of politicians' selection. Voters can reward public-good provision only when they perceive it accurately. As voter knowledge improves, the incumbent's chances of re-election become more tightly linked to his performance. Ex ante, his career concerns are heightened so he extracts lower rents ($\partial\rho/\partial\theta_j < 0$). Ex post, skilled politicians are more likely to be re-elected and unskilled ones more likely to be replaced. Electoral screening improves and so does the average ability of ruling politicians ($\partial\mathbb{E}\hat{\eta}_{p,t}/\partial\theta_j > 0$).[7]

The key result in Lemma 1 is that rent extraction is decreasing but convex in voter information ($\partial^2\rho/\partial\theta_j^2 > 0$).[8] Decreasing returns to monitoring follow from the dynamic nature of the politicians' problem. The immediate impact of voter information on rent extraction is linear (equation 10). For a given value of re-election R, more informed voters induce one-to-one more investment and lower political rents. A transitory one-period increase in voter information would have no other effect.[9] In a dynamic setting, however, a permanent increase in voter information also has an indirect effect. Politicians understand they will

[7] Voters have no incentives to acquire information in order to improve governance because of the rational-voter paradox. Each voter has a negligible likelihood of determining the outcome of the election. His strategic incentives to become informed are likewise negligible. Therefore, information θ_j reflects exogenous voter characteristics. E.g., social capital reflects civic involvement and people's willingness both to acquire information through newspaper readership (Putnam 1993) and to share it in a wide social network (Ponzetto and Troiano 2014).

[8] Other determinants of the quality of government are straightforward. More patient politicians are more willing to reduce rent extraction in order to raise their chances of re-election ($\partial\rho/\partial\delta < 0$). A higher variance of politicians' ability raises the gains from screening ($\partial\mathbb{E}\hat{\eta}_{p,t}/\partial\sigma^2 > 0$). Both incentives and screening improve when voters are keener on competence than other determinants of political popularity ($\partial\rho/\partial\phi < 0$ and $\partial\mathbb{E}\hat{\eta}_{p,t}/\partial\phi > 0$).

[9] Accordingly, this is the only effect of greater voter information in a simplified one-shot model of career concerns (Persson and Tabellini 2000).

be monitored more closely if they are re-elected. Therefore, the expected future rents from holding office decrease. Their decline reduces the incentives to refrain from immediately extracting rents. The direct effect of improved monitoring is mitigated. Current rent extraction is more sensitive to the expectation of future rents when voters' average information is higher. Thus, a marginal increase in voters' information causes a lower decline in rent extraction when the average voter is more informed to begin with.[10]

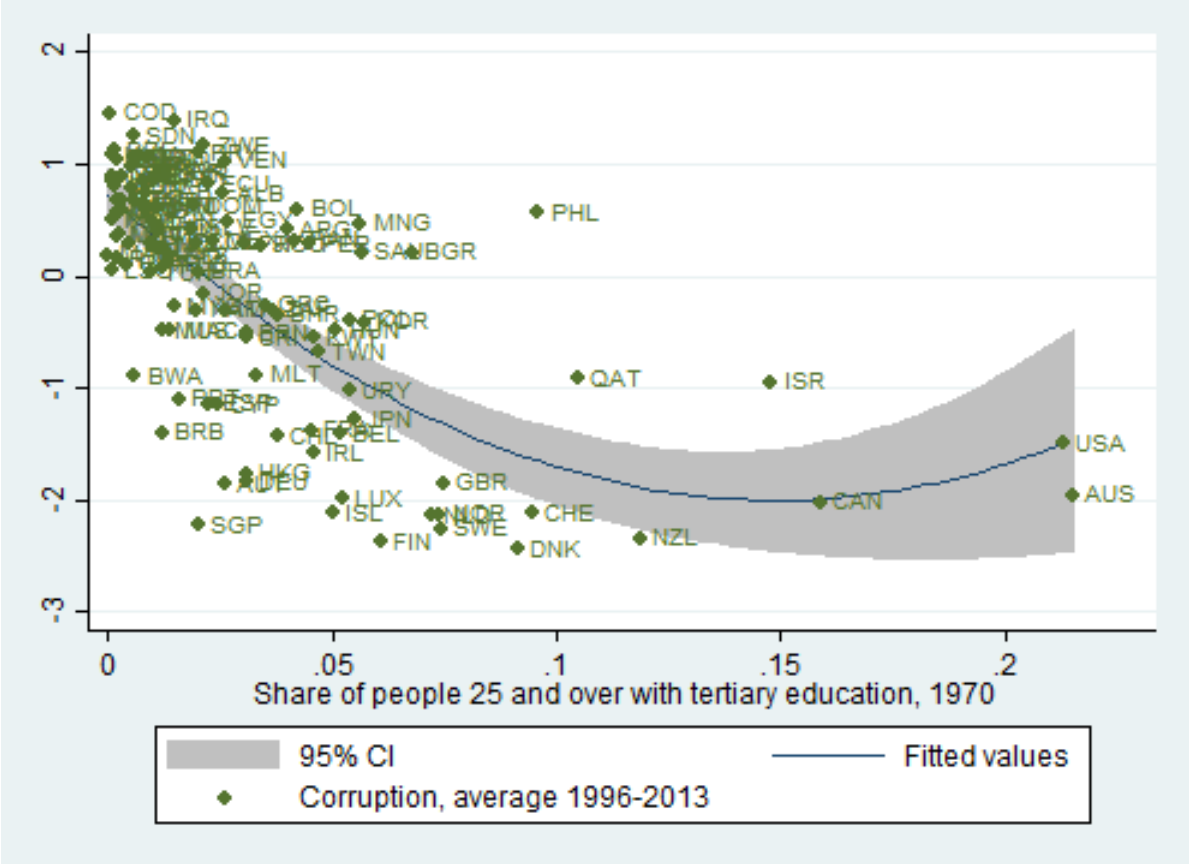
A large body of empirical evidence confirms that the quality of government is higher if citizens are more educated and politicians are subject to greater scrutiny by the media (Besley and Burgess 2002; Adserà, Boix, and Payne 2003; Glaeser et al. 2004; Svensson 2005; Glaeser and Saks 2006; Ferraz and Finan 2008; Snyder and Strömberg 2010). While none of these studies have explored specifically the concavity of this relationship, the data provide suggestive empirical support for our prediction. Svensson's (2005) documents that low human capital is the best predictor of high corruption across countries. Consistent with Lemma 1, Figure 1 shows that corruption is not only a decreasing but also a convex function of the share of people with a tertiary education. A similar relationship emerges in Figure 2, where we proxy information with newspaper circulation instead. Both results are robusts to controlling for income.[11]

Our finding that government accountability is an increasing but concave function of voter information has a broader theoretical underpinning than Lemma 1. The mechanism behind our result applies to any determinant of political discipline in a dynamic setting, including voter information but also, e.g., citizens' civic spirit (Nannicini et al. 2012). Information, however, has the distinctive feature that voters can share it. Such sharing is an additional source of concavity. We can interpret the fraction Θ_t^j of voters with full knowledge of policy outcomes $\mathbf{g}_t$ as the result of a two-stage process (Ponzetto 2011; Ponzetto and Troiano 2014). First, it includes those who acquired information directly, e.g., because they read

[10] Extreme cases highlight decreasing returns to monitoring with particular clarity. If no voters are aware of public-good provision, career concerns are absent and rent extraction is unchecked ($\theta = 0 \Rightarrow \rho = 1$). Introducing a little monitoring induces a forceful reaction by politicians who are afraid of losing very large rents. Conversely, if all voters perfectly observe public-good provision, career concerns are at their strongest but rent-extraction cannot be reduced to zero ($\theta = 1 \Rightarrow \rho > 0$). Incumbents always extract some rents because only the appeal of future rents induces them to make any productive investment. Marginally worsening perfect monitoring causes a small loss.

[11] The multivariate regressions are respectively $\rho_l = \underset{(.5)}{2.4} - \underset{(.06)}{.23} \ln y_l - \underset{(5)}{26}\theta_l + \underset{(27)}{82}\,\theta_l^2 + \varepsilon_l$ for education (across 118 countries) and $\rho_l = \underset{(.5)}{1.6} - \underset{(.07)}{.11} \ln y_l - \underset{(2)}{12}\theta_l + \underset{(3)}{13}\theta_l^2 + \varepsilon_l$ for newspaper circulation (across 100 countries). Corruption $\rho_l \in [-2.5, 2.5]$ is the opposite of the Control of Corruption index is from the World Governance Indicators (Kaufmann, Kraay and Mastruzzi 2010), averaging across available years (1996-2013). Real GDP per capita is from the Penn World Table 8.0 (Feenstra, Inklaar and Timmer 2013), measured in 1970 following Svensson (2005). The share of people over 25 with a tertiary education is from Barro and Lee's (2010) dataset version 2.0, also measured in 1970. Newspaper circulation per capita is from the World Development Indicators, averaging across available years (1997-2005).

Figure 1: Corruption and Education

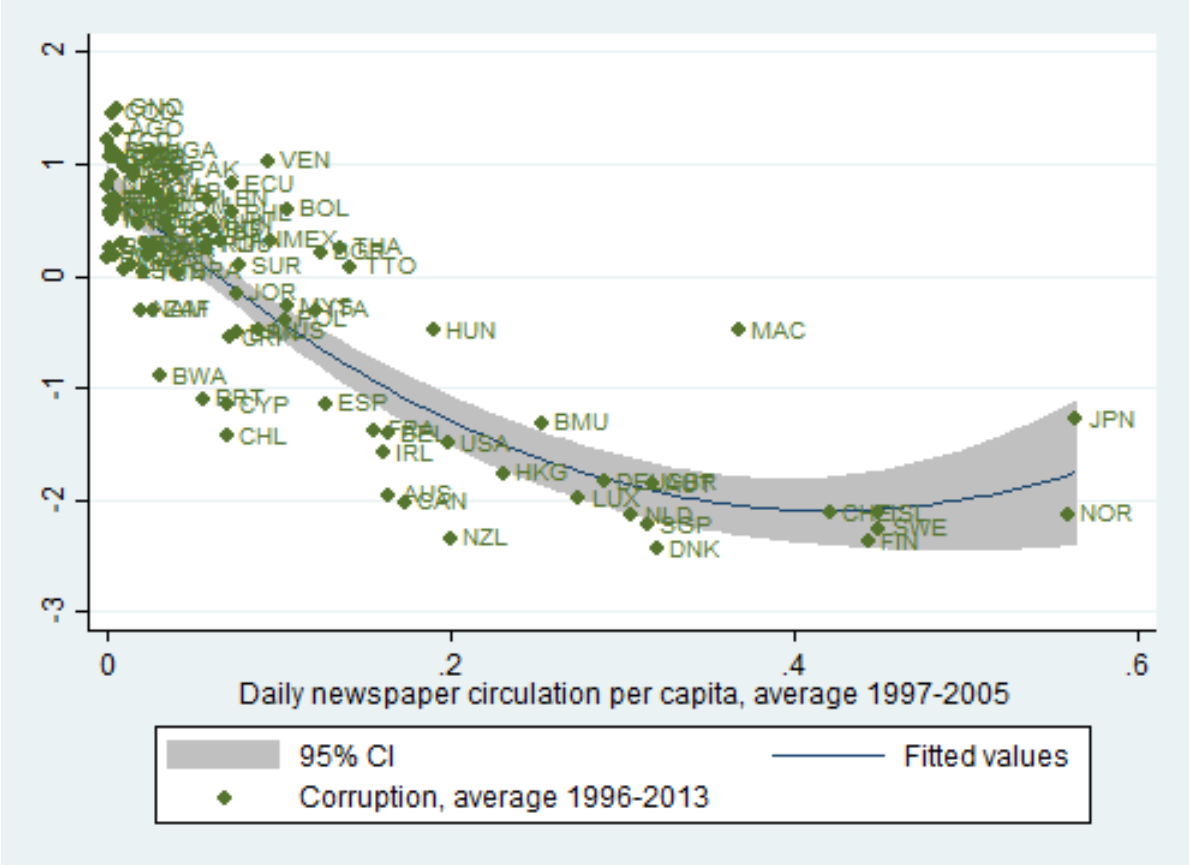


Notes: Corruption is the opposite of the Control of Corruption index is from the World Governance Indicators. The share of people over 25 with a BA degree in 1970 is from Barro and Lee (2010).

newspapers or because their education enables them to grasp the precise role of politicians in providing public goods. Second, it includes those who did not directly acquire information, but obtained it from an informed neighbor. Then the expected number θ_j of ultimately informed agents is an increasing and concave function of the probability that each agent directly acquires information because one voter's knowledge has greater spillovers if his neighbors are less informed.[12]

[12]If each agent obtains information directly with probability $\underline{\theta}_j$ and shares it in a group of n neighbors, his eventual probability of being informed is $\theta_j = 1 - \left(1 - \underline{\theta}_j\right)^n$ such that $\partial\theta_j/\partial\underline{\theta}_j > 0 > \partial^2\theta_j/\partial\underline{\theta}_j^2$.

Figure 2: Corruption and Newspaper Circulation



Notes: Corruption is the opposite of the Control of Corruption index is from the World Governance Indicators. Newspaper circulation per capita is from the World Development Indicators.

3 Should Government Be Decentralized?

We turn now to our motivating question. Should different regions have different governments whenever there are no spillovers, in accordance with Oates's (1972) classic Decentralization Theorem? When can we expect decentralization to deliver the benefits Salmond touted to Scotland's voters? When will a local government with a homogeneous population turn out instead to be a fiasco, as in Detroit under Young or Slovakia under Mečiar? Key to this question is that regions can differ along several dimensions. They can have different preferences but also different levels of voter information.

We consider an economy composed of L regions, each populated by a unit measure of voters. Preferences are homogeneous within each region, but heterogeneous across regions (Tiebout 1956; Oates 1972). E.g., conservative residents of "red states" may prefer greater

spending on defence, justice and police, while progressive residents of "blue states" may prefer instead environmental protection, public education, and welfare spending.[13] Formally, we assume that each region's preference vector $\boldsymbol{\alpha}^l$ is an independent draw from a common distribution that is symmetric across goods.[14] Symmetry entails that the marginal distribution α_p^l is the same for all p and has mean $\mathbb{E}\alpha_p^l = 1/P$. Differences in preferences across regions are summarized by the homogeneity parameter $\upsilon > 0$. In the limit as $\upsilon \to 0$, preferences are maximally heterogeneous: each region desires only one specific public good. Therefore, the probability that the same public good provides utility to two different regions is $1/P$, which is negligible when P is large. Conversely, in the limit as $\upsilon \to \infty$ preferences are perfectly homogeneous. Every region values all public goods identically, so all have the same ideal uniform basket ($\alpha_p^l = 1/P$ for all p and l). The distribution of preferences contracts smoothly as υ increases, in the sense that any decrease in υ entails a mean-preserving spread of α_p^l.[15]

Our novel contribution lies in studying at the same time differences in voter information. E.g., states with more educated residents are likely to have a greater share of voters with full knowledge of government performance, while voters in less educated states are less likely to succeed at inferring correctly incumbents' ability. Formally, we assume that each region's expected share of informed voters θ_l is an independent draw from a common distribution with mean $\mathbb{E}\theta_l = \bar{\theta} \in (0,1)$. Information is independent of preferences, and its variation across regions is summarized by the homogeneity parameter $\iota > 0$. In the limit as $\iota \to 0$, information is maximally heterogeneous. Each region is either always perfectly informed ($\theta_l = 1$, with probability $\bar{\theta}$) or always completely uninformed ($\theta_l = 0$, with probability $1-\bar{\theta}$). Conversely, in the limit as $\iota \to \infty$, information is perfectly homogeneous. Every region has the same expected share of informed voters $\theta_l = \bar{\theta}$. The distribution of information contracts smoothly as ι increases, in the sense that any decrease in ι entails a mean-preserving spread of θ_l.[16]

In a decentralized system, each region forms a separate constituency with a share of

[13] We focus on different preferences over the allocation of resources across public goods. Previous studies have typically neglected this dimension and focused instead of different preferences over the amount of public goods provided (Lockwood 2002, 2008; Besley and Coate 2003; Alesina, Angeloni, and Etro 2005; Harstad 2007; Tommasi and Weinschelbaum 2007). In reality, preferences vary on both dimensions (Alesina and Spolaore 1997). For simplicity, we consider only the allocation problem. This restriction preserves the equivalence between the allocation of the government budget and the allocation of effort by ruling politicians (Alesina and Tabellini 2008).

[14] We abstract from differences between the sample distribution and the population distribution by considering the limit case of a continuum of regions.

[15] These properties are satisfied if the preference vector $\boldsymbol{\alpha}^l$ has a symmetric Dirichlet distribution on the regular $(P-1)$-simplex with concentration parameter $\upsilon > 0$. Then $\mathrm{Var}\left(\alpha_p^l\right) = (P-1)/\left[P^2(1+\upsilon P)\right]$. None of our results relies on this particular specification.

[16] These properties are satisfied for instance if information has a beta distribution $\theta_l \sim B\left(\bar{\theta}\iota, \left(1-\bar{\theta}\right)\iota\right)$. Then $\mathrm{Var}\left(\theta_l\right) = \bar{\theta}\left(1-\bar{\theta}\right)/(1+\nu)$. None of our findings requires this specific distribution.

informed voters θ_l. It has an independent local government that allocates the regional budget b. Local politicians with skills $\eta^D_{l,p,t}$ invest in the provision of local public goods $x^D_{l,p,t}$ and extract rent $r^D_{l,t} = b - \sum_{p=1}^{P} x^D_{l,p,t}$.

Under centralization, instead, the central government is elected by a single unified constituency whose share of informed voters equals the average across regions $\sum_{l=1}^{L} \theta_l / L$. We rule out economies of scale: the central-government budget equals the sum bL of the regional budgets.[17] Central politicians with skills $\eta^C_{p,t}$ choose expenditures $x^C_{l,p,t}$ for each public good p in each region l and extract rent $r^C_t = bL - \sum_{l=1}^{L} \sum_{p=1}^{P} x^C_{l,p,t}$. The central government may be required to provide public goods uniformly across regions ($g^C_{l,p,t} = g^C_{p,t}$ for all l), either by a technological or by a constitutional constraint (Oates 1972; Alesina and Spolaore 1997). Conversely, it may be able to allocate spending across regions with complete discretion (Lockwood 2002; Besley and Coate 2003).

Different government structures admit the following ranking in terms of aggregate social welfare.

Proposition 1 *Aggregate social welfare is higher under decentralization than under centralization without a uniformity constraint. It is highest under centralization with a uniformity constraint if and only if preferences are sufficiently homogeneous ($\upsilon \geq \bar{\upsilon} > 0$). Centralization is more likely to be optimal when information is more heterogeneous ($\partial \bar{\upsilon} / \partial \iota > 0$) and politicians' ability less variable ($\partial \bar{\upsilon} / \partial \sigma > 0$).*

Centralization is unambiguously welfare-reducing if the central government can operate without a uniformity constraint that requires public goods to be provided identically in all regions. Office-seeking politicians target government spending disproportionately to the most politically influential regions. In our model, influence stems from information. Therefore, absent a uniformity constraint central-government spending in different regions is proportional to their voters' information:

$$\frac{\sum_{p=1}^{P} x^C_{l,p,t}}{\sum_{p=1}^{P} x^C_{m,p,t}} = \frac{\theta_l}{\theta_m} \text{ for all } l \text{ and } m. \tag{12}$$

This equilibrium allocation is a kind of harmful regressive redistribution. Independent local governments provide more public goods in more informed regions and instead extract larger rents in less informed ones. Centralization without uniformity increases public-good provision in regions with above-average information but reduces it in those with below-average

[17] If we interpret rent extraction as slacking, absence of scale economies requires politicians' ego rents to be proportional to the number of regions they control. Non-linear ego rents correspond to economies or diseconomies of scale and represent an exogenous driver of centralization or decentralization (Seabright 1996; Persson and Tabellini 2000).

information. As a consequence, aggregate social welfare declines relative to decentralization.

The key result in Proposition 1 is that, instead, centralization with a uniformity constraint is welfare maximizing when voter information varies substantially across regions, despite the absence of externalities or economies of scale. Political integration reduces aggregate rent extraction whenever voters in different regions have different information. By merging heterogeneous regions into a single polity, centralization leads to an overall level of voter information equal to the average of information across regions. Total rents decline because rent extraction is a decreasing and convex function of voter information, as established in Lemma 1:

$$\frac{1}{L}\sum_{l=1}^{L}\rho\left(\theta_l\right) \geq \rho\left(\frac{1}{L}\sum_{l=1}^{L}\theta_l\right). \tag{13}$$

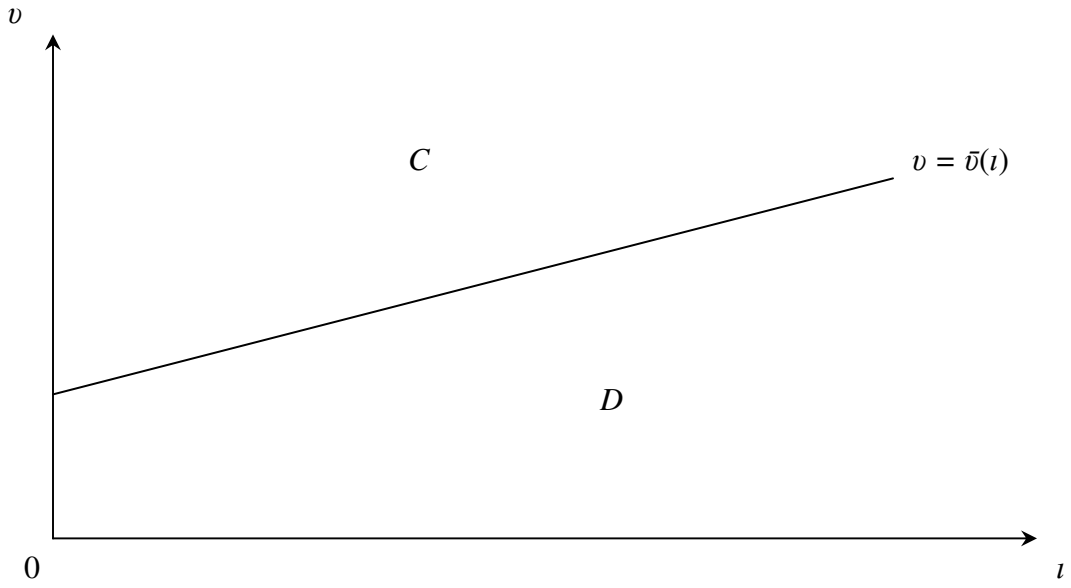
Not only does centralization entail an unambiguous decrease in rent extraction. With a uniformity constraint it is also a kind of beneficial progressive redistribution. The central government extracts slightly higher rents than local governments in regions with above-average information. It extracts much lower rents than local governments in regions with below-average information. The uninformed gain from integration because they can outsource their governance to politicians who are held accountable by better-informed voters in other regions. The informed conversely suffer from a dilution of their government accountability. However, not only do rents fall more in uninformed regions than they rise in informed ones. The marginal utility of public goods is also lower in the latter because the quality of their local government is higher.

If differences in voter information across regions are greater, so is the decline in rent extraction when the most informed regions monitor government performance for everyone. Heterogeneity in voter information is a centripetal force. The more regions differ in their government accountability, the greater the benefits of political integration. Conversely, heterogeneity in voter preferences is a centrifugal force. The central government must be constrained to provide public goods uniformly. Centralization thus sacrifices the ability to tailor local public goods to local preferences. The more regions differ in their ideal allocation, the greater the costs of political integration.

Figure 3 represents this trade off between greater preference-matching under decentralization and greater accountability under centralization is. The graph depicts the regions in which aggregate social welfare is maximized by centralization (C) or conversely by decentralization (D). The two dimensions of heterogeneity drive the trade off in opposite directions. When voters in different regions differ more in their information than in their preferences ($\upsilon \geq \bar{\upsilon}$), accountability gains dominate and centralization is optimal. If instead heterogeneity mostly concerns preferences, the cost of uniformity prevails and decentralization is

optimal ($\partial\bar{v}/\partial\iota > 0$).

Figure 3: Regional differences and optimal political integration



Centralization requires a uniformity constraint on the allocation of resources. It necessarily entails uniformity in government competence too. Under decentralization, each region selects—to the best of its imperfect screening ability—ruling politicians who are most talented at providing those public goods the region finds most important. The central government, instead, has average skills that try to satisfy all regions but truly fit none. When the variance of politicians' ability is greater, so is the cost of such uniformity. In Figure 3, the locus $\upsilon = \bar{\upsilon}$ shifts up ($\partial\bar{\upsilon}/\partial\sigma > 0$).

In Proposition 1, centralization influences the profile of politician's skills but has no impact on average screening across regions and public goods:

$$\sum_{p=1}^{P} \mathbb{E}\hat{\eta}_p^C = \phi\sigma^2\bar{\theta} = \frac{1}{L}\sum_{l=1}^{L}\sum_{p=1}^{P} \mathbb{E}\hat{\eta}_{l,p}^D. \tag{14}$$

This invariance result obtains because voter information about public goods (θ_l) is independent of the level of government that provides them. This assumption is realistic in so far as heterogeneous voter knowledge reflects differences in education, cognitive ability, social capital, or civic engagement. On the other hand, voter information also reflects differences in media coverage of policy outcomes, which plausibly varies with political integration. The media may be more likely to report on centralized policies because they concern a broader

audience (Gentzkow 2006; Snyder and Strömberg 2010). Such an increase in reporting would entail additional efficiency gains from centralization, through better selection as well as better incentives (Glaeser and Ponzetto 2014). Then, greater variance in politicians' ability might make political integration more appealing, rather than less.

Do the theoretical results in Proposition 1 have counterparts in the real world? We are certainly not able to render an empirical verdict on whether the European Union or an increasing national share of U.S. government spending is good or bad. We do, however, believe there is evidence supporting the key points in our model: discretionary spending by the central government can lead to short-changing less informed groups; decentralized control has often been associated with corruption and limited political accountability; the benefits of centralization often seem to be greater for less informed populations; and decentralization has been more successful where accountability is less heterogeneous across regions.

Strömberg (2004) studies the allocation of discretionary government spending during the New Deal and documents that state governors favored counties with a greater share of radio listeners, and so with better informed voters. If one accepts Strömberg's (2004) identification assumption that ground connectivity and woodland cover have no direct effect on the effectiveness of government expenditures, then it also follows that voter information alone is driving these differences in public spending across space. The tendency of discretionary spending to follow knowledge is precisely why Proposition 1 finds that discretion is bad.

The downsides of discretion may also explain why uniformity seems so common in many government policies. It may seem counterintertuitive that U.S. federal housing policy should offer similar subsidies to building in areas where supply is constrained, like New York City, and areas where supply seems almost unlimited, like Houston. One explanation for the spatial uniformity that seems like an unwritten element in U.S. budgets is that the tendency of locational discretion to harm particular regions is well understood.

The fundamental downside of decentralization in our model is that it leads to less accountability and more corruption. We know of no clear studies that illustrate the relative corruption of national versus local governments in the United States and Europe, but at the turn of the twentieth century American state and local governments were infamous for their corruption (Steffens 1904) and there is evidence suggesting that greater federal involvement with local government during the New Deal generally reduced local corruption (Wallis 2006).

One of the major themes of the U.S. progressive movement was the corruption that marked America's large urban and state governments at the turn of the century. The epitome of systematic corruption in local government was William Tweed, the boss of the formidable Tammany Hall machine in New York City (Ackerman 2005). The New York County Courthouse, better known as the Tweed Courthouse, became a veritable monument to or-

ganized graft. Its construction took over twenty years and cost $12 million, with overbilling of comical proportions. A Tammany ring member was paid $133,187 (around $2 million in present-day terms) for two days' work as a plasterer. Hardly less famous was the case of Chicago's street railways. The city council granted franchises on such favorable terms that in 1893 the entire system returned a mere $50,000 to the city. Instead, traction magnate Charles Yerkes spent $1 million in bribes to get through the state legislature an 1897 law enabling Chicago aldermen to grant franchises for no less than fifty years and without any compensation to the city (Junger 2010). This urban experience seems very far from Tiebout's (1956) and Oates's (1972) vision of local governments responding tightly to the desires of their residents.

The corruption and political manipulation that had characterized U.S. local politics were eradicated by federal intervention in the context of welfare spending. Until the Great Depression, poverty-relief programs managed by states and localities were bywords for patronage and graft. The New Deal—the most dramatic episode of centralization in the history of the United States—introduced strict federal oversight of welfare spending. One consequence was a striking decrease in corruption (Wallis 2000, 2006; Wallis, Fishback, and Kantor 2006).

While city politics cleaned up after the New Deal, state governments remained notorious for corruption (Wilson 1966). Since the Second World War, no less than ten governors and nine members of state executives have been convicted for official corruption and sentenced to jail. Two former governors—Don Siegelman of Alabama and Rod Blagojevich of Illinois—are currently serving sentences in federal prison. No member of the federal cabinet, let alone a president, has been charged with crimes investigated by the Department of Justice as part of the federal prosecution of public corruption.

International comparisons have yielded conflicting results on the relationship between decentralization and corruption. Treisman (2000) finds that a federal structure is associated with higher corruption, but Fisman and Gatti (2002) find that a larger sub-national share of government spending is associated with lower corruption. Both finding are not robust to changes in the sample or the addition of control variables (Treisman 2007). Consistent with our model, Fan, Lin and Treisman (2009) find that having more numerous and smaller local-government units is associated with more corruption. Overall, contemporary cross-country studies remain inconclusive.

Historical evidence from around the world, however, supports the view that political integration has a positive impact on government accountability. Centralized political institutions in precolonial Africa reduced corruption and fostered the rule of law. They caused a long-lasting increase in the provision of public goods that endured into the postcolonial period (Gennaioli and Rainer 2007a, b). Fiscal centralization was a key element in the moderniza-

tion of European states. It proved a necessary step for the consolidation of state capacity, which was in turn a critical determinant of economic and political development (Dincecco 2009, 2011; Besley and Persson 2011; Dincecco and Katz 2015; Gennaioli and Voth 2015). Blanchard and Shleifer (2001) argue that China grew faster than Russia in recent decades thanks to the greater strength of its central government vis à vis local politicians.

Proposition 1 predicts not only that centralization should reduce rent extraction, but that these accountability benefits should flow mostly to the least informed regions, as long as the central government enacts a uniform policy. Empirical evidence on reforms to public education systems bears out this prediction of our model. In the early 1990s, Argentina transferred control of federal secondary schools to provincial governments. This decentralization affected a third of existing public schools and half of all students in the public system. Five years after the reform, student test scores had risen in richer municipalities, but had failed to rise or even fallen in poor ones (Galiani, Gertler and Schargrodsky 2008). Decentralization increased inequality and harmed those already disadvantaged. Similarly, a 1998 university reform in Italy transferred responsibility for faculty hiring from the national ministry to individual universities. After this reform, faculty hires became significantly more nepotistic in provinces with low newspaper readership. Those with higher readership experienced at best a marginal improvement (Durante, Labartino, and Perotti 2014). Decentralization worsened the quality of academic recruitment and hurt the least informed regions the most.

Environmental policy also provides suggestive evidence supporting our theoretical prediction. In the United States, the Clean Air Act of 1970 transferred responsibility for pollution regulation from the state and local governments to the federal Environmental Protection Agency. Relative to pre-existing trends, pollutant emissions began to decline considerably faster in states with lower newspaper circulation (we provide a formal difference-in-differences analysis in Boffa, Piolatto and Ponzetto [2014]). In Europe, an EU directive introducing uniform standards for packaging waste "was less stringent than the existing German, Danish and Dutch laws, but was significantly stricter than the Greek, Irish and Portuguese requirements" (Fredriksson and Gaston 2000, p. 508).

The conclusion of Proposition 1 is that decentralization is desirable only if accountability is relatively homogeneous across regions. Our finding is consistent with historical evidence on the formation of unified nation-states in Germany and Italy. Both countries were unified in the second half of the nineteenth century: the Kingdom of Italy was established in 1861 and the German Empire in 1871. Before unification, Germany comprised many modern and well-functioning states. In Italy, the quality of pre-unitary institutions was lower and more heterogeneous. The Kingdom of Sardinia, which led the process of unification, could be considered the only efficient modern state. Consistent with our theory, Ziblatt (2006) argues

that precisely these different patterns of institutional development before unification explain why Germany was conceived as a federal nation-state and Italy as a unitary one. Remarkably, both the degree of centralization and the underlying heterogeneity in accountability have remained larger in Italy than in Germany up to the present day—excepting the tragic parenthesis of German centralization under Nazism.

4 How Many Levels of Government Should There Be?

The classic theory of fiscal federalism studies "which functions and instruments are best centralized and which are best placed in the sphere of decentralized levels of government" (Oates 1999, p. 1120). This standard approach suggests that there should be as many levels of government as there are geographic units a function is optimally tied to. Evidence from local governments in the United States, however, paints a different picture. Special-purpose districts managing individual public services for different and overlapping areas have performed poorly in terms of efficiency and accountability (Berry 2009). In this section, we explain why the proliferation of government tiers can harm welfare and we study when it is optimal to create a federal structure in which some policy decisions are centralized and other decentralized.

The distribution of voter information is the same as in Proposition 1, with mean $\bar{\theta}$ and a homogeneity parameter $\iota > 0$. However, we now consider two kinds of public goods at the opposite extremes of preference heterogeneity. First, there is a set of public goods for which all regions have perfectly homogeneous preferences ($\upsilon \to \infty$). By Proposition 1, these public goods would best be provided by a central government if there were no other policy choices. For the second set of public goods, preferences are completely idiosyncratic ($\upsilon \to 0$ and $P \to \infty$). Each region benefits exclusively from its own ideal variety, and derives no utility at all from any of the $L-1$ ideal varieties of the other regions. Absent other policies, Proposition 1 established that these idiosyncratic public goods should be provided by decentralized local governments. With both types of public goods, a resident i of region l has utility

$$u_t^i = \tilde{u}_t^i + \alpha_0 \log g_{l,0,t} + (1-\alpha_0) \log g_{l,l,t}, \tag{15}$$

where g_0 is a composite bundle of all the homogeneously desired public goods, while g_l is region l's desired variety of idiosyncratic public goods. The ideal share $\alpha_0 \in (0,1)$ provides a measure of preference homogeneity in this setting.

The structure of government is described by an allocation of powers and budgets to the two levels of government, local and central. As before, full decentralization means that each

local government provides the residents of its region l with both the homogeneously desired public goods ($g_{l,0}$) and their ideal variety of idiosyncratic public goods ($g_{l,l}$). Conversely, the government is fully centralized if the central government is tasked with providing all public goods to residents of all regions.

An intermediate possibility is the creation of a federal system. The central government provides homogeneously desired public goods ($g_{l,0}$) to all regions, while every region has its own local government provide the idiosyncratic public good $g_{l,l}$.[18] The overall budget remains exogenously fixed at Lb. Consistent with our focus on the expenditure side, we assume that all regions must contribute equally to the central-government budget. Its size b_C then suffices to characterize the budget allocation. Local-government budgets are determined residually as $b_D = b - b_C/L$ for every region.

The central government may be required to provide any public good uniformly. The uniformity constraint is imposed independently on each good. It may apply to some goods and not others. It may not, however, apply to an aggregate of goods. This restriction is immediate for a technological constraint because every good is distinct. The aggregate amount of public goods provided to a region ($\sum_{p=0}^{L} g_{l,p,t}$) cannot be constrained constitutionally either. The quantities of different goods cannot be properly compared by an impartial auditor, so it is unfeasible to require the provision of "separate but equal" public goods to different regions.

The welfare-maximizing structure of government admits the following characterization.

Proposition 2 *A federal system is optimal if differences in voter information are large enough ($\iota < \bar{\iota}$) while differences in preferences are neither too small nor too large ($\alpha_0 \in (\bar{\alpha}_{D\sim F}, \bar{\alpha}_{F\sim C})$). A federal system is more likely to be optimal when information is more heterogeneous ($\partial\bar{\alpha}_{D\sim F}/\partial\iota > 0$ and $\partial\bar{\alpha}_{F\sim C}/\partial\iota < 0$) and politicians' ability more variable ($\partial\bar{\iota}/\partial\sigma > 0$ and $\partial\bar{\alpha}_{F\sim C}/\partial\sigma > \partial\bar{\alpha}_{D\sim F}/\partial\sigma = 0$).*

Full centralization is optimal if differences in preferences are small ($\iota < \bar{\iota}$ and $\alpha \geq \bar{\alpha}_{F\sim C}$, or $\iota \geq \bar{\iota}$ and $\alpha \geq \bar{\alpha}_{D\sim C}$). Full decentralization is optimal if differences in preferences are large ($\iota < \bar{\iota}$ and $\alpha \leq \bar{\alpha}_{D\sim F}$, or $\iota \geq \bar{\iota}$ and $\alpha < \bar{\alpha}_{D\sim C}$). Full centralization is less likely to be optimal when politicians' ability is more variable ($\partial\bar{\alpha}_{D\sim C}/\partial\sigma > 0$).

Our model of accountability reverses the standard logic of fiscal federalism. The existence of some policy instruments that are best centralized and some others that are best decentralized does not immediately imply that the government should be structured on federal lines. On the contrary, if there are no differences in voter information across regions ($\iota \to \infty$) a

[18] A federal system with the opposite allocation of powers is theoretically possible but intuitively undesirable. We prove in the appendix that it can never be welfare-maximizing.

single level of government is unambiguously optimal: either a unitary central government, or independent unitary regional governments.

This result reflect endogenous economies of scope in government accountability. Politicians with little power also have low-powered incentives. Their skills have a lower impact on voters' utility, so other factors are more likely to determine re-election and their career concerns are weaker. In equilibrium, incumbents try to demonstrate each skill in proportion to its welfare value. E.g., they invest $x_0 = \alpha_0 \bar{\theta} \phi R$ if they have to provide the homogeneously desired good to voters with average information $\bar{\theta}$. The budget allocation would be independent of the division of powers if so were each politician's value of re-election R. Dividing government powers, however, requires dividing the public-sector budget, and the value of re-election is proportional to the budget a politician controls. Therefore, each government extracts a lower share of its budget as rents when it is responsible for providing a larger set of public goods.

Centralization minimizes aggregate rent extraction because it exploits both these economies of scope and the efficiency benefits of delegating government monitoring to the best monitors. As in Proposition 1, however, the central government fails to match idiosyncratic local needs. Under full centralization, each region unavoidably receives its ideal variety of idiosyncratic public goods in proportion to its residents' information:

$$\frac{x^C_{l,l,t}}{x^C_{m,m,t}} = \frac{\theta_l}{\theta_m} \text{ for all } l \text{ and } m. \tag{16}$$

The optimal provision of homogeneously desired public goods is uniform across regions, so a uniformity constraint suffices to ensure it. On the contrary, requiring uniform provision of idiosyncratic public goods only makes misallocation worse. The central government keeps catering disproportionately to the preferences of the informed, but it has to provide their ideal variety to other regions that derive no benefit from it. This uniformity constraint is so wasteful it makes every region worse off than discretionary central provision of idiosyncratic public goods.

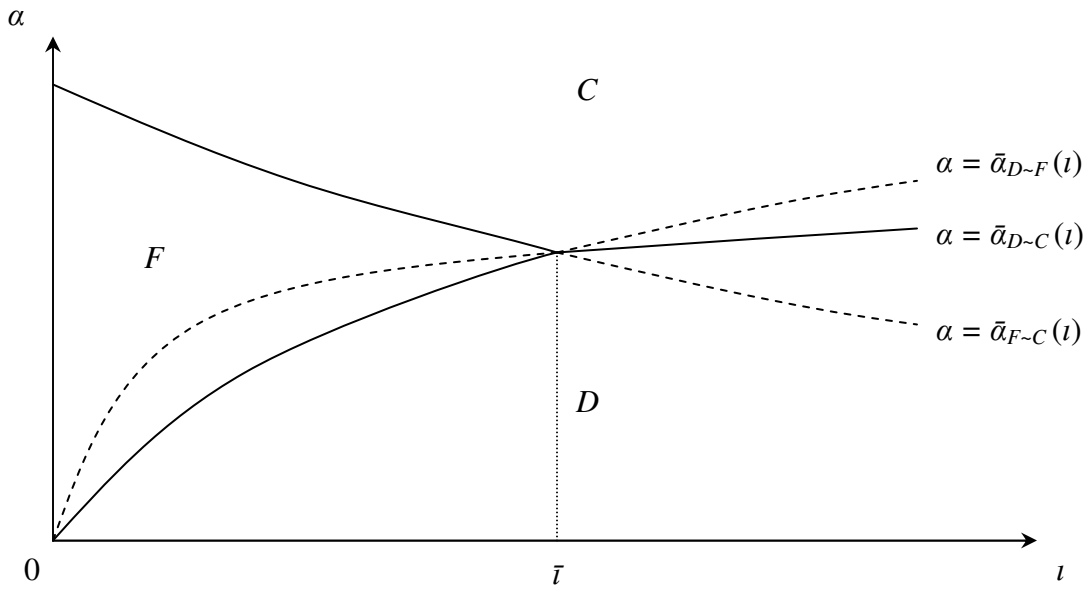
Preference heterogeneity then has a natural effect on the optimal structure of government. If preferences are highly idiosyncratic ($\alpha_0 \to 0$) decentralization is optimal because local governments are best at matching idiosyncratically preferences. If preferences are highly homogeneous ($\alpha_0 \to 1$), centralization is optimal because preference-matching is unimportant and only rent-minimization matters. In both extreme cases, one class of public goods is marginal, so it is worth sacrificing its optimal provision in order to exploit economies of scope and raise accountability in the provision of the dominant kind of public goods.

When preference heterogeneity is intermediate, both idiosyncratic and homogeneously

desired public goods are important. The key result in Proposition 2 is that a federal system is then optimal if and only if differences in voter information across regions are large enough. When the information gap is larger, uninformed regions gain more from delegating monitoring to informed ones. Hence, there are greater benefits from having a central government provide homogeneously desired public goods ($\partial\bar{\alpha}_{D\sim F}/\partial\iota > 0$). Greater heterogeneity also implies that uninformed regions lose more from ceding power to informed ones. Thus, there are greater costs of having the central government provide idiosyncratic public goods too ($\partial\bar{\alpha}_{F\sim C}/\partial\iota < 0$).

When differences in voter information are large, it is worth sacrificing economies of scope to reap the large benefits of a progressive transfer of accountability without paying the large costs of a regressive transfer of power. Figure 4 represents graphically the optimal structure of government. The larger the difference in information, the larger the region F in which a federal system is optimal.

Figure 4: Optimal federalism



As in Proposition 1, a downside of centralization is the uniformity of central politicians' skills. Thus, greater variation in the pool of political talent reduces the appeal of full centralization ($\partial\bar{\alpha}_{D\sim C}/\partial\sigma > 0$ and $\partial\bar{\alpha}_{F\sim C}/\partial\sigma > 0$). As a consequence, not only decentralization but also a federal system become more attractive ($\partial\bar{\iota}/\partial\sigma > 0$). In Figure 4, the continuous locus $\alpha = \max\{\bar{\alpha}_{F\sim C}, \bar{\alpha}_{D\sim C}\}$ shifts up and so does its intersection $\bar{\iota}$ with the locus $\alpha = \bar{\alpha}_{D\sim F}$.

Proposition 2 shows that multiple levels of government come at the cost of reduced government efficiency and accountability, even if they may be desirable for preference-matching and distributive reasons. The experience of local government in the United States bears out empirically our theoretical prediction. Both the number and the size of local governments have grown dramatically since World War II. Many states now have overlapping layers of county governments, municipal governments, and multiple special-purpose governments, such as elected school districts and independent districts managing specific public utilities. Yet, the performance record of special-purpose governments has been disappointing and they have proved prone to capture by special interests (Berry 2009). The employees of the special-purpose district are often the key voting block in its elections. Public libraries provide a telling example of systematic inefficiency. Berry (2009) analyzes empirically more than 8,000 public library systems from 1992 to 2004. Directly elected special-purpose library districts have larger budgets, but neither more visitors nor higher circulation. On the contrary, they hold fewer books and fewer of their employees are actually librarians.

Evidence from Europe confirms that the multiplication of government tiers has detrimental effects. In England, the most common structure of local government has two levels, with powers divided between counties and districts. A sizeable minority of areas are governed instead by a unitary authority entrusted with all local-government tasks. Unitary authorities are more efficient, particularly because the two-tier structure is associated to lower labor productivity and excess employment (Andrews and Boyne 2009).

France has a complex system that includes three nested tiers of sub-national governments (regions, departments and municipalities) as well as multiple kinds of aggregations of municipalities. This multiplicity of layers has proven a source of inefficiency and institutional weakness, especially at intermediate levels (Le Galès and John 1997; Seifert and Nieswand 2014). In its two latest reports on local government finances, the French Court of Auditors (2013, 2014) stresses that the proliferation of sub-national government tiers determines unproductive public employment. It also highlights inadequate governance mechanisms and advocates intervention by the national parliament to set directly goals and standards for local governments. Unsurprisingly, pruning of the local-government structure is on the French government agenda. The Attali Commission (2008) recommended abolishing the departmental tier within ten years as one of its twenty "fundamental decisions." President Hollande has proposed abolishing elected departmental councils by 2020.

Germany is also undergoing an analogous simplification. Three states (Rhineland-Palatinate, Saxony-Anhalt and Lower Saxony) have abolished one level of local government since 2000. Italy has abolished elected provincial councils in 2014, and the government has proposed a constitutional reform to abolish provinces altogether. The inefficiency of a three-tier

subnational structure (regions, provinces and municipalities) has been widely recognized. Indeed, Dente (1988) argues that it was designed specifically as a way for political parties to provide sinecures to their members and patronage to their supporters.

Cross-country evidence also supports the predictions of Proposition 2. In countries with more levels of government firms report having to pay more frequent and costlier bribes. The positive correlation between corruption and the number of government tiers is particularly robust. In developing countries, it is also extremely significant. Fan, Lin and Treisman (2009, p. 32) conclude that "[o]ther things equal, in a country with six tiers of government (such as Uganda) the probability that firms reported 'never' being expected to pay bribes was .32 lower than the same probability in a country with two tiers (such as Slovenia)."

While there is clear evidence that the multiplication of government tiers dilutes accountability, we know of no equally clear evidence of the distributive benefits of federalism. Nonetheless, the pattern of political discourse in the United States is suggestively consistent with our theoretical prediction that the least informed regions benefit the most from a federal structure relative to either unitary alternative. On average, Southern states have less educated voters and lower newspaper readership. They also have lower quality of government, as measured by official corruption (Glaeser and Saks 2006). The distributive predictions of our model can then help explain why the South is at the same time particularly patriotic—e.g., it provides a disproportionate share of U.S. military personnel—but also the region most supportive of appeals to curb the expansion of federal power and preserve the states' independent policy-making responsibilities.

When neither full centralization nor full decentralization is optimal, we can characterize the precise structure of the optimal federal system.

Corollary 1 *In the optimal federal system, the budget, productivity and accountability of the central government are lower when differences in preferences are larger ($\partial b_C^*/\partial\alpha_0 > 0$, $\partial\mathbb{E}\hat{\eta}_0^C/\partial\alpha_0 > 0$ and $\partial\rho^C/\partial\alpha_0 < 0$).*

The budget, productivity and accountability of local governments are higher when differences in preferences are larger ($\partial b_D^*/\partial\alpha_0 < 0$, $\partial\mathbb{E}\hat{\eta}_{l,l}^D/\partial\alpha_0 < 0$ and $\partial\rho_l^D/\partial\alpha_0 > 0$). Rent extraction by local governments increases with differences in information ($\partial\left(\sum_{l=1}^L \rho_l^D/L\right)/\partial\iota < 0$).

Overall rent extraction increases with differences in information. It is a concave function of preference heterogeneity and it reaches a maximum at the value $\check{\alpha}_0 \in (0, 1/2)$ for which local governments have on average the same accountability as the central government ($\alpha_0 = \check{\alpha}_0 \Leftrightarrow \rho^C = \sum_{l=1}^L \rho_l^D/L$). The difference in preferences associated with maximum rents increases with differences in information ($\partial\check{\alpha}_0/\partial\iota > 0$).

The comparative statics on each level of government highlight the fundamental strength of a federal system. Resources flow to the level of government where they are most useful. The efficient budget allocation reflects this logic most directly. All regions prefer the same optimal budget for the central government when they all contribute identically to it. The unique efficient allocation gives each level of government a budget proportional to the ideal share of the public good it is responsible for providing:

$$b_C^* = \alpha_0 bL \text{ and } b_D^* = (1-\alpha_0)\, b. \tag{17}$$

In equilibrium, voter monitoring of politicians obeys a similar allocation. Screening for competence is proportional to the welfare weight of the public goods each politician is in charge of providing:

$$\mathbb{E}\hat{\eta}_0^C = \alpha_0 \phi \sigma^2 \bar{\theta} \text{ and } \mathbb{E}\hat{\eta}_{l,l}^D = (1-\alpha_0)\, \phi \sigma^2 \theta_l. \tag{18}$$

Hence, incentives improve and rent extraction declines when a politician has more important responsibilities:

$$\rho^C = \left[1 + 2\alpha_0 \delta \left(2-\delta\right)^{-1} \phi \bar{\theta}\right]^{-1} \text{ and } \rho_l^D = \left[1 + 2\left(1-\alpha_0\right) \delta \left(2-\delta\right)^{-1} \phi \theta_l\right]^{-1}, \tag{19}$$

such that $\partial \rho^C / \partial \alpha_0 < 0 < \partial \rho_l^D / \partial \alpha_0$.

The impact of preference heterogeneity on aggregate rent extraction reflects instead the weakness of a federal system established by Proposition 2. For highly skewed values of α_0, one level of government accounts for most public-good provision. Then, it is both in charge of lion's share of the budget and exposed to substantial voter monitoring. This allocation implies low aggregate rent extraction because one level of government is large and accountable, while the other is relatively unaccountable but small. When this logic (and the value of α_0) is brought to an extreme, a federal structure becomes undesirable: the small and unaccountable level of government is best abolished, following Proposition 2. Proposition 1 reflects the second-best nature of the optimal government structure. Federalism is welfare-maximizing for intermediate values of α_0, which are those corresponding to the largest rents. Intuitively, rent extraction is highest when both levels of government are equally accountable ($\rho^C = \mathbb{E}\rho_l^D$). Then, if either grew more important it would at the same time control a larger budget and extract proportionally fewer rents from it.

Maximum rent extraction always occurs when the central government is smaller than the local ones ($\check{\alpha}_0 < 1/2$). This is a natural consequence of greater accountability at the central level in the presence of heterogeneous information. As differences in voter information grow larger, so does the inefficiency of local governments, and thus of a federal system that includes

them ($\partial \mathbb{E}\rho_l^D/\partial\iota < 0$ and $\partial\bar{\rho}^F/\partial\iota < 0$). Accordingly, the peak of rent extraction is associated with a greater importance of local governments ($\partial\check{\alpha}_0/\partial\iota > 0$).

5 What Should Determine the Boundaries of Governments?

Government structure is not entirely described by the number of tiers. The size of subnational jurisdictions can also vary. Is it better to have few large local governments or many small ones? Our model can be applied directly to study the optimal boundaries of governments. Proposition 1 considered for simplicity a symmetric setting in which it is optimal either to integrate all regions or to let each have its independent government. The intuition, however, generalizes to the asymmetric case. Regional boundaries should be drawn so that people with similar preferences but different information share a government, while those with different preferences but similar information do not. In Section 7 below we consider from this perspective the potential for state mergers in the United States.

In this section, we extend our model by relaxing the standard assumption that voters are exogenously sorted into geographic regions with internally homogeneous preferences. To study optimal boundaries when ideological groups do not naturally coincide with geographic regions, we assume a simple two-fold partition of voters by ideology and information.

Voters have ideological preferences for two distinct public goods L and R. Left-wingers desire the former and have utility $u_{L,t}^i = \tilde{u}_t^i + \log g_{l,L,t}$. Right-wingers desire the latter and have utility $u_{R,t}^i = \tilde{u}_t^i + \log g_{l,R,t}$. This simple preference structure provides a stylized model of local government consistent with Proposition 2. Preferences over locally provided public goods are highly heterogeneous because public goods that all voters desire homogeneously should instead be provided by the federal government.

Each ideological group comprises voters with different levels of information. Better informed voters succeed at inferring the incumbent's competence from realized policy outcomes with probability θ_I. Relatively uninformed voters have a lower probability of learning $\theta_U < \theta_I$.

A country is then characterized by the sizes of the four groups $\lambda_{L,I}$, $\lambda_{L,U}$, $\lambda_{R,I}$ and $\lambda_{R,U}$. We consider partitions of this overall population into autonomous regions or federal states. Each region is endowed with a budget of b units per resident, so there are no economies of scale. Moreover, a region is the minimal administrative unit, so the regional government is subject to a technological uniformity constraint: it cannot differentiate the provision of public goods across residents.

We begin by characterizing the optimal regional structure when there are no constraints on how citizens can be partitioned into regions.

Proposition 3 *Optimal regions are perfectly separated by preferences and perfectly mixed by information (every region l has either $\lambda_{l,L,I} = \lambda_{l,L,U} = 0$ and $\lambda_{l,R,I}/\lambda_{l,R,U} = \lambda_{R,I}/\lambda_{R,U}$, or $\lambda_{l,R,I} = \lambda_{l,R,U} = 0$ and $\lambda_{l,L,I}/\lambda_{l,L,U} = \lambda_{L,I}/\lambda_{L,U}$).*

In the absence of exogenous constraints, the optimal partition resolves intuitively the two forces highlighted by Proposition 1. Preference heterogeneity is a centrifugal force that can be accommodated by separating groups with different ideal allocations. Such optimal segregation reflects Tiebout's (1956) classic intuition. It is typically optimal when there are no economies of scale and no constraints on creating as many regions as there are desired bundles of public goods (Bewley 1981). The novelty of our model lies in the centripetal force caused by differences in information. A partition that achieves homogeneous preferences within each region can nonetheless be highly suboptimal. Optimality also requires the perfect mixing of like-minded voters with different levels of information. Citizens suffer from sharing a government with others with opposite preferences who cause a distributional conflict. They suffer no less from being cut off from better-informed voters with the same preferences, whose influence is necessary to keep the local government accountable.

Proposition 3 highlights that an ideologically homogeneous but uniformly uninformed region is plagued by bad governance. Its government reflects the preferences of local residents, but it is also unaccountable, inefficient and corrupt. This prediction of our model is consistent with evidence from local governments in the United States. City politicians have at times succeeded in creating large local majorities of their poorer and less educated supporters by encouraging the out-migration of a rival higher-status group. The detrimental consequences of his process are best illustrated by the long career of Boston mayor James Michael Curley (Glaeser and Shleifer 2005). Both his policies and his stark rhetoric championed the poor Irish community against the richer Anglo-Saxon Protestants that had previously dominated the city. The end of Brahmin dominance pleased Boston's Irish and removed the discrimination they had suffered from. However, Curley's administration was inefficient and corrupt; Boston declined under his government. Similar patterns emerge in other cases of populist local politics catering to particular ethnic and socioeconomic constituencies, such as African-Americans in Detroit under Coleman Young.

The optimal partition described by Proposition 3 has two contrasting features. Tension between the two can entail a welfare loss when groups with different preferences are separated. Proposition 1 characterized one set of circumstances leading to this outcome. When voters' preferences are not completely distinct, separation is undesirable if differences in voter

information are large enough.

Another possibility is that perfect separation à la Tiebout is technologically impossible because residents with different preferences are mixed in a narrow area such as a city or a county. In reality, most Americans live in a county that includes a substantial share of supporters of either party (Glaeser and Ward 2006). If perfect separation is impossible, is partial separation desirable, or is it even worse than perfect integration?

Consider two symmetric atomistic locations. Their total population is identical, but the first location has a majority of left-wing residents and the second a majority of right-wing residents. The distribution of the population is characterized by a degree of ideological sorting $\tau \in (0,1)$ such that

$$\lambda_{1,L} = \lambda_{2,R} = \frac{1+\tau}{4} \text{ and } \lambda_{1,R} = \lambda_{2,L} = \frac{1-\tau}{4}. \tag{20}$$

In the limit as $\tau \to 0$ residents with different preferences are perfectly mixed, while in the limit as $\tau \to 1$ there is perfect sorting.

Voter information is also symmetric, but not homogeneous across locations. Voters with either preferences have an average probability θ of learning from realized policy outcomes in the location in which they belong to the majority. In the location where they are a minority, their learning probability is reduced to $\theta(1-\zeta)$ for a coefficient $\zeta \in (0,1)$ of information disadvantage. The lower information of the minority reflects, in particular, endogenous media slant (Gentzkow and Shapiro 2010). Local media in each location choose an ideological bias that matches the preferences of the local majority. As a consequence, news consumption becomes more appealing for the majority, and less for the minority.

The following result characterizes formally whether political integration or partial separation is optimal when perfect segregation by preferences is impossible.

Proposition 4 *Aggregate social welfare is higher under political integration than under separation if minorities suffer from a high information disadvantage ($\zeta \geq \bar{\zeta} > 0$). Integration is more likely to be optimal when ideological sorting is less complete ($\partial\bar{\zeta}/\partial\tau > 0$) and politicians' ability less variable ($\partial\bar{\zeta}/\partial\sigma > 0$).*

Intra-regional heterogeneity entails a new trade off, related to but distinct from the one presented by Proposition 1. The novel centripetal force is a different kind of information heterogeneity. In Proposition 4 there are no differences in average information across regions. Accordingly, aggregate rent extraction is invariant. There are, however, differences in information between the majority and the minority within each location. Under separation, uninformed minorities are dominated by better informed local majorities. Political integra-

tion restores even power to the two ideological groups. Each uninformed minority gains political influence thanks to the like-minded informed majority in the other location. Thus, political integration can raise welfare even if the efficiency gains from delegated monitoring are absent.

These distributive welfare gains are monotone increasing in the information disadvantage of the minority. If information is homogeneous, separation is the constrained optimum ($\bar{\zeta} > 0$). Imperfect ideological segregation remains costly, and minorities naturally bear a greater share of this cost. Yet, political integration merely worsens overall preference matching. At the opposite extreme, if a minority is completely uninformed it is essentially disenfranchised. Then utilitarian welfare maximization requires political integration to protect the minority ($\bar{\zeta} < 1$ for all $\tau < 1$).

More generally, ideological sorting provides a countervailing centrifugal force. As groups with opposite preferences are more and more segregated, the difference in preferences across regions increases. The appeal of political separation increases smoothly with the degree of ideological separation ($\partial\bar{\zeta}/\partial\tau > 0$). In the limit, political separation is optimal if ideological sorting is complete, as Proposition 3 already established ($\lim_{\tau\to 1}\bar{\zeta} = 1$). Finally, just as in Proposition 1, greater variance in politicians' ability makes integration less attractive because of distortions in the allocation of talent ($\partial\bar{\zeta}/\partial\sigma > 0$).[19]

Our results speak directly to proposals for the partition of California, which have been put forward several times—most recently, venture capitalist Tim Draper attempted to introduce for 2016 a ballot initiative to split the state in six. By far the largest state in the union, California is composed of several distinct regions (Baldassarre 2000; Gimpel and Schuknecht 2003). The most traditional divide is between North and South (Douzet and Miller 2008), but today the most salient divide is between East and West. The differences are both partisan and ideological: Western California is more liberal, even among Republican voters and politicians; Eastern California considerably more conservative (Kousser 2009). At a first glance, such a political divide might suggest that a break up of coastal and inland California would be optimal on preference-matching grounds.

Proposition 4, however, cautions against this superficial assessment. Both the southeastern Inland Empire and the San Joaquin Valley contain a large Hispanic population that overwhelmingly prefers the Democratic party (Michelson 2005). This group is much less educated, less politically knowledgeable, and less likely to vote than Republican supporters in

[19]The effect of political integration on screening would be opposite if majorities were systematically less informed than minorities. Aside from comparative statics, however, the trade off presented by Proposition 1 remains in this less intuitive case. If an uninformed local majority is dominated by an informed minority, a fortiori political integration has the benefit of equalizing the power of the two groups. It raises welfare if and only if sorting is sufficiently imperfect.

the region, who are on average older, whiter, and wealthier.[20] At the same time, the left-wing Hispanic working class in the Valley shares the political leanings of highly educated liberals on the coast. This ideological alignment goes beyond mere partisanship and includes shared preferences over policies: "whether they ride in limousines, Volvos, or buses, Democrats in the blue areas of the state share similar policy views" (Kousser 2009, p. 2).

As a consequence, our model suggests that the political integration of California is welfare maximizing. For relatively uneducated inland minorities to have a government corresponding to their preferences, it is essential that they share a state with ideologically aligned liberal elites in the Bay area. Right-wing Californians, instead, are sufficiently educated and influential to have a voice in state-wide politics, despite being in the minority: California had a Republican governor for twenty-one of the past thirty years.

The lesson of Proposition 4 applies more broadly. Disadvantaged ethnic minorities—which are less educated and often politically underrepresented—should belong whenever possible to the same polity as better educated and higher-status voters having similar political preferences. Only then are politicians effectively held accountable to both groups.

6 Will the Informed Support Political Integration?

Our analysis has focused on the welfare consequences of government structure. Differences in information across regions make political integration desirable both because it yields efficiency gains from increased accountability and because it is a form of progressive redistribution. Uninformed regions reap large gains while informed ones suffer small losses, as shown in Proposition 1. Such distributional effects of centralization are appealing from the perspective of aggregate social welfare, but they raise a question of feasibility: will informed regions oppose and block optimal integration? This question is particularly relevant in Europe. Propositions 1 and 2 suggest that a federal structure in the European Union may be optimal due to the large disparities in accountability across member states (Charron, Dijkstra, and Lapuente 2014). But why would Danes and Germans agree to a federation whose benefits accrue to Greeks and Italians?

In this section, we extend our model in two directions that show how political integration can receive unanimous support. First, we allow for public-good spillovers across regions, a classic element of the fiscal-federalism literature since Oates (1972). In our model, externalities imply not only—mechanically—that the informed care about public goods in uninformed regions, but also that centralization may increase government efficiency in informed regions

[20] Hispanic immigrants are also more likely not to have the right to vote, but a substantial majority of hispanic residents of southeastern California are U.S. citizens.

too. Alternatively, we discuss how unanimity can be obtained at the expense of welfare maximization, by combining centralization with partial discretionality in public-good provision.

6.1 Public-Good Spillovers

We introduce externalities with a simple symmetric specification that preserves constant aggregate returns to scale. There is a single composite public good ($P = 1$) and a resident i of region l has utility

$$u_t^i = \tilde{u}_t^i + (1 - \xi) \log g_{l,t} + \frac{\xi}{L} \sum_{m=1}^{L} \log g_{m,t}, \tag{21}$$

where the index $\xi \in [0, 1]$ measures interregional spillovers. Citizens' mobility within the United States or the European Union provides an intuitive interpretation of this setup. Each agent has a probability ξ of moving, and conditional on a move he has equal probability of moving to each region.

Public-good spillovers entail systematic differences between the productivity of the central government and that of local governments.

Proposition 5 *Suppose there are spillovers in public goods across regions ($\xi > 0$). Then the expected competence of ruling politicians is on average higher under centralization than decentralization ($\mathbb{E}\hat{\eta}^C > \sum_{l=1}^{L} \mathbb{E}\hat{\eta}_l^D / L$). Aggregate rent extraction is lower under centralization than decentralization regardless of differences in voter information ($\rho^C < \sum_{l=1}^{L} \rho_l^D / L$). Both efficiency advantages of centralization are increasing in the extent of spillovers ($\partial(\mathbb{E}\hat{\eta}^C - \sum_{l=1}^{L} \mathbb{E}\hat{\eta}_l^D / L)/\partial\xi > 0$ and $\partial \left(\sum_{l=1}^{L} \rho_l^D / L - \rho^C \right) / \partial\xi > 0$).*

Internalizing spillovers through centralization raises the screening value of elections and thus the expected productivity of elected politicians. Informed voter may support an incompetent incumbent because of his personal likability or ideological affinity, but they are less likely to be swayed by such factors when politicians' skills are more important. Public-good spillovers imply that competence is more important for the central than the local government. The ability of local politicians influences local public goods only; that of central politicians also determines spillovers from other regions. Therefore, voters are keener on screening for competence at the central than at the local level. This sharper voter focus on competence improves the monitoring as well as the screening value of elections. As a result, public-good spillovers strengthen the accountability gains from centralization: rent extraction declines with political integration even when regions have identical information. Both efficiency advantages of centralization are monotone increasing in the extent of spillovers.

The improvement in politicians' selection and incentives described by Proposition 5 is distinct from the benefits of policy coordination that Oates (1972) highlighted as a rationale for centralization. Coordination is reflected in an improvement in resource allocation rather than in government productivity. This additional classic element is also present in our model when we consider both a public good g that generates inter-regional spillovers $\xi > 0$ and another public good h whose benefits are purely local. Then, a resident i of region l has utility

$$u_t^i = \tilde{u}_t^i + \alpha_g \left[(1-\xi) \log g_{l,t} + \frac{\xi}{L} \sum_{m=1}^{L} \log g_{m,t} \right] + (1-\alpha_g) \log h_{l,t}, \tag{22}$$

where $\alpha_g \in (0,1)$ is the share of resources that would be allocated to the spillover-generating public good by a benevolent planner. Then the equilibrium allocation of resources across public goods is systematically different under centralization and decentralization

Corollary 2 *Centralization induces the socially optimal allocation resources across public goods ($\beta_g^C = \alpha_g$). Decentralization induces an insufficient allocation of resources to the spillover-generating public good ($\beta_{g,l}^D < \alpha_g$ for all l). Underprovision is increasing in the size of spillovers ($\partial \beta_{g,l}^D / \partial \xi < 0$).*

Incumbents provide public goods merely to showcase their ability to their own constituents. Under centralization, all beneficiaries of each public good vote for the incumbent's re-election. Then career concerns are exactly aligned with social welfare across goods. Resources are allocated to public goods in proportion to the full social value of each investment and each skill. Under decentralization, instead, career concerns induce every local politician to ignore all spillovers. Externality-inducing goods are under-provided and purely local goods are over-provided instead. Incumbents are uninterested in demonstrating their ability at generating welfare for regions that do not vote for their re-election. As a consequence, centralization entails endogenous gains from policy coordination.

Oates (1972) assumed that local governments maximize local residents' welfare but are exogenously incapable of cooperating to reach Pareto improvements. Such a cooperation failure can be microfounded through frictions in bargaining between benevolent local governments (Harstad 2007). Corollary 2 provides the complementary microfoundation. If bargaining is frictionless but local politicians are rent-seeking instead of benevolent, career concerns provide them with no incentives to cooperate in the pursuit of aggregate social welfare. Cooperation is irrelevant for the pursuit of their own goal, reelection.

6.2 Partial Discretionality

If spillovers are modest or absent, is it ever possible to obtain unanimous support for the transfer of powers to a central government? In this context, the regressive distributive consequences of centralization without a uniformity constraint have a silver lining. Discretionality transfers power to the informed. This transfer is welfare-reducing, but it can be the price to pay to buy their support for an efficiency-increasing institutional reforms.

Consider homogeneous, symmetric preferences ($\upsilon \to \infty$) over a measure-one continuum of public goods. A resident i of region l has utility

$$u_t^i = \tilde{u}_t^i + \int_0^1 \log g_{l,t}(p)\, dp. \tag{23}$$

Centralization is characterized by an index of discretionality $\omega \in [0,1]$ such that goods $p \in [0,\omega]$ are not subject to the uniformity constraint, while goods $p \in [\omega,1]$ are. By a straightforward extension of Proposition 1, social welfare is maximized by full uniformity ($\omega^* = 0$) and declines as discretionality increases. On the other hand, we can establish the following result.

Proposition 6 *Suppose that the variance of politicians' ability is not too high ($\sigma^2 \leq \bar{\sigma}^2$). Then there is a level of discretionality $\tilde{\omega} \in \left(\rho^C, 1\right)$ such that centralization with discretionality $\tilde{\omega}$ is preferred to decentralization by every region. The minimum discretionality required for centralization to enjoy unanimous support is lower when voters are more informed ($\partial\tilde{\omega}/\partial\bar{\theta} < 0$) and politicians' ability less variable ($\partial\tilde{\omega}/\partial\sigma^2 > 0$).*

Better incentives for central politicians reduce aggregate rent extraction and thus create an overall surplus. Proposition 6 shows that the incentives of the central government can be fine-tuned so that all regions share in the efficiency gains from centralization, irrespective of the distribution of voter information. Centralization transfers power over the allocation of a share ω of public goods from uninformed to informed regions. It also transfers accountability from informed regions to uninformed regions, inducing a uniform rent extraction ρ^C.

The uninformed gain more from reducing local rents to ρ^C than the informed lose from raising local rents to ρ^C. Then, if $\omega \geq \rho^C$ the gain in power is worth more to the informed than their local decline in accountability. But if $\omega \leq \rho^C$ the loss of power is worth less to the uninformed than their local increase in accountability. When rent extraction and discretionality are exactly matched ($\omega = \rho^C$), all regions with $\theta_l \neq \bar{\theta}$ strictly prefer the endogenous allocation of resources under centralization to the one under decentralization (a region with exactly average information is indifferent). Higher voter information implies lower rent extraction by the central government. Then, informed regions require less discretionality to

support centralization ($\partial\tilde{\omega}/\partial\bar{\theta} < 0$).

Political integration is also redistributive with respect to screening. Central politicians have average skills above local politicians in uninformed regions, but below local politicians in informed ones. Unanimity requires informed regions to gain enough power to offset this progressive transfer through government selection. Therefore, the required discretionality is $\tilde{\omega} > \rho^C$, and it increases monotonically with the importance of political screening. If the variance of ability were too high, unanimous support for centralization might prove impossible ($\sigma^2 > \bar{\sigma}^2$). However, we view as a natural benchmark the case in which moral hazard is a greater problem in political agency than adverse selection.

The political debate within the European Union, whose treaties are adopted by unanimity of the member states, is consistent with the patterns described by Proposition 6. "Core" countries such as Austria, Finland, Germany and the Netherlands complain about the low institutional quality and the ineffective and corrupt politicians in "peripheral" countries such as Greece, Italy, Portugal, and Spain. Such complaints chime with our prediction of declining government accountability and productivity for the more informed regions. At the same time, peripheral countries complain that European policy is largely dictated by core countries and disproportionately caters to their needs and interests. Again, this accords with our prediction of declining policy-making power for the less informed regions. Proposition 6 suggests that intra-European frictions may be manifestations of a Pareto-improving agreement that makes the Union beneficial for all members, albeit not welfare-maximizing.

7 Rethinking State Borders in the United States

The United States display a striking geographic heterogeneity in culture and ideology (Glaeser and Ward 2006). The classic logic of Tiebout (1956) and Oates (1972) suggests that it is optimal for such differences to be reflected in a partition of the country into relatively homogeneous red and blue states. Our model, however, sounds a cautionary note. Government accountability also varies widely across the United States (Glaeser and Saks 2006). The most corrupt states, such as Louisiana and Mississippi, witness five times as many federal corruption convictions per capita as the least corrupt ones, such as Oregon and Washington. Proposition 1 highlights the risk that America could be fragmented in an excessive number of states that differ less in their residents' preferences than in their ability to monitor their government.

In this section, we investigate the potential benefits of state consolidations by applying our model to the 48 contiguous United States. We consider the 105 pairs of states that share a border and compute the welfare gains or losses from removing the border by merging the

two contiguous states. To quantify our model, we need to calibrate two variables and three parameters. Each state is characterized by preferences $\boldsymbol{\alpha}_l$ and voter information θ_l. Electoral discipline depends on the discount factor δ, the variance of politicians' competence σ^2, and voter's keenness on competence relative to other determinants of political popularity, ϕ.

We measure voter preferences by the average vote shares of the Democratic and Republican parties in the six latest presidential elections, from 1992 to 2012 inclusive. Formally, we consider two composite public goods: a conservative and a liberal bundle. The ideal share α_l^R of conservative public goods in state l is proxied by the Republican share of the two-party vote total. This measure is appealing because the divide between the parties largely reflects cultural and religious issues (Glaeser, Ponzetto and Shapiro 2005; Glaeser and Ward 2006). Averaging over the last twenty years yields and intuitive list of red and blue states. The five most conservative states are Utah, Wyoming, Idaho, Nebraska and Oklahoma, with Republican shares above 60%. The most liberal states are Rhode Island, Massachusetts, Vermont, and Maryland—the latter has a Republican share of 40%.

We measure voter information by the share of college graduates among the population over 25, averaging CPS data for the six election years 1992-2012. Education is admittedly a coarse proxy for the ability to monitor politicians, but a high share of educated voters is the best predictor of a less corrupt government, both within the United States and internationally (Svensson 2005; Glaeser and Saks 2006). The ranking of states by education is also intuitive. Massachusetts, Colorado, Connecticut, Maryland and New Jersey are the most educated states: more than 32% of their residents have a BA. West Virginia, Arkansas, Indiana, Mississippi and Kentucky are the least educated, with no more than 20% of residents having a BA.

We set the discount factor to $\delta = .85$ for a four-year term, which corresponds to a real interest rate of about 4% on an annual basis. The two remaining parameters can be calibrated from a measure of state governors' performance and its impact on their likelihood of winning re-election. Choosing a benchmark to assess government performance is difficult and controversial. We pick an uncontroversial proxy for voter welfare: the growth rate of state income per capita. This is also a reasonable if coarse proxy for governors' performance because it significantly predicts their probability of re-election (Besley 2006). Voters appear to infer governors' talent from economic growth under their watch. On the other hand, income growth clearly reflects factors beyond a governor's control. Accordingly, we decompose the growth rate γ_t into an exogenous component $\tilde{\gamma}_t$ and the welfare impact of government policy $\log g_t$. The basis of our calibration is the welfare function

$$\gamma_t = \tilde{\gamma}_t + \log g_t = \varepsilon_t + \varepsilon_{t-1} + \log \bar{x} + \tilde{\gamma}_t, \tag{24}$$

which implies a probability of re-election

$$\pi_t = \frac{1}{2} + \phi\theta \left(\log g_t - \log \bar{x} - \varepsilon_{t-1}\right) = \frac{1}{2} + \phi\theta\varepsilon_t. \tag{25}$$

We calibrate our model to Besley's (2006) empirical analysis of the determinants of governors' re-election in the 48 contiguous United States from 1950 to 2000. Our setup fits his linear probability model, but it implies that his regressor γ_t is a noisy measurement of the true determinant of re-election ε_t. Thus, the coefficient estimate suffers from attenuation bias. If the volatility of income growth reflects the variance of the governor's competence innovation with a signal ratio $s \in (0, 1]$, then

$$\sigma^2 = s \operatorname{Var}(\gamma_t) \text{ and } \phi = \frac{\hat{\beta}}{\theta s}. \tag{26}$$

In Besley's (2006) data, the variance of the growth rate of state income per capita in the two years before an election is $\operatorname{Var}(\gamma_t) = 0.405\%$. The estimated impact of income growth on the probability of re-election is $\hat{\beta} = 1.808$. We set $\theta = 0.138$ to match the average college share across the 48 states and 6 census years, 1950-2000.

Rather than taking a stand on the precise extent to which government policy affects economic volatility, we report our results for different parametrizations of the signal-extraction ratio s. The screening value of elections is independent of this parameter, since it is proportional to $\phi\sigma^2 = \operatorname{Var}(\gamma_t)\hat{\beta}/s$. A lower signal ratio s only entails a higher estimate of voters' responsiveness to the incumbent's skill, and thus better incentives for politicians. If $s = 1/2$, then our calibration implies rents ranging from 6.6% to 14.3% of the state budget, with a mean of 9.4%. If instead $s = 1/8$, rents range from 1.8% to 4% and average 2.5%.

Table 1 presents the list of mergers of contiguous states that would yield the largest welfare gains in light of our model. For each state, we list only the single best merger involving it, in keeping with our focus on pairwise mergers. Column 1 shows the ranking for $s = 1/2$ and column 2 for $s = 1/8$.

The desirable state mergers are remarkably intuitive. All pairs of states in Table 1 share not only closely aligned partisan preferences, but also a similar culture more broadly. By far the most desirable change of borders would be the reunion of Virginia and West Virginia. The two states have very similar vote shares (51% Republican for Virginia and 52% for West Virginia) but dramatically different levels of human capital (respectively 31% and 16% college graduates). Another welfare-increasing merger would constitute a return to past integration: Vermont seceded from New York during the Revolution, just as West Virginia seceded from Virginia during the Civil War. Consolidation of states in New England (such

Table 1: Most desirable state mergers

Rank	$s = 1/2$	$s = 1/8$
1	Virginia - West Virginia	Virginia - West Virginia
2	Arkansas - Missouri	Arkansas - Missouri
3	New Jersey - Pennsylvania	Georgia - Tennessee
4	Kansas - Oklahoma	Kansas - Oklahoma
5	Iowa - Minnesota	Massachusetts - Rhode Island
6	Colorado - New Mexico	Iowa - Minnesota
7	Maine - New Hampshire	New Jersey - Pennsylvania
8	Georgia - Tennessee	Delaware - Maryland
9	Massachusetts - New York	Idaho - Wyoming
10	Nevada - Oregon	New York - Vermont

Notes: Preferences α_l are measured by the average Republican share of the two-party vote in presidential elections 1992-2012. Information θ_l is measured by the share of people over 25 with a BA degree, from March CPS data in election years 1992-2012. Four-year discount factor $\delta = .85$. Voters' keenness on politicians' ability ϕ and its variance σ^2 are calibrated to Besley's (2006) analysis of governors' re-election as a function of the growth rate (γ_t) of state income per capita in the two years before an election, 1950-2000. $s = \sigma^2 / \operatorname{Var}(\gamma_t)$ is the signal ratio of governor's ability in observed income growth.

as Maine and New Hampshire, or Massachusetts and Rhode Island) was proposed by the historian Frederick Jackson Turner (1921), and by Connecticut governor Wilbur Cross in 1931 and again in 1939.

The set of desirable mergers is rather robust to changes in the signal ratio s. Nonetheless, differences in the ranking for $s = 1/2$ and $s = 1/8$ have a precise explanation. The higher s, the worse our estimate of electoral discipline, the higher the implied rent-extraction and the larger the accountability benefits from integration. If s is high the most appealing mergers are those between states with larger differences in education, even if their political preferences are not perfectly aligned: e.g., Colorado and New Mexico have respectively 35% and 26% college graduates and Republican vote shares of 50% and 46%.

If instead s is low, our estimate of accountability is already very high without mergers, so further integration is less desirable. Mergers that increase welfare if $s = 1/8$ are a strict subset of those that increase it for $s = 1/2$: in the latter case, 36 of the potential 105 mergers are desirable, but in the former only 20. The ranking then privileges safer mergers of states with very similar preferences, even if their levels of human capital are less far apart: e.g., Delaware and Maryland have Republican vote shares of 42% and 40% respectively, and 26% and 34% college graduates. All in all, we interpret the second column of Table 1 as a

conservative set of mergers that pose little risk of reduced preference-matching, while they offer the potential for accountability gains.

How large are the potential net gains? In our calibration, welfare is measured in terms of income growth rates. The benefits of merging Virginia and West Virginia equal those from an increase in the annual growth rate of real income per capita by 30.9 basis points if $s = 1/2$ and 8.5 basis points if $s = 1/8$. For Iowa and Minnesota—the median merger in both lists—the gains are respectively 5.5 or 0.8 basis points.

These quantitative results are admittedly coarse, but the qualitative lesson of our model seems clear. Redrawing state borders may not deserve a very high priority as a political reform in the United States, but the existing 50 states are most likely too many. Excessive fragmentation contributes to inefficiency and corruption in state governments, and a few state mergers would improve the American political landscape.

8 Conclusion

Should different people have different governments? The idea has gained wide currency, from European Union law enshrining the principle of subsidiarity to independence movements in Québec, Scotland or Catalonia and recurring proposals to split California into separate liberal and conservative states. The classic theory of fiscal federalism supports and formalizes the intuitive appeal of this notion: according to the seminal Decentralization Theorem (Oates 1972), decentralization is more efficient than centralization whenever regions are not identical and there are no policy spillovers.

This paper has offered a different perspective by focusing on a key overlooked dimension of regional heterogeneity: voters' ability to monitor politicians and hold them accountable. Our model explains why local governments with homogeneous constituencies can end up as political failures (Glaeser and Shleifer 2005) and why decentralization works better in a country with fairly homogeneous accountability like Germany than in one with gaping regional disparities like Italy (Ziblatt 2006).

When voter information varies across regions, political integration yields aggregate gains in accountability. The central government is monitored mainly by the most informed regions and as consequence it has better incentives to serve its citizens than the average local government. At the same time, however, its incentives are disproportionately to serve the informed and neglect the uninformed, so it must be forced to provide public goods uniformly in order to avoid unacceptable distributive distortions. The same mechanism thus drives two opposing forces: preference heterogeneity prompts decentralization, as in the standard theory; information heterogeneity, however, prompts centralization instead.

As a consequence, we have also shown that the borders of governments should not reflect only the classic Tiebout (1956) logic of separating people with different preferences. In addition to clustering by tastes, it is also crucial to ensure diversity of information. In particular, uninformed voters are caught between the hammer of unaccountable politicians and the anvil of better informed voters with contrasting policy priorities. A concern for social welfare requires them to share a government with highly informed voters with similar preferences. Thus, California should not be broken up: the benefits of separating a liberal local majority on the coast from a conservative local majority inland are likely to be smaller than those of grouping together the coastal liberal elite with the working-class left-wing minority in the Central Valley.

In fact, our analysis suggests that the main problem with state boundaries in the United States is not that states like California are too big and diverse, but on the contrary that many states are too small and insufficiently diverse. We have calibrated our model to the post-war pattern of gubernatorial elections (Besley 2006) and shown that around a quarter of possible pairwise mergers of states sharing a border would be welfare-increasing. Although only heterogeneity, not size per se, determines optimal integration in our model, our quantitative results indicate that merging away the smallest states would provide the most obvious benefits. Out of twelve states with less than two million inhabitants (excluding Alaska and Hawaii), we have found that at least half should not remain separate. Mergers involving Wyoming, Vermont, Delaware, Rhode Island, Idaho, and especially West Virginia are robustly among the most attractive.

Our framework has also offered novel insights on federal systems with multiples level of government. The standard logic of fiscal federalism suggests there should be many government layers, so that every policy instrument is tied to its optimal geographic unit. Instead, we have shown that government accountability exhibits economies of scope: a unitary government that controls a large budget and multiple policy instruments suffers less from moral hazard than many special-purpose governments, each controlling a specific policy and its separate budget. Our model thus explains the observed inefficiency of special-purpose districts in the United States (Berry 2009) and ongoing reforms to reduce the number of government tiers in European countries.

Furthermore, we have found that a federal structure can be desirable only if information heterogeneity is large enough. This results speaks in particular to the ongoing European debate. Since the start of the crisis in the Euro area, there have been suspicions that differences in institutional quality across member states might be too large for the smooth working of the European Union. How can the Union include virtuous "core" countries like Germany, the Netherlands, or Finland, and at the same time the troubled Euro periphery

of Italy, Spain, Portugal and Greece, not to mention post-communist Eastern Europe? Our model shows that such differences in government accountability are not a weakness but instead a motivating strength of the European project. They explain why we can expect overall efficiency gains from transferring powers to EU institutions, but also why substantial policy instruments should remain at the national level.

The European case is also consistent with our results on the trade-off between welfare maximization and unanimous support for integration. In our model, the best informed regions favor centralization, even without externalities, if the can gain control of some union-wide policy. Unanimous centralization is, in effect, an exchange of power for accountability. Accordingly, in the European discourse core countries complain of low institutional quality in the periphery and its detrimental impact on the whole union; at the same time the periphery complains that EU policy is disproportionately shaped by the needs and preferences of the core. Both complaints may well be justified. Our theory shows they reflect the distributive consequences of centralization under partial discretionality, which ensures all member states gain from European integration but also leaves each of them with something to complain about.

While we have not extended our analysis quantitatively in this direction, our framework has the potential to help explain another enduring European puzzle: why the European Union does precisely what it does (Alesina, Angeloni and Schuknecht 2005). The allocation of policy instruments between the European institutions and the member states is not entirely accounted for by the classic considerations of externalities and taste heterogeneity. Our theory suggests two more considerations are equally crucial. Efficiency is maximized by centralizing policies for which citizens' monitoring ability varies most starkly across countries. Political feasibility may require striking a balance between instruments that transfer power to the core and others that transfer accountability to the periphery.

References

[1] Ackerman, Kenneth D. 2005. *Boss Tweed: The Rise and Fall of the Corrupt Pol Who Conceived the Soul of Modern New York*. New York: Carroll & Graf.

[2] Adserà, Alícia, Carles Boix, and Mark Payne. 2003. "Are You Being Served? Political Accountability and the Quality of Government." *Journal of Law, Economics, and Organization* 19 (2): 445–90.

[3] Alesina, Alberto, Ignazio Angeloni, and Federico Etro. 2005. "International Unions." *American Economic Review* 95 (3): 602–15.

[4] Alesina, Alberto, Ignazio Angeloni, and Ludger Schuknecht. 2005. "What Does the European Union Do?" *Public Choice* 123 (3): 275–319.

[5] Alesina, Alberto, and Enrico Spolaore. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economics* 112 (4): 1027–56.

[6] ——. 2003. *The Size of Nations*. Cambridge, MA: MIT Press.

[7] Alesina, Alberto, and Guido Tabellini. 2008. "Bureaucrats or Politicians? Part II: Multiple Policy Tasks." *Journal of Public Economics* 92: 426–47.

[8] Andrews, Rhys, and George Boyne. 2009. "Size, Structure and Administrative Overheads: An Empirical Analysis of English Local Authorities." *Urban Studies* 46: 739–59.

[9] Ashworth, Scott, and Ethan Bueno de Mesquita. 2008. "Electoral Selection, Strategic Challenger Entry, and the Incumbency Advantage." *Journal of Politics* 70 (4): 1006–25.

[10] Attali Commission. 2008. *300 décisions pour changer la France* [300 Decisions to Change France]. Paris: XO

[11] Baldassarre, Mark. 2000. *California in the New Millennium: The Changing Social and Political Landscape*. Berkeley, CA: University of California Press.

[12] Banks, Jeffrey S., and Rangarajan K. Sundaram. 1993. "Adverse Selection and Moral Hazard in a Repeated Elections Model." In *Political Economy: Institutions, Competition and Representation*, edited by William A. Barnett, Melvin Hinich, and Norman Schofield, 295–311. Cambridge: Cambridge University Press.

[13] ——. 1998. "Optimal Retention in Agency Problems." *Journal of Economic Theory* 82: 293–323.

[14] Bardhan, Pranab, and Dilip Mookherjee. 2000. "Capture and Governance at Local and National Levels." *American Economic Review* 90 (2): 135–9.

[15] ——. 2006a. "Decentralisation and Accountability in Infrastructure Delivery in Developing Countries." *Economic Journal* 116(1): 101–27.

[16] ——. 2006b. "Decentralization, Corruption, and Government Accountability." In *International Handbook on the Economics of Corruption*, edited by Susan Rose-Ackerman, 161–88. Cheltenham: Edward Elgar.

[17] Baron, David P. 1994. "Electoral Competition with Informed and Uninformed Voters." *American Political Science Review* 88: 33–47.

[18] Barro, Robert, and Jong-Wha Lee. 2010. "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics* 104: 184–98.

[19] Belleflamme, Paul, and Jean Hindriks. 2005. "Yardstick Competition and Political Agency Problems." *Social Choice and Welfare* 24 (1): 155–69.

[20] Berry, Christopher B. 2009. *Imperfect Union: Representation and Taxation in Multilevel Governments*. Cambridge: Cambridge University Press.

[21] Besley, Timothy. 2006. *Principled Agents? The Political Economy of Good Government*. Oxford: Oxford University Press.

[22] Besley, Timothy, and Robin Burgess. 2002. "The Political Economy of Government Responsiveness: Theory and Evidence from India." *Quarterly Journal of Economics* 117 (4): 1415–51.

[23] Besley, Timothy, and Anne Case. 1995. "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition." *American Economic Review* 85 (1): 25–45.

[24] Besley, Timothy, and Stephen Coate. 2003. "Centralized Versus Decentralized Provision of Local Public Goods: A Political Economy Approach." *Journal of Public Economics* 87: 2611–37.

[25] Besley, Timothy, and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton, NJ: Princeton University Press.

[26] Besley, Timothy, and Michael Smart. 2007. "Fiscal Restraints and Voter Welfare." *Journal of Public Economics* 91: 755–73.

[27] Bewley, Truman F. 1981. "A Critique of Tiebout's Theory of Local Public Expenditures." *Econometrica* 49 (3): 713–40.

[28] Blanchard, Olivier, and Andrei Shleifer. 2001. "Federalism With and Without Political Centralization: China Versus Russia." *IMF Staff Papers* 48 (4): 171–9.

[29] Boffa, Federico, Amedeo Piolatto, and Giacomo A. M. Ponzetto. 2014. "Political Centralization and Government Accountability." CEPR Discussion Paper No. 9514.

[30] Callander, Steve, and Bard Harstad. 2015. "Experimentation in Federal Systems." *Quarterly Journal of Economics,* online access.

[31] Charron, Nicholas, Lewis Dijkstra, and Victor Lapuente. 2014. "Regional Governance Matters: A Study on Regional Variation in Quality of Government within the EU." *Regional Studies* 48 (1): 68–90.

[32] Court of Auditors. 2013. *Les Finances publiques locales: Rapport public thématique* [Local Government Finances: Public Thematic Report]. Paris: Court of Auditors of France.

[33] Court of Auditors. 2014. *Les Finances publiques locales: Rapport public thématique* [Local Government Finances: Public Thematic Report]. Paris: Court of Auditors of France.

[34] Dente, Bruno. 1988. "Local Government Reform and Legitimacy." In *The Dynamics of Institutional Change: Local Government Reorganization in Western Democracies*, edited by Bruno Dente and Francesco Kjellberg. London: Sage.

[35] Dincecco, Mark. 2009. "Fiscal Centralization, Limited Government, and Public Revenues in Europe, 1650-1913." *Journal of Economic History* 69: 48-103.

[36] ——. 2011. *Political Transformations and Public Finances*. Cambridge: Cambridge University Press.

[37] Dincecco, Mark, and Gabriel Katz. 2015. "State Capacity and Long-Run Performance." *Economic Journal*, forthcoming.

[38] Douzet, Frederick, and Kenneth P. Miller. 2008. "California's East-West Divide." In *The New Political Geography of California*, edited by Frederick Douzet, Thad Kousser, and Kenneth P. Miller, 9–43. Berkeley, CA: Berkeley Public Policy Press.

[39] Durante, Ruben, Giovanna Labartino, and Roberto Perotti. 2014. "Academic Dynasties: Decentralization and Familism in the Italian Academia." NBER Working Paper No 17572.

[40] Fan, C. Simon, Chen Lin, and Daniel Treisman. 2009. "Political Decentralization and Corruption: Evidence from Around the World." *Journal of Public Economics* 93 (1-2): 14–34.

[41] Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer. 2013. "The Next Generation of the Penn World Table." Available for download at www.ggdc.net/pwt.

[42] Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123 (2): 703–45.

[43] Fisman, Raymond, and Roberta Gatti. 2002. "Decentralization and Corruption: Evidence Across Countries." *Journal of Public Economics* 91: 2261–90.

[44] Fredriksson, Per G., and Noel Gaston. 2000. "Environmental Governance in Federal Systems: The effects of Capital Competition and Lobby Groups." *Economic Inquiry* 38(3): 501–14.

[45] Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky. 2008. "School Decentralization: Helping the Good Get Better, But Leaving the Poor Behind." *Journal of Public Economics* 92: 2106–20.

[46] Gennaioli, Nicola, and Ilia Rainer. 2007a. "The Modern Impact of Precolonial Centralization in Africa." *Journal of Economic Growth* 12 (3): 185–234.

[47] ——. 2007b. "Precolonial Centralization and Institutional Quality in Africa." In *Institutions and Norms in Economic Development*, edited by Mark Gradstein and Kai Konrad. Cambridge, MA: MIT Press.

[48] Gennaioli, Nicola, and Hans-Joachim Voth. 2015. "State Capacity and Military Conflict." *Review of Economic Studies*, forthcoming.

[49] Gentzkow, Matthew. 2006. "Television and voter turnout." *Quarterly Journal of Economics* 121 (3): 931–72.

[50] Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 78(1): 35–71.

[51] Gimpel, James G., and Jason E. Schuknecht. 2003. *Patchwork Nation: Sectionalism and Political Change in American Politics*. Ann Arbor, MI: University of Michigan Press.

[52] Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9 (3): 271–303.

[53] Glaeser, Edward L., and Giacomo A. M. Ponzetto. 2014. "Shrouded Costs of Government: The Political Economy of State and Local Public Pensions." *Journal of Public Economics* 116: 89–105.

[54] Glaeser, Edward L., Giacomo A. M. Ponzetto, and Jesse M. Shapiro. 2005. "Strategic Extremism: Why Republicans and Democrats Divide on Religious Values." *Quarterly Journal of Economics* 120 (4): 1283–330.

[55] Glaeser, Edward L., and Raven E. Saks. 2006. "Corruption in America." *Journal of Public Economics* 90: 1053–72.

[56] Glaeser, Edward L., and Andrei Shleifer. 2005. "The Curley Effect: The Economics of Shaping the Electorate." *Journal of Law, Economics, and Organization* 21 (1): 1–19.

[57] Glaeser, Edward L., and Bryce A. Ward. 2006. "Myths and Realities Of American Political Geography." *Journal of Economic Perspectives* 20 (2): 119–44.

[58] Grossman, Gene M., and Elhanan Helpman. 1996. "Electoral Competition and Special-Interest Politics." *Review of Economic Studies* 63: 265–86.

[59] Harstad, Bård. 2007. "Harmonization and Side Payments in Political Cooperation." *American Economic Review* 97(3): 871–89.

[60] Hindriks, Jean, and Ben Lockwood. 2009. "Decentralization and Electoral Accountability: Incentives, Separation and Voter Welfare." *European Journal of Political Economy* 25: 385–397.

[61] Holmström, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies* 66 (1): 169–82.

[62] Joanis, Marcelin. 2014. "Shared Accountability and Partial Decentralization in Local Public Good Provision." *Journal of Development Economics* 107: 28–37.

[63] Junger, Richard. 2010. *Becoming the Second City: Chicago's Mass News Media, 1833-1898*. Urbana, IL: University of Illinois Press.

[64] Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi. 2010. "The Worldwide Governance Indicators: A Summary of Methodology, Data and Analytical Issues." World Bank Policy Research Working Paper No. 5430.

[65] Kotsogiannis, Christos, and Robert Schwager. 2006. "On the Incentives to Experiment in Federations." *Journal of Urban Economics* 60 (3): 484–97.

[66] Kousser, Thad. 2009. "How Geopolitics Cleaved California's Republicans and United Its Democrats." *California Journal of Politics and Policy* 1 (1).

[67] Le Galès, Patrick, and Peter John. 1997. "Is the Grass Greener on the Other Side? What Went Wrong with French Regions, and the Implications for England." *Policy & Politics* 25 (1): 51–60.

[68] Lindbeck, Assar, and Jörgen W. Weibull. 1987. "Balanced-Budget Redistribution as the Outcome of Political Competition." *Public Choice* 52: 273–97.

[69] Lockwood, Ben. 2002. "Distributive Politics and the Costs of Centralization." *Review of Economic Studies* 69: 313–37.

[70] ——. 2006. "The Political Economy of Decentralization." In *Handbook of Fiscal Federalism*, edited by Ehtisham Ahmad and Giorgio Brosio, 33–60. Cheltenham: Edward Elgar.

[71] ——. 2008. "Voting, Lobbying, and the Decentralization Theorem." *Economics and Politics* 20: 416–31.

[72] Michelson, Melissa. 2005. "Does Ethnicity Trump Party? Competing Vote Cues and Latino Voting Behavior." *Journal of Political Marketing* 4: 1–26.

[73] Nannicini, Tommaso, Andrea Stella, Guido Tabellini and Ugo Troiano. 2012. "Social Capital and Political Accountability." *American Economic Journal: Economic Policy* 5 (2): 222–50.

[74] Oates, Wallace E. 1972. *Fiscal Federalism*. New York, NY: Harcourt Brace Jovanovich.

[75] ——. 1999. "An Essay on Fiscal Federalism." *Journal of Economic Literature* 37: 1120–49.

[76] Persson, Torsten, and Guido Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge, MA: MIT Press.

[77] Ponzetto, Giacomo A. M. 2011. "Heterogeneous Information and Trade Policy." CEPR Discussion Paper No. 8726.

[78] Ponzetto, Giacomo A. M., and Ugo Troiano. 2014. "Social Capital, Government Expenditures, and Growth." CEPR Discussion Paper No. 9891.

[79] Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

[80] Salmond, Alex. 2013. Preface to *Scotland's Future: Your Guide to an Independent Scotland*, produced by APS Group Scotland. Edinburgh: The Scottish Government.

[81] Seabright, Paul. 1996. "Accountability and Decentralization in Government: An Incomplete Contracts Model." *European Economic Review* 40: 61–89.

[82] Seifert, Stefan, and Maria Nieswand. 2014. "What Drives Intermediate Local Governments' Spending Efficiency: the Case of French *Départements*." *Local Government Studies* 40 (5): 766–90.

[83] Snyder, James M., and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy* 118(2): 335–408.

[84] Steffens, Lincoln. 1904. *The Shame of the Cities*. New York: McClure.

[85] Strömberg, David. 2004. "Radio's Impact on Public Spending." *Quarterly Journal of Economics* 99: 189–221.

[86] Svensson, Jakob. 2005. "Eight Questions about Corruption." *Journal of Economic Perspectives* 19 (3): 19–42.

[87] Tiebout, Charles M. 1956. "A Pure Theory of Local Public Expenditures." *Journal of Political Economy* 64 (5): 416–24.

[88] Tommasi, Mariano, and Federico Weinschelbaum. 2007. "Centralization vs. Decentralization: A Principal-Agent Analysis." *Journal of Public Economic Theory* 9(2): 369–89.

[89] Treisman, Daniel. 2000. "The Causes of Corruption: A Cross-National Study." *Journal of Public Economics* 76: 399–457.

[90] ——. 2007. "What Have We Learned About the Causes of Corruption From Ten Years of Cross-National Empirical Research?" *Annual Review of Political Science* 10: 211–44.

[91] Turner, Frederick J. 1921. *The Frontier in American History*. New York, NY: Holt.

[92] Wallis, John J. 2000. "American Government Finance in the Long Run: 1790 to 1990." *Journal of Economic Perspectives*, 14 (1): 61-82.

[93] ——. 2006. "The Concept of Systematic Corruption in American History" in *Corruption and Reform: Lessons from America's Economic History*, edited by Edward L. Glaeser and Claudia Goldin. Cambridge, MA: National Bureau of Economic Research.

[94] Wallis, John J., Price Fishback, and Shawn Kantor. 2006. "Politics, Relief, and Reform: Roosevelt's Efforts to Control Corruption and Political Manipulation during the New Deal." In *Corruption and Reform: Lessons from America's Economic History*, edited by Edward L. Glaeser and Claudia Goldin. Cambridge, MA: National Bureau of Economic Research.

[95] Whitt, Ward. 1985. "Uniform Conditional Variability Ordering of Probability Distributions." *Journal of Applied Probability* 22 (3): 619–33.

[96] Wilson, James Q. 1966. "Corruption: The Shame of the States." *The Public Interest* 2: 28–38.

[97] Ziblatt, Daniel. 2006. *Structuring the State: The Formation of Italy and Germany and the Puzzle of Federalism.* Princeton, NJ: Princeton University Press.

A Appendix

A.1. Proof of Lemma 1

Taking into account that the realizations of the uniform idiosyncratic shock ψ^i are independent across voters, the share of members of group j who vote for the incumbent conditional on the realizations of $\mathbf{g}_t$, Ψ_t and Θ_j equals

$$v_t^j\left(\mathbf{g}_t,\Psi_t,\Theta_j\right)=\Theta_j\Pr\left(\psi_t^i\leq\Delta_1^j\left(\mathbf{g}_t\right)-\Psi_t\right)+\left(1-\Theta_j\right)\Pr\left(\psi_t^i\leq-\Psi_t\right)$$
$$=\frac{1}{2}+\frac{1}{2\bar{\psi}}\left[\Theta_j\sum_{p=1}^{P}\alpha_p^j\mathbb{E}\left(\varepsilon_{p,t}|g_{p,t}\right)-\Psi_t\right]. \quad \text{(A1)}$$

Taking into account the uniform aggregate shock Ψ_t, the incumbent's probability of re-election conditional on the realizations of public-good provision $\mathbf{g}_t$ equals

$$\pi\left(\mathbf{g}_t\right)=\Pr\left(\sum_{j=1}^{J}\lambda_j v_t^j\left(\mathbf{g}_t,\Psi_t\right)\geq\frac{1}{2}\right)=\Pr\left(\Psi_t\leq\sum_{j=1}^{J}\Theta_j\lambda_j\sum_{p=1}^{P}\alpha_p^j\mathbb{E}\left(\varepsilon_{p,t}|g_{p,t}\right)\right)$$
$$=\mathbb{E}\left[\frac{1}{2}+\phi\sum_{j=1}^{J}\Theta_j\lambda_j\sum_{p=1}^{P}\alpha_p^j\mathbb{E}\left(\varepsilon_{p,t}|g_{p,t}\right)\right]=\frac{1}{2}+\phi\sum_{j=1}^{J}\theta_j\lambda_j\sum_{p=1}^{P}\alpha_p^j\mathbb{E}\left(\varepsilon_{p,t}|g_{p,t}\right)$$
$$=\frac{1}{2}+\phi\sum_{j=1}^{J}\theta_j\lambda_j\sum_{p=1}^{P}\alpha_p^j\left(\log g_{p,t}-\log\bar{x}_p-\varepsilon_{p,t-1}\right). \quad \text{(A2)}$$

Taking into account the mean-zero competence shocks $\varepsilon_{p,t}$, the incumbent's probability of re-election conditional on his policy choices $\mathbf{x}_t$ (and residually r_t) equals

$$\pi\left(\mathbf{x}_t\right)=\mathbb{E}\left[\pi\left(\mathbf{g}_t\right)|\mathbf{x}_t\right]=\frac{1}{2}+\phi\sum_{j=1}^{J}\theta_j\lambda_j\sum_{p=1}^{P}\alpha_p^j\left(\log x_{p,t}-\log\bar{x}_p\right). \quad \text{(A3)}$$

The trade-off between current rent extraction and a value R of re-election leads to policy choices

$$\mathbf{x}\left(R\right)=\arg\max_{\mathbf{x}_t}\left\{b-\sum_{p=1}^{P}x_{p,t}+R\pi\left(\mathbf{x}_t\right)\right\}, \quad \text{(A4)}$$

namely

$$x_p\left(R\right)=\phi R\sum_{j=1}^{J}\theta_j\lambda_j\alpha_p^j\text{ for all }p=1,...,P, \quad \text{(A5)}$$

and thus current rent extraction

$$r\left(R\right)=b-\phi\sum\nolimits_{j=1}^{J}\lambda_j\theta_j R. \quad \text{(A6)}$$

For ease of notation, let

$$\Phi \equiv \frac{2\delta}{2-\delta}\phi. \tag{A7}$$

By equation (11), equilibrium rent-extraction is

$$r = b\left(1 + \Phi \sum\nolimits_{j=1}^{J} \lambda_j \theta_j\right)^{-1}, \tag{A8}$$

which is decreasing and convex in θ_j.

The equilibrium allocation of resources across public goods follows the shares

$$\beta_p \equiv \frac{x_p}{(1-\rho)\, b} = \sum_{j=1}^{J} \frac{\theta_j}{\bar{\theta}} \lambda_j \alpha_p^j. \tag{A9}$$

The incumbent is re-elected if and only if

$$\Psi_t \leq \sum_{j=1}^{J} \theta_j \lambda_j \sum_{p=1}^{P} \alpha_p^j \varepsilon_{p,t}. \tag{A10}$$

Let χ_t be an indicator variable for this condition. The competence of ruling politicians evolves according to

$$\hat{\boldsymbol{\eta}}_t = \chi_{t-1}\left(\boldsymbol{\varepsilon}_{t-1}^I + \boldsymbol{\varepsilon}_t^I\right) + \left(1-\chi_{t-1}\right)\left(\boldsymbol{\varepsilon}_{t-1}^C + \boldsymbol{\varepsilon}_t^C\right), \tag{A11}$$

where the superscripts I and C refer to the incumbent and challenger in the election at the end of period $t-1$.

The cumulative distribution function of ability $\hat{\eta}_{p,t}$ is

$$\begin{aligned}
\Pr\left(\hat{\eta}_{p,t} \leq \eta\right) &= \Pr\left[\chi_{t-1}\left(\varepsilon_{p,t-1}^I + \varepsilon_{p,t}^I\right) + \left(1-\chi_{t-1}\right)\left(\varepsilon_{p,t-1}^C + \varepsilon_{p,t}^C\right) \leq \eta\right] \\
&= \Pr\left(\chi_{t-1} = 1 \wedge \varepsilon_{p,t-1}^I + \varepsilon_{p,t}^I \leq \eta\right) + \Pr\left(\chi_{t-1} = 0 \wedge \varepsilon_{p,t-1}^C + \varepsilon_{p,t}^C \leq \eta\right) \\
&= \Pr\left(\Psi_{t-1} \leq \sum_{j=1}^{J} \lambda_j \theta_j \sum_{q=1}^{P} \alpha_q^j \varepsilon_{q,t-1} \wedge \varepsilon_{p,t-1}^I + \varepsilon_{p,t}^I \leq \eta\right) + \frac{1}{2}\Pr\left(\varepsilon_{p,t-1}^C + \varepsilon_{p,t}^C \leq \eta\right) \\
&= \int_{-\infty}^{\infty} \left(1 + \varepsilon\phi \sum_{j=1}^{J} \lambda_j \theta_j \alpha_p^j\right) F_\varepsilon\left(\eta - \varepsilon\right) f_\varepsilon\left(\varepsilon\right) d\varepsilon,
\end{aligned} \tag{A12}$$

where $F_\varepsilon\left(\varepsilon\right)$ is the cumulative distribution function of $\varepsilon_{p,t}$ and $f_\varepsilon\left(\varepsilon\right)$ its probability density function. Since

$$\int_{-\infty}^{\infty} \varepsilon F_\varepsilon\left(\eta - \varepsilon\right) f_\varepsilon\left(\varepsilon\right) d\varepsilon = \mathbb{E}\left[\varepsilon F_\varepsilon\left(\eta - \varepsilon\right)\right] < \mathbb{E}\varepsilon \mathbb{E}\left[F_\varepsilon\left(\eta - \varepsilon\right)\right] = 0, \tag{A13}$$

an increase in $\sum_{j=1}^{J} \lambda_j \theta_j \alpha_p^j$ induces an increase in $\hat{\eta}_p$ in the sense of first-order stochastic dominance.

The unconditional expectation of ability $\hat{\eta}_{p,t}$ is

$$\mathbb{E}\hat{\eta}_{p,t} = \mathbb{E}\left(\chi_{t-1}\varepsilon_{p,t-1}\right) = \int_{-\infty}^{\infty}\left(\frac{1}{2} + \phi\sum_{j=1}^{J}\lambda_j\theta_j\sum_{q=1}^{P}\alpha_q^j\varepsilon_q\right)\varepsilon_p f_\varepsilon\left(\varepsilon_p\right)d\varepsilon_p$$
$$= \phi\sigma^2\sum_{j=1}^{J}\lambda_j\theta_j\alpha_p^j. \quad \text{(A14)}$$

The equilibrium utility of each member of group j equals

$$\mathbb{E}u_j = \sum_{p=1}^{P}\alpha_p^j\mathbb{E}\log g_{p,t} = \log b + \log\left(1-\rho\right) + \sum_{p=1}^{P}\alpha_p^j\left(\mathbb{E}\hat{\eta}_p + \log\beta_p\right). \quad \text{(A15)}$$

A.2. Proof of Proposition 1

In a polity composed of L regions there are LP public goods: $g_{l,p,t}$ is the provision of public good p in region l at time t. Residents of each region l derive utility from public goods in their own region only: $\alpha_{l,p}^l = \alpha_p^l$ while $\alpha_{m,p}^l = 0$ for $l \neq m$.

Under decentralization, in each region l a local politician with ability $\eta_{l,p,t}^D$ independently invests in the provision of public goods $x_{l,p,t}^D$ and extracts rent $r_{l,t}^D = b - \sum_{p=1}^{P} x_{l,p,t}^D$. Equilibrium rent extraction is

$$\rho_l^D = \left(1 + \Phi\theta_l\right)^{-1}, \quad \text{(A16)}$$

the expected ability of a local politician is

$$\mathbb{E}\hat{\eta}_{l,p}^D = \phi\sigma^2\alpha_p^l\theta_l, \quad \text{(A17)}$$

and the relative shares of each local public good are

$$\beta_{l,p}^D \equiv \frac{x_{l,p}}{\left(1-\rho_l^D\right)b} = \alpha_p^l. \quad \text{(A18)}$$

Welfare in region l is

$$\mathbb{E}u_l^D = \log b + \log\left(1-\rho_l^D\right) + \sum_{p=1}^{P}\alpha_p^l\left(\mathbb{E}\hat{\eta}_{l,p}^D + \log\beta_{l,p}^D\right), \quad \text{(A19)}$$

and aggregate welfare is

$$W^D = \log b + \log\left(1-\rho_l^D\right) + \frac{1}{L}\sum_{l=1}^{L}\sum_{p=1}^{P}\alpha_p^l\left(\mathbb{E}\hat{\eta}_{l,p}^D + \log\beta_{l,p}^D\right). \quad \text{(A20)}$$

Under centralization a single politician with ability $\eta_{p,t}^C$ chooses investment in public goods $x_{l,p,t}^C$ for all l. and extracts rents $r_t^C = bL - \sum_{l=1}^{L}\sum_{p=1}^{P} x_{l,p,t}^C$. We partition the P public goods into two sets. The set $\mathcal{U}$ consists of public goods whose centralized provision is subject to

a uniformity constraint $g^C_{l,p,t} = g^C_{p,t}$ for all l. This constraint coincides with a constraint on resource allocation $x^C_{l,p,t} = x^C_{p,t}$ for all l because ability $\eta^C_{p,t}$ is common. The complementary set $\mathcal{D}$ consists instead of public goods that the central government can provide in different amounts to different regions. Regardless of this partition, equilibrium rent extraction is

$$\rho^C = \left(1 + \Phi\bar{\theta}\right)^{-1} \text{ for } \bar{\theta} = \frac{1}{L}\sum_{l=1}^{L}\theta_l, \tag{A21}$$

and the expected ability of a central politician is

$$\mathbb{E}\hat{\eta}^C_p = \frac{\phi\sigma^2}{L}\sum_{l=1}^{L}\theta_l\alpha^l_p. \tag{A22}$$

For expositional convenience, we characterize the allocation of resources under centralization by the shares

$$\beta^C_{l,p} \equiv \frac{x^C_{l,p}}{\left(1-\rho^C\right)b} \tag{A23}$$

relative to a region's equal share of net aggregate resources, rather than to the total $\left(1-\rho^C\right)bL$. Thus, $\beta^C_{l,p}$ lies in $[0, L]$ instead of $[0, 1]$. Then relative shares of each local public good are

$$\beta^C_p = \frac{1}{L}\sum_{l=1}^{L}\frac{\theta_l}{\bar{\theta}}\alpha^l_p \text{ for } p \in \mathcal{U} \tag{A24}$$

and

$$\beta^C_{l,p} = \frac{\theta_l}{\bar{\theta}}\alpha^l_p \text{ for } p \in \mathcal{D}. \tag{A25}$$

Welfare in region l is

$$\mathbb{E}u^C_l = \log b + \log\left(1-\rho^C\right) + \sum_{p=1}^{P}\alpha^l_p\mathbb{E}\hat{\eta}^C_p + \sum_{p\in\mathcal{U}}\alpha^l_p\log\beta^C_p + \sum_{p\in\mathcal{D}}\alpha^l_p\log\beta^C_{l,p} \tag{A26}$$

and aggregate welfare is

$$W^C = \log b + \log\left(1-\rho^C\right) + \sum_{p=1}^{P}\bar{\alpha}_p\mathbb{E}\hat{\eta}^C_p + \sum_{p\in\mathcal{U}}\bar{\alpha}_p\log\beta^C_p + \frac{1}{L}\sum_{l=1}^{L}\sum_{p\in\mathcal{D}}^{P}\alpha^l_p\log\beta^C_{l,p}, \tag{A27}$$

for

$$\bar{\alpha}_p = \frac{1}{L}\sum_{l=1}^{L}\alpha^l_p. \tag{A28}$$

Letting $\mathbb{E}$ denote the expected value across a continuum of regions, aggregate welfare

under decentralization is

$$W^D = \log b + \mathbb{E}\log\frac{\Phi\theta_l}{1+\Phi\theta_l} + \phi\sigma^2\sum_{p=1}^{P}\mathbb{E}\left[\theta_l\left(\alpha_p^l\right)^2\right] + \sum_{p=1}^{P}\mathbb{E}\left(\alpha_p^l\log\alpha_p^l\right), \tag{A29}$$

while under centralization it is

$$W^C = \log b + \log\frac{\Phi\mathbb{E}\theta_l}{1+\Phi\mathbb{E}\theta_l} + \phi\sigma^2\sum_{p=1}^{P}\mathbb{E}\left(\theta_l\alpha_p^l\right)\mathbb{E}\alpha_p^l + \sum_{p\in\mathcal{U}}\mathbb{E}\alpha_p^l\log\mathbb{E}\left(\theta_l\alpha_p^l\right) + \sum_{p\in\mathcal{D}}\mathbb{E}\left[\alpha_p^l\log\left(\theta_l\alpha_p^l\right)\right] - \log\mathbb{E}\theta_l. \tag{A30}$$

The welfare comparison can be decomposed into three elements.

1. Centralization with heterogeneous information induces a reduction in rent extraction:

$$\log\left(1-\rho^C\right) = \log\frac{\Phi\mathbb{E}\theta_l}{1+\Phi\mathbb{E}\theta_l} > \mathbb{E}\log\left(1-\rho_l^D\right) = \mathbb{E}\log\frac{\Phi\theta_l}{1+\Phi\theta_l} \tag{A31}$$

   by Jensen's inequality.

2. Centralization with heterogeneous preferences induces a misallocation of ability:

$$\mathbb{E}\left(\theta_l\alpha_p^l\right)\mathbb{E}\alpha_p^l = \mathbb{E}\theta_l\left(\mathbb{E}\alpha_p^l\right)^2 < \mathbb{E}\theta_l\mathbb{E}\left[\left(\alpha_p^l\right)^2\right] = \mathbb{E}\left[\theta_l\left(\alpha_p^l\right)^2\right] \text{ for all } p \tag{A32}$$

   because information θ_l and preferences $\boldsymbol{\alpha}^l$ are independent.

3. Centralization with heterogeneous preferences and information induces a misallocation of resources:

$$\begin{aligned}&\sum_{p\in\mathcal{U}}\mathbb{E}\alpha_p^l\log\mathbb{E}\left(\theta_l\alpha_p^l\right) + \sum_{p\in\mathcal{D}}\mathbb{E}\left[\alpha_p^l\log\left(\theta_l\alpha_p^l\right)\right] - \log\mathbb{E}\theta_l = \\ &\sum_{p=1}^{P}\mathbb{E}\left(\alpha_p^l\log\alpha_p^l\right) - \sum_{p\in\mathcal{U}}\left[\mathbb{E}\left(\alpha_p^l\log\alpha_p^l\right) - \mathbb{E}\alpha_p^l\log\mathbb{E}\alpha_p^l\right] - \sum_{p\in\mathcal{D}}\mathbb{E}\alpha_p^l\left(\log\mathbb{E}\theta_l - \mathbb{E}\log\theta_l\right) \\ &< \sum_{p=1}^{P}\mathbb{E}\left(\alpha_p^l\log\alpha_p^l\right)\end{aligned} \tag{A33}$$

   because information θ_l and preferences $\boldsymbol{\alpha}^l$ are independent.

Since the distribution of preferences is symmetric across goods, it is welfare-maximizing to apply the uniformity constraint either to all or to none. If no uniformity constraint is applied ($\mathcal{U} = \varnothing$) then centralization is welfare-reducing because the gain from reduced rent-seeking is less than the loss from resource misallocation, even before taking into account the

misallocation of ability:

$$\lim_{\sigma^2 \to 0} \left(W^D - W^C \right) = \log\left(1 + \Phi \mathbb{E}\theta_l\right) - \mathbb{E}\log\left(1 + \Phi\theta_l\right) \geq 0. \tag{A34}$$

Centralization with uniformity ($\mathcal{D} = \varnothing$) is preferable to decentralization ($W^C \geq W^D$) if and only if

$$\mathbb{E}\log\left(1 + \frac{1}{\Phi\theta_l}\right) - \log\left(1 + \frac{1}{\Phi\mathbb{E}\theta_l}\right) \geq \log P + P\mathbb{E}\left(\alpha_p^l \log \alpha_p^l\right) + \phi\sigma^2 \mathbb{E}\theta_l P \operatorname{Var}\left(\alpha_p^l\right). \tag{A35}$$

For a given mean of the distribution of information $\mathbb{E}\theta_l = \bar{\theta}$, the left-hand side can be written as $\mathbb{E} f_L\left(\theta_l; \bar{\theta}\right)$ for a function

$$f_L\left(\theta_l; \bar{\theta}\right) \equiv \log\left(1 + \frac{1}{\Phi\theta_l}\right) - \log\left(1 + \frac{1}{\Phi\mathbb{E}\theta_l}\right) \tag{A36}$$

such that

$$\frac{\partial^2 f_L}{\partial \theta_l^2} = \frac{1 + 2\Phi\theta_l}{\left[\left(1 + \Phi\theta_l\right)\theta_l\right]^2} > 0. \tag{A37}$$

Therefore, a mean-preserving spread of θ_l increases the left-hand side of equation A35 while leaving the right-hand side unchanged: centralization with uniformity is then more likely to be welfare-maximizing.

The marginal distribution of preferences for p necessarily has mean $\mathbb{E}\alpha_p^l = 1/P$. The right hand side of equation A35 can be written as $\mathbb{E} f_R\left(\alpha_p^l; \bar{\theta}\right)$ for a function

$$f_R\left(\alpha_p^l; \bar{\theta}\right) \equiv P\left[\alpha_p^l \log \alpha_p^l + \bar{\theta}\phi\sigma^2\left(\alpha_p^l\right)^2\right] - \frac{\bar{\theta}\phi\sigma^2}{P} + \log P \tag{A38}$$

such that

$$\frac{\partial^2 f_R}{\partial\left(\alpha_p^l\right)^2} = P\left(\frac{1}{\alpha_p^l} + 2\bar{\theta}\phi\sigma^2\right) > 0. \tag{A39}$$

Therefore, a mean-preserving spread of α_p^l increases the right-hand side of equation A35 while leaving the left-hand side unchanged: decentralization is then more likely to be welfare-maximizing.

If $\theta_l \sim B\left(\bar{\theta}\iota, \left(1 - \bar{\theta}\right)\iota\right)$, a decrease in the homogeneity parameter $\iota > 0$ entails a mean-preserving spread of information. If $\boldsymbol{\alpha}^l$ has a symmetric Dirichlet distribution with concentration υ its marginal distribution is beta-distributed with homogeneity parameter υP: $\alpha_p^l \sim B\left(\upsilon, \upsilon\left(P - 1\right)\right)$. Thus a decrease in υ entails a mean-preserving spread of preferences.

In both cases, a decrease in the homogeneity parameter entails mean-preserving spread because a beta distribution with mean $\mu \in (0, 1)$ and homogeneity $\nu > 0$ has density

$$f\left(x; \mu, \nu\right) = \frac{1}{B\left(\mu\nu, \left(1 - \mu\right)\nu\right)} x^{\mu\nu - 1}\left(1 - x\right)^{(1-\mu)\nu - 1} \text{ for } x \in [0, 1]. \tag{A40}$$

The density ratio of two beta-distributed random variables X and Y with equal means μ and concentration parameters $\nu_X > \nu_Y$ equals

$$\frac{f(x;\mu,\nu_X)}{f(x;\mu,\nu_Y)} = \frac{B(\mu\nu_Y,(1-\mu)\nu_Y)}{B(\mu\nu_X,(1-\mu)\nu_X)}\left[x^{\mu}(1-x)^{1-\mu}\right]^{\nu_X-\nu_Y}, \tag{A41}$$

a log-concave function of x:

$$\frac{\partial^2}{\partial x^2}\log\frac{f(x;\mu,\nu_X)}{f(x;\mu,\nu_Y)} = -(\nu_X-\nu_Y)\left[\frac{\mu}{x^2}+\frac{1-\mu}{(1-x)^2}\right] < 0 \tag{A42}$$

Therefore, Y is a mean-preserving spread of X (Whitt 1985).

In the limit as $\iota \to 0$, the distribution of θ_l converges to a Bernoulli distribution with $\Pr(\theta_l = 1) = \bar{\theta}$. In the limit as $\iota \to \infty$, θ_l converges to the deterministic value $\bar{\theta}$. Thus the left-hand side of equation A35 is monotone decreasing in ι from infinity to zero.

In the limit as $\upsilon \to 0$, the distribution of α_p^l converges to a Bernoulli distribution with $\Pr\left(\alpha_p^l = 1\right) = 1/P$. In the limit as $\upsilon \to \infty$, α_p^l converges to the deterministic value $1/P$. Thus the right-hand side of equation A35 is monotone decreasing in υ from $\log P + \bar{\theta}\phi\sigma^2(1-1/P)$ to zero

As a consequence, there exists a finite threshold $\bar{\upsilon}(\iota,\sigma) > 0$ such that centralization with uniformity is preferable to decentralization if and only if $\upsilon \geq \bar{\upsilon}$. The threshold is increasing in ι. It is increasing in σ^2 because so is the right-hand side of equation A35.

A.3. Proof of Proposition 2

The division of powers is described by two indicator variables: $\chi_0 = 1$ if and only if the central government is tasked with providing the homogeneously desired good; $\chi_1 = 1$ if and only if it provides the idiosyncratically preferred good.

From equations (A5) and (11), equilibrium rent extraction by a local politician in region l is

$$\rho_l^D = \left\{1 + \Phi\theta_l\left[(1-\chi_0)\alpha_0 + (1-\chi_1)(1-\alpha_0)\right]\right\}^{-1}. \tag{A43}$$

The politician's expected abilities are

$$\mathbb{E}\hat{\eta}_{l,0}^D = (1-\chi_0)\alpha_0\phi\sigma^2\theta_l \text{ and } \mathbb{E}\hat{\eta}_{l,l}^D = (1-\chi_1)(1-\alpha_0)\phi\sigma^2\theta_l, \tag{A44}$$

and $\mathbb{E}\hat{\eta}_{l,m}^D = 0$ for all $m \neq l$. He chooses shares

$$\beta_{l,0}^D = \frac{(1-\chi_0)\alpha_0}{(1-\chi_0)\alpha_0 + (1-\chi_1)(1-\alpha_0)} \tag{A45}$$

and

$$\beta_{l,l}^D = \frac{(1-\chi_1)(1-\alpha_0)}{(1-\chi_0)\alpha_0 + (1-\chi_1)(1-\alpha_0)},\ \beta_{l,m}^D = 0 \text{ for all } m \neq l \tag{A46}$$

for the allocation of his budget $b_D = b - b_D/L$.

Equilibrium rent extraction by a central politician is

$$\rho^C = \left\{1 + \Phi\bar{\theta}\left[\chi_0\alpha_0 + \chi_1\left(1-\alpha_0\right)\right]\right\}^{-1}. \tag{A47}$$

His expected abilities are

$$\mathbb{E}\hat{\eta}_0^C = \chi_0\alpha_0\phi\sigma^2\bar{\theta} \text{ and } \mathbb{E}\hat{\eta}_l^C = \chi_1\left(1-\alpha_0\right)\phi\sigma^2\frac{\theta_l}{L} \text{ for } l = 1, 2, ..., L. \tag{A48}$$

His budget shares given his budget b_C are defined again with the convention that

$$\beta_{l,p}^C \equiv \frac{x_{l,p}^C}{\left(1-\rho^C\right)b_C}. \tag{A49}$$

If he is entrusted with providing the homogeneously desired good he chooses a budget share

$$\beta_0^C = \frac{\chi_0\alpha_0}{\chi_0\alpha_0 + \chi_1\left(1-\alpha_0\right)} \text{ if } 0 \in \mathcal{U}, \tag{A50}$$

or budget shares

$$\beta_{l,0}^C = \frac{\chi_0\alpha_0}{\chi_0\alpha_0 + \chi_1\left(1-\alpha_0\right)}\frac{\theta_l}{\bar{\theta}} \text{ if } 0 \in \mathcal{D}. \tag{A51}$$

If he is entrusted with providing the idiosyncratically preferred good, he sets a budget share

$$\beta_l^C = \frac{\chi_1\left(1-\alpha_0\right)}{\chi_0\alpha_0 + \chi_1\left(1-\alpha_0\right)}\frac{1}{L}\frac{\theta_l}{\bar{\theta}} \text{ if } l \in \mathcal{U}, \tag{A52}$$

or budget shares

$$\beta_{l,l}^C = \frac{\chi_1\left(1-\alpha_0\right)}{\chi_0\alpha_0 + \chi_1\left(1-\alpha_0\right)}\frac{\theta_l}{\bar{\theta}} \text{ and } \beta_{m,l}^C = 0 \text{ for all } m \neq l \text{ if } l \in \mathcal{D}. \tag{A53}$$

Welfare in region l can be decomposed into four components

$$\mathbb{E}u_l \equiv u_l^b + u_l^\beta + u_l^\rho + \mathbb{E}u_l^\eta. \tag{A54}$$

The allocation of resources between the two levels of government has a welfare impact

$$u_l^b = \left[\left(1-\chi_0\right)\alpha_0 + \left(1-\chi_1\right)\left(1-\alpha_0\right)\right]\log\left(b - \frac{b_C}{L}\right) + \left[\chi_0\alpha_0 + \chi_1\left(1-\alpha_0\right)\right]\log\frac{b_C}{L}. \tag{A55}$$

The allocation of each government's budget has a welfare impact

$$u_l^\beta = \left(1-\chi_0\right)\alpha_0\log\beta_{l,0}^D + \left(1-\chi_1\right)\left(1-\alpha_0\right)\log\beta_{l,l}^D + \chi_0\alpha_0\log\beta_{l,0}^C + \chi_1\left(1-\alpha_0\right)\log\beta_{l,l}^C. \tag{A56}$$

Rent extraction by the different levels of government has a welfare impact

$$u_l^{\rho} = \left[(1-\chi_0)\,\alpha_0 + (1-\chi_1)\,(1-\alpha_0)\right]\log\left(1-\rho_l^D\right) + \left[\chi_0\alpha_0 + \chi_1\,(1-\alpha_0)\right]\log\left(1-\rho^C\right). \quad \text{(A57)}$$

The selection of politicians according to their skills has a welfare impact

$$\mathbb{E}u_l^{\eta} = (1-\chi_0)\,\alpha_0\mathbb{E}\hat{\eta}_{l,0}^D + (1-\chi_1)\,(1-\alpha_0)\,\mathbb{E}\hat{\eta}_{l,l}^D + \chi_0\alpha_0\mathbb{E}\hat{\eta}_0^C + \chi_1\,(1-\alpha_0)\,\mathbb{E}\hat{\eta}_l^C. \quad \text{(A58)}$$

The allocation of the budget between the two levels of government affects welfare only through the therm u_l^b. Every region desires the unique Pareto efficient allocation

$$b_C^* = \arg\max u_l^b\,(b_C) = \left[\chi_0\alpha_0 + \chi_1\,(1-\alpha_0)\right]bL, \quad \text{(A59)}$$

such that the local-government budget is

$$b_D^* = \left[(1-\chi_0)\,\alpha_0 + (1-\chi_1)\,(1-\alpha_0)\right]b. \quad \text{(A60)}$$

Uniformity constraints affect welfare only through the therm u_l^{β}. If $\chi_0 = 1$, imposing a uniformity constraint on centralized provision of the homogeneously desired public good increases aggregate social welfare by

$$\alpha_0\left(\log\bar{\theta} - \frac{1}{L}\sum_{l=1}^{L}\log\theta_l\right) \geq 0. \quad \text{(A61)}$$

If $\chi_1 = 1$, imposing a uniformity constraint on centralized provision of the idiosyncratically preferred public good reduces welfare in every region by

$$-\,(1-\alpha_0)\log L \leq 0. \quad \text{(A62)}$$

With the efficient central-government budget and the welfare-maximizing uniformity constraints,

$$u_l^b + u_l^{\beta} = \log b + \alpha_0\log\alpha_0 + (1-\alpha_0)\log(1-\alpha_0) + \chi_1\,(1-\alpha_0)\left(\log\theta_l - \log\bar{\theta}\right). \quad \text{(A63)}$$

With equilibrium rent extraction,

$$u_l^{\rho} = \left[(1-\chi_0)\,\alpha_0 + (1-\chi_1)\,(1-\alpha_0)\right]\log\frac{\left[(1-\chi_0)\,\alpha_0 + (1-\chi_1)\,(1-\alpha_0)\right]\Phi\theta_l}{1+\left[(1-\chi_0)\,\alpha_0 + (1-\chi_1)\,(1-\alpha_0)\right]\Phi\theta_l} + \left[\chi_0\alpha_0 + \chi_1\,(1-\alpha_0)\right]\log\frac{\left[\chi_0\alpha_0 + \chi_1\,(1-\alpha_0)\right]\Phi\bar{\theta}}{1+\left[\chi_0\alpha_0 + \chi_1\,(1-\alpha_0)\right]\Phi\bar{\theta}}. \quad \text{(A64)}$$

With the equilibrium skill of incumbent politicians,

$$\mathbb{E}u_l^{\eta} = \phi\sigma^2\left\{\alpha_0^2\left[(1-\chi_0)\,\theta_l + \chi_0\bar{\theta}\right] + (1-\alpha_0)^2\left(1 - \frac{L-1}{L}\chi_1\right)\theta_l\right\}. \quad \text{(A65)}$$

Abstracting from differences between sample distributions and population distributions thanks to the assumption of a continuum of regions ($L \to \infty$), aggregate social welfare is

$$\begin{aligned} W = \log b + \alpha_0 \log \alpha_0 + (1-\alpha_0) \log (1-\alpha_0) - \chi_1 (1-\alpha_0) (\log \mathbb{E}\theta_l - \mathbb{E} \log \theta_l) \\ + \left[(1-\chi_0) \alpha_0 + (1-\chi_1)(1-\alpha_0) \right] \mathbb{E} \log \frac{\left[(1-\chi_0) \alpha_0 + (1-\chi_1)(1-\alpha_0) \right] \Phi \theta_l}{1 + \left[(1-\chi_0) \alpha_0 + (1-\chi_1)(1-\alpha_0) \right] \Phi \theta_l} \\ + \left[\chi_0 \alpha_0 + \chi_1 (1-\alpha_0) \right] \log \frac{\left[\chi_0 \alpha_0 + \chi_1 (1-\alpha_0) \right] \Phi \mathbb{E} \theta_l}{1 + \left[\chi_0 \alpha_0 + \chi_1 (1-\alpha_0) \right] \Phi \mathbb{E} \theta_l} \\ + \left[\alpha_0^2 + (1-\chi_1)(1-\alpha_0)^2 \right] \phi \sigma^2 \mathbb{E} \theta_l. \quad \text{(A66)} \end{aligned}$$

Under full decentralization ($\chi_0 = \chi_1 = 0$) welfare is

$$\begin{aligned} W_D = \log b + \alpha_0 \log \alpha_0 + (1-\alpha_0) \log (1-\alpha_0) \\ + \mathbb{E} \log \frac{\Phi \theta_l}{1 + \Phi \theta_l} + \left[\alpha_0^2 + (1-\alpha_0)^2 \right] \phi \sigma^2 \mathbb{E} \theta_l. \quad \text{(A67)} \end{aligned}$$

Under a federal system ($\chi_0 = 1$ and $\chi_1 = 0$) it is

$$\begin{aligned} W_F = \log b + \alpha_0 \log \alpha_0 + (1-\alpha_0) \log (1-\alpha_0) \\ + (1-\alpha_0) \mathbb{E} \log \frac{(1-\alpha_0) \Phi \theta_l}{1 + (1-\alpha_0) \Phi \theta_l} + \alpha_0 \log \frac{\alpha_0 \Phi \mathbb{E} \theta_l}{1 + \alpha_0 \Phi \mathbb{E} \theta_l} + \left[\alpha_0^2 + (1-\alpha_0)^2 \right] \phi \sigma^2 \mathbb{E} \theta_l. \quad \text{(A68)} \end{aligned}$$

Under full centralization ($\chi_0 = \chi_1 = 1$) it is

$$\begin{aligned} W_C = \log b + \alpha_0 \log \alpha_0 + (1-\alpha_0) \log (1-\alpha_0) \\ - (1-\alpha_0) (\log \mathbb{E}\theta_l - \mathbb{E} \log \theta_l) + \log \frac{\Phi \mathbb{E} \theta_l}{1 + \Phi \mathbb{E} \theta_l} + \alpha_0^2 \phi \sigma^2 \mathbb{E} \theta_l. \quad \text{(A69)} \end{aligned}$$

Under a reverse federal system ($\chi_0 = 0$ and $\chi_1 = 1$) welfare would be

$$\begin{aligned} W_{-F} = \log b + \alpha_0 \log \alpha_0 + (1-\alpha_0) \log (1-\alpha_0) - (1-\alpha_0) (\log \mathbb{E}\theta_l - \mathbb{E} \log \theta_l) \\ + \alpha_0 \mathbb{E} \log \frac{\alpha_0 \Phi \theta_l}{1 + \alpha_0 \Phi \theta_l} + (1-\alpha_0) \log \frac{(1-\alpha_0) \Phi \mathbb{E} \theta_l}{1 + (1-\alpha_0) \Phi \mathbb{E} \theta_l} + \alpha_0^2 \phi \sigma^2 \mathbb{E} \theta_l < W_C, \quad \text{(A70)} \end{aligned}$$

so this arrangement is dominated by full centralization.

To compare the three undominated government structures, it is convenient to rescale welfare by an additive constant $\log b + \alpha_0 \log \alpha_0 + (1-\alpha_0) \log (1-\alpha_0) + \left[\alpha_0^2 + (1-\chi_1)(1-\alpha_0)^2 \right] \phi \sigma^2 \mathbb{E} \theta_l$. Then welfare under full decentralization is

$$W_D = \mathbb{E} \log \frac{\Phi \theta_l}{1 + \Phi \theta_l} \quad \text{(A71)}$$

independent of α_0 up to the rescaling.

Welfare under a federal system is

$$W_F = (1-\alpha_0)\,\mathbb{E}\log\frac{(1-\alpha_0)\,\Phi\theta_l}{1+(1-\alpha_0)\,\Phi\theta_l} + \alpha_0\log\frac{\alpha_0\Phi\mathbb{E}\theta_l}{1+\alpha_0\Phi\mathbb{E}\theta_l}, \tag{A72}$$

with limits

$$\lim_{\alpha_0\to 0} W_F = \mathbb{E}\log\frac{\Phi\theta_l}{1+\Phi\theta_l} < \lim_{\alpha_0\to 1} W_F = \log\frac{\Phi\mathbb{E}\theta_l}{1+\Phi\mathbb{E}\theta_l}. \tag{A73}$$

Its derivative with respect to α_0 is

$$\frac{\partial W_F}{\partial\alpha_0} = -\mathbb{E}\left[\log\frac{(1-\alpha_0)\,\Phi\theta_l}{1+(1-\alpha_0)\,\Phi\theta_l} + \frac{1}{1+(1-\alpha_0)\,\Phi\theta_l}\right] + \log\frac{\alpha_0\Phi\mathbb{E}\theta_l}{1+\alpha_0\Phi\mathbb{E}\theta_l} + \frac{1}{1+\alpha_0\Phi\mathbb{E}\theta_l}, \tag{A74}$$

with limits

$$\lim_{\alpha_0\to 0}\frac{\partial W_F}{\partial\alpha_0} = -\infty \text{ and } \lim_{\alpha_0\to 1}\frac{\partial W_F}{\partial\alpha_0} = +\infty. \tag{A75}$$

It is a globally convex function of α_0:

$$\frac{\partial^2 W_F}{\partial\alpha_0^2} = \frac{1}{1-\alpha_0}\mathbb{E}\left[1+(1-\alpha_0)\,\Phi\theta_l\right]^{-2} + \frac{1}{\alpha_0}\left(1+\alpha_0\Phi\mathbb{E}\theta_l\right)^{-2} > 0. \tag{A76}$$

Welfare under full centralization ($\chi_0 = \chi_1 = 1$) is

$$W_C = \log\frac{\Phi\mathbb{E}\theta_l}{1+\Phi\mathbb{E}\theta_l} - (1-\alpha_0)\left(\log\mathbb{E}\theta_l - \mathbb{E}\log\theta_l\right) - (1-\alpha_0)^2\,\phi\sigma^2\mathbb{E}\theta_l, \tag{A77}$$

with limits

$$\lim_{\alpha_0\to 0} W_C = \log\frac{\Phi\mathbb{E}\theta_l}{1+\Phi\mathbb{E}\theta_l} - \log\mathbb{E}\theta_l + \mathbb{E}\log\theta_l - \phi\sigma^2\mathbb{E}\theta_l < \lim_{\alpha_0\to 1} W_C = \log\frac{\Phi\mathbb{E}\theta_l}{1+\Phi\mathbb{E}\theta_l}. \tag{A78}$$

It is a monotone increasing and concave function of α_0:

$$\frac{\partial W_C}{\partial\alpha_0} = \log\mathbb{E}\theta_l - \mathbb{E}\log\theta_l + 2\,(1-\alpha_0)\,\phi\sigma^2\mathbb{E}\theta_l > 0 > \frac{\partial^2 W_C}{\partial\alpha_0^2} = -2\phi\sigma^2\mathbb{E}\theta_l. \tag{A79}$$

Its first derivative has limits

$$\lim_{\alpha_0\to 0}\frac{\partial W_C}{\partial\alpha_0} = \log\mathbb{E}\theta_l - \mathbb{E}\log\theta_l + 2\phi\sigma^2\mathbb{E}\theta_l > \lim_{\alpha_0\to 1}\frac{\partial W_C}{\partial\alpha_0} = \log\mathbb{E}\theta_l - \mathbb{E}\log\theta_l. \tag{A80}$$

There is a threshold $\bar{\alpha}_{D\sim C}\in(0,1)$ defined by $W_C\left(\bar{\alpha}_{D\sim C}\right) = W_D$ such that complete centralization yields higher welfare than complete decentralization if and only if $\alpha > \bar{\alpha}_{D\sim C}$. There is a second threshold $\bar{\alpha}_{D\sim F}\in(0,1)$ defined by $\bar{\alpha}_{D\sim F} > 0$ and $W_F\left(\bar{\alpha}_{D\sim F}\right) = W_D$

such that a federal allocation of powers yields higher welfare than complete decentralization if and only if $\alpha_0 > \bar{\alpha}_{D\sim F}$. There is a threshold $\bar{\alpha}_{F\sim C} \in (0,1)$ defined by $\bar{\alpha}_{F\sim C} < 1$ and $W_C(\bar{\alpha}_{F\sim C}) = W_F(\bar{\alpha}_{F\sim C})$ such that complete centralization yields higher welfare than a federal allocation of powers if and only if $\alpha_0 > \bar{\alpha}_{F\sim C}$.

Since W_D is independent of α_0, $W_F(\alpha_0)$ convex and $W_C(\alpha_0)$ concave, with $W_D(0) = W_F(0) > W_C(0)$ and $W_F(1) = W_C(1) > W_D(1)$, two cases are possible:

1. If $\bar{\alpha}_{D\sim F} < \bar{\alpha}_{D\sim C} < \bar{\alpha}_{F\sim C}$ then complete decentralization is optimal for $\alpha_0 \in [0, \bar{\alpha}_{D\sim F}]$, a federal allocation of powers for $\alpha_0 \in [\bar{\alpha}_{D\sim F}, \bar{\alpha}_{F\sim C}]$, and complete centralization for $\alpha_0 \in [\bar{\alpha}_{F\sim C}, 1]$.

2. If $\bar{\alpha}_{F\sim C} \leq \bar{\alpha}_{D\sim C} \leq \bar{\alpha}_{D\sim F}$ then complete decentralization is optimal for $\alpha_0 \in [0, \bar{\alpha}_{D\sim C}]$ and complete decentralization for $\alpha_0 \in [\bar{\alpha}_{D\sim C}, 1]$, while a federal allocation of powers is dominated.

For a given mean of the distribution of information $\mathbb{E}\theta_l = \bar{\theta}$, the definition of $\bar{\alpha}_{D\sim F}$ can be written $\mathbb{E}f_{D\sim F}\left(\theta_l, \bar{\alpha}_{D\sim F}; \bar{\theta}\right) = 0$, where

$$f_{D\sim F}\left(\theta_l, \alpha; \bar{\theta}\right) \equiv (1-\alpha)\log\frac{(1-\alpha)\Phi\theta_l}{1+(1-\alpha)\Phi\theta_l} + \alpha\log\frac{\alpha\Phi\bar{\theta}}{1+\alpha\Phi\bar{\theta}} - \log\frac{\Phi\theta_l}{1+\Phi\theta_l}, \tag{A81}$$

such that

$$\frac{\partial^2 f_{D\sim F}}{\partial\theta_l^2} = \alpha\frac{1+2(2-\alpha)\Phi\theta_l + 3(1-\alpha)(\Phi\theta_l)^2}{\{\theta_l(1+\Phi\theta_l)[1+(1-\alpha)\Phi\theta_l]\}^2} > 0. \tag{A82}$$

Therefore, a mean-preserving spread of θ_l increases $\mathbb{E}f_{D\sim F}\left(\theta_l, \bar{\alpha}_{D\sim F}; \bar{\theta}\right)$. At the same time, $\partial\mathbb{E}f_{D\sim F}\left(\theta_l, \bar{\alpha}_{D\sim F}; \bar{\theta}\right)/\partial\alpha > 0$ because $\partial W_F(\bar{\alpha}_{D\sim F})/\partial\alpha > 0 = \partial W_D/\partial\alpha$. Hence, $\partial\bar{\alpha}_{D\sim F}/\partial\iota > 0$.

The definition of $\bar{\alpha}_{F\sim C}$ can be written $\mathbb{E}f_{F\sim C}\left(\theta_l, \bar{\alpha}_{F\sim C}; \bar{\theta}, \sigma\right) = 0$, where

$$\begin{aligned} f_{F\sim C}\left(\theta_l, \alpha; \bar{\theta}, \sigma\right) \equiv \log\frac{\Phi\bar{\theta}}{1+\Phi\bar{\theta}} - (1-\alpha)\left(\log\bar{\theta} - \log\theta_l\right) - (1-\alpha)^2\phi\sigma^2\bar{\theta} \\ - (1-\alpha)\,\mathbb{E}\log\frac{(1-\alpha)\Phi\theta_l}{1+(1-\alpha)\Phi\theta_l} - \alpha\log\frac{\alpha\Phi\bar{\theta}}{1+\alpha\Phi\bar{\theta}}, \end{aligned} \tag{A83}$$

such that

$$\frac{\partial^2 f_{F\sim C}}{\partial\theta_l^2} = -\frac{(1-\alpha)^3\Phi^2}{[1+(1-\alpha)\Phi\theta_l]^2} < 0. \tag{A84}$$

Therefore, a mean-preserving spread of θ_l decreases $\mathbb{E}f_{F\sim C}\left(\theta_l, \bar{\alpha}_{F\sim C}; \bar{\theta}, \sigma\right)$. At the same time, $\partial\mathbb{E}f_{F\sim C}\left(\theta_l, \bar{\alpha}_{F\sim C}; \bar{\theta}, \sigma\right)/\partial\alpha > 0$ because $\partial W_C(\bar{\alpha}_{F\sim C}) > \partial W_F(\bar{\alpha}_{F\sim C})$. Hence, $\partial\bar{\alpha}_{F\sim C}/\partial\iota < 0$.

In the limit as $\iota \to \infty$ information becomes perfectly homogeneous ($\theta_l = \bar{\theta}$ for all l), so

$$\lim_{\iota\to\infty} W_D = \log\frac{\Phi\bar{\theta}}{1+\Phi\bar{\theta}}, \tag{A85}$$

while

$$\lim_{\iota\to\infty} W_F = (1-\alpha_0)\log\frac{(1-\alpha_0)\,\Phi\bar{\theta}}{1+(1-\alpha_0)\,\Phi\bar{\theta}} + \alpha_0\log\frac{\alpha_0\Phi\bar{\theta}}{1+\alpha_0\Phi\bar{\theta}}, \tag{A86}$$

which is symmetric around its minimum $\alpha_0 = 1/2$, and

$$\lim_{\iota\to\infty} W_C = \log\frac{\Phi\bar{\theta}}{1+\Phi\bar{\theta}} - (1-\alpha_0)^2\,\phi\sigma^2\bar{\theta}. \tag{A87}$$

Thus

$$\lim_{\iota\to\infty}\bar{\alpha}_{D\sim C} = \lim_{\iota\to\infty}\bar{\alpha}_{D\sim F} = 1 > \lim_{\iota\to\infty}\bar{\alpha}_{F\sim C}. \tag{A88}$$

In the limit as $\iota\to 0$ information becomes maximally heterogeneous ($\Pr(\theta_l = 1) = \bar{\theta}$ and $\Pr(\theta_l = 0) = 1-\bar{\theta}$). Then $\lim_{\iota\to 0} W_D = \lim_{\iota\to 0} W_F = \lim_{\iota\to 0} W_C = -\infty$, with well-defined ratios

$$\lim_{\iota\to 0}\frac{W_F}{W_D} = \lim_{\iota\to 0}\frac{W_C}{W_D} = 1-\alpha < \lim_{\iota\to 0}\frac{W_C}{W_F} = 1. \tag{A89}$$

Intuitively, a fraction $1-\bar{\theta}$ of regions unavoidably tend towards no provision of their ideal variety of the idiosyncratically preferred public good, but they also tend towards no provision of the homogeneously desired good if and only if its provision is decentralized. Thus

$$\lim_{\iota\to 0}\bar{\alpha}_{D\sim F} = \lim_{\iota\to 0}\bar{\alpha}_{D\sim C} = 0 < \lim_{\iota\to\infty}\bar{\alpha}_{F\sim C}. \tag{A90}$$

Thus, there exists a finite threshold $\bar{\iota}(\sigma) > 0$ such that $\bar{\alpha}_{F\sim C} \leq \bar{\alpha}_{D\sim C} \leq \bar{\alpha}_{D\sim F}$ if and only if $\iota \geq \bar{\iota}$. The threshold is increasing in σ because an increase in σ shifts down W_C while leaving W_D and W_F unaffected. Hence, $\partial\bar{\alpha}_{F\sim C}/\partial\sigma > 0$ and and $\partial\bar{\alpha}_{D\sim C}/\partial\sigma > 0$, while $\partial\bar{\alpha}_{D\sim F}/\partial\sigma = 0$.

A.4. Proof of Corollary 1

In a federal system $\chi_0 = 1$ and $\chi_1 = 0$. Therefore, equilibrium rent extraction is

$$\rho^C = \left(1+\alpha_0\Phi\bar{\theta}\right)^{-1} \text{ and } \rho_l^D = \left[1+(1-\alpha_0)\,\Phi\theta_l\right]^{-1}. \tag{A91}$$

The expected skills of incumbents are

$$\mathbb{E}\hat{\eta}_0^C = \alpha_0\phi\sigma^2\bar{\theta} \text{ and } \mathbb{E}\hat{\eta}_{l,l}^D = (1-\alpha_0)\,\phi\sigma^2\theta_l, \text{ while } \mathbb{E}\hat{\eta}_l^C = \mathbb{E}\hat{\eta}_{l,0}^D = 0. \tag{A92}$$

The efficient budget allocation is

$$b_C^* = \alpha_0 bL \text{ and } b_D^* = (1-\alpha_0)\,b. \tag{A93}$$

Aggregate rent extraction is

$$\bar{\rho}^F = \alpha_0\rho^C + (1-\alpha_0)\,\mathbb{E}\rho_l^D, \tag{A94}$$

such that

$$\frac{\partial \bar{\rho}^F}{\partial \alpha_0} = \left(1+\alpha_0\Phi\bar{\theta}\right)^{-2} - \mathbb{E}\left\{\left[1+(1-\alpha_0)\,\Phi\theta_l\right]^{-2}\right\} = \left(\rho^C\right)^2 - \mathbb{E}\left[\left(\rho_l^D\right)^2\right] \tag{A95}$$

and

$$\frac{\partial^2 \bar{\rho}^F}{\partial \alpha_0^2} = 2\left[\rho^C \frac{\partial \rho^C}{\partial \alpha_0} - \mathbb{E}\left(\rho_l^D \frac{\partial \rho_l^D}{\partial \alpha_0}\right)\right] < 0. \tag{A96}$$

Thus, aggregate rent extraction $\bar{\rho}^F$ reaches a maximum at $\check{\alpha}_0$ such that

$$\left(1+\check{\alpha}_0\Phi\bar{\theta}\right)^{-2} = \mathbb{E}\left\{\left[1+(1-\check{\alpha}_0)\,\Phi\theta_l\right]^{-2}\right\}. \tag{A97}$$

For a given mean of the distribution of information, the definition of $\check{\alpha}_0$ can be written $\mathbb{E}f_F\left(\theta_l, \check{\alpha}_0; \bar{\theta}\right) = 0$, where

$$f_F\left(\theta_l, \alpha; \bar{\theta}\right) \equiv \left[1+(1-\alpha)\,\Phi\theta_l\right]^{-2} - \left(1+\alpha\Phi\bar{\theta}\right)^{-2}, \tag{A98}$$

such that

$$\frac{\partial^2 f_F}{\partial \theta_l^2} = 6\left[(1-\alpha)\,\Phi\right]^2\left[1+(1-\alpha)\,\Phi\theta_l\right]^{-4} > 0. \tag{A99}$$

Therefore, a mean-preserving spread of θ_l increases $\mathbb{E}f_F\left(\theta_l, \alpha; \bar{\theta}\right)$. At the same time, $\partial \mathbb{E}f_F(\theta_l, \alpha; \bar{\theta})/\partial\alpha > 0$. Hence $\partial\check{\alpha}_0/\partial\iota > 0$. In the limit case of no heterogeneity, $\lim_{\iota\to\infty}\check{\alpha}_0 = 1/2$. In the limit case of maximum heterogeneity $\lim_{\iota\to 0}\check{\alpha}_0 > 0$ because the threshold satisfies $\left(1+\check{\alpha}_0\Phi\bar{\theta}\right)^{-2} = \left(1-\bar{\theta}\right) + \bar{\theta}\left[1+(1-\check{\alpha}_0)\,\Phi\right]^{-2}$.

A mean-preserving spread of θ_l also increases average rent extraction by local governments

$$\mathbb{E}\rho_l^D = \mathbb{E}\left\{\left[1+(1-\alpha_0)\,\Phi\theta_l\right]^{-1}\right\} \tag{A100}$$

because ρ_l^D is a convex function of θ_l. It does not affect ρ^C. Therefore, $\partial\mathbb{E}\rho_l^D/\partial\iota < 0$ and $\partial\bar{\rho}^F/\partial\iota < 0$.

A.5. Proof of Proposition 3

Let $\left(\lambda_{IL}^l, \lambda_{UL}^l, \lambda_{IR}^l, \lambda_{UR}^l\right)$ denote the relative shares of the four groups in region l's population: $\lambda_{ip}^l \equiv \lambda_{l,i,p}/\sum_{i,p}\lambda_{l,i,p}$. Taking into account rent extraction and the resolution of distributional conflict, the equilibrium allocation of resources to each public good $p \in \{L, R\}$ in region l is

$$x_{l,p} = \frac{b\Phi\left(\theta_I\lambda_{Ip}^l + \theta_U\lambda_{Up}^l\right)}{1+\Phi\left[\theta_I\left(\lambda_{IL}^l + \lambda_{IR}^l\right) + \theta_U\left(\lambda_{UL}^l + \lambda_{UR}^l\right)\right]}. \tag{A101}$$

The regional government has expected competence at providing each public good equal to

$$\mathbb{E}\hat{\eta}_{l,p} = \phi\sigma^2\left(\theta_I\lambda_{Ip}^l + \theta_U\lambda_{Up}^l\right). \tag{A102}$$

The expected utility of a resident of region l with partisan preferences $p \in \{L, R\}$ is

$$\mathbb{E}u_p^l = \log x_{l,p} + \mathbb{E}\hat{\eta}_{l,p} \tag{A103}$$

whose derivatives with respect to the shares of like-minded residents are

$$\frac{\partial \mathbb{E}u_p^l}{\partial \lambda_{ip}^l} = \frac{\theta_i}{\theta_I \lambda_{Ip}^l + \theta_U \lambda_{Up}^l} \frac{1 + \Phi\left(\theta_I \lambda_{I\neg p}^l + \theta_U \lambda_{U\neg p}^l\right)}{1 + \Phi\left[\theta_I\left(\lambda_{IL}^l + \lambda_{IR}^l\right) + \theta_U\left(\lambda_{UL}^l + \lambda_{UR}^l\right)\right]} + \phi\sigma^2\theta_i > 0 \text{ for } i \in \{I, U\} \tag{A104}$$

and with respects to the shares of opposite partisans

$$\frac{\partial \mathbb{E}u_p^l}{\partial \lambda_{i\neg p}^l} = -\frac{\Phi\theta_i}{1 + \Phi\left[\theta_I\left(\lambda_{IL}^l + \lambda_{IR}^l\right) + \theta_U\left(\lambda_{UL}^l + \lambda_{UR}^l\right)\right]} < 0 \text{ for } i \in \{I, U\}. \tag{A105}$$

Thus, any Pareto-efficient unconstrained partition is perfectly separated by preferences: $n_{IL}^l = n_{UL}^l = 0$ or $n_{IR}^l = n_{IR}^l = 0$.

Welfare in region l with homogeneous preferences p and a share λ_I^l of better-informed voters is

$$\mathbb{E}u_p^l = \log \frac{b\Phi\left[\theta_U + (\theta_I - \theta_U)\lambda_I^l\right]}{1 + \Phi\left[\theta_U + (\theta_I - \theta_U)\lambda_I^l\right]} + \phi\sigma^2\left[\theta_U + (\theta_I - \theta_U)\lambda_I^l\right], \tag{A106}$$

such that

$$\frac{\partial \mathbb{E}u_p^l}{\partial \lambda_I^l} = (\theta_I - \theta_U)\left(\frac{1}{\left[\theta_U + (\theta_I - \theta_U)\lambda_I^l\right]\left\{1 + \Phi\left[\theta_U + (\theta_I - \theta_U)\lambda_I^l\right]\right\}} + \phi\sigma^2\right) > 0 \tag{A107}$$

and

$$\frac{\partial^2 \mathbb{E}u_p^l}{\partial\left(\lambda_I^l\right)^l} = -\frac{(\theta_I - \theta_U)^2\left\{1 + 2\Phi\left[\theta_U + (\theta_I - \theta_U)\lambda_I^l\right]\right\}}{\left[\theta_U + (\theta_I - \theta_U)\lambda_I^l\right]^2\left\{1 + \Phi\left[\theta_U + (\theta_I - \theta_U)\lambda_I^l\right]\right\}^2} < 0. \tag{A108}$$

Thus, the welfare-maximizing unconstrained partition equalizes the share of better-informed voters across regions with the same preferences.

A.6. Proof of Proposition 4

Let the total population be exogenously distributed into regions $l \in \{1, 2\}$ and preferences $p \in \{L, R\}$ according to the probability distribution $P_{l,p}$. Let the average information of each group be $\theta_{l,r}$. Under separation, the expected utility of each citizen is

$$\mathbb{E}u_{l,p}^S = \log b + \log \frac{\Phi\mathbb{E}(\theta|l)}{1 + \Phi\mathbb{E}(\theta|l)} + \log\left[P(p|l)\frac{\theta_{l,p}}{\mathbb{E}(\theta|l)}\right] + \phi\sigma^2 P(p|l)\theta_{l,p}, \tag{A109}$$

while under integration it is

$$\mathbb{E}u_{l,p}^I = \log b + \log \frac{\Phi\mathbb{E}\theta}{1 + \Phi\mathbb{E}\theta} + \log\left[P(p)\frac{\mathbb{E}(\theta|p)}{\mathbb{E}\theta}\right] + \phi\sigma^2 P(p)\mathbb{E}(\theta|p). \tag{A110}$$

Thus, welfare under separation is

$$W_S = \log b + \mathbb{E}\log\frac{\Phi\mathbb{E}\left(\theta|l\right)}{1+\Phi\mathbb{E}\left(\theta|l\right)} + \mathbb{E}\log P\left(p|l\right) + \mathbb{E}\log\theta - \mathbb{E}\log\mathbb{E}\left(\theta|l\right) + \phi\sigma^2\mathbb{E}\left[\theta P\left(p|l\right)\right], \quad \text{(A111)}$$

while under integration it is

$$W_I = \log b + \log\frac{\Phi\mathbb{E}\theta}{1+\Phi\mathbb{E}\theta} + \mathbb{E}\log P\left(p\right) + \mathbb{E}\log\mathbb{E}\left(\theta|p\right) - \log\mathbb{E}\theta + \phi\sigma^2\mathbb{E}\left[\theta P\left(p\right)\right]. \quad \text{(A112)}$$

Let the distribution of population be

$$P_{1,L} = P_{2,R} = \frac{1+\tau}{4} \text{ and } P_{1,R} = P_{2,L} = \frac{1-\tau}{4} \quad \text{(A113)}$$

and information

$$\theta_{1,L} = \theta_{2,R} = \theta \text{ and } \theta_{1,R} = \theta_{2,L} = \theta\left(1-\zeta\right). \quad \text{(A114)}$$

Then, welfare under separation is

$$W_S = \log b + \log\frac{\Phi\theta\left[1-\left(1-\tau\right)\zeta/2\right]}{1+\Phi\left[1-\left(1-\tau\right)\zeta/2\right]} - \log 2 + \frac{1}{2}\left[\left(1+\tau\right)\log\left(1+\tau\right) + \left(1-\tau\right)\log\left(1-\tau\right)\right] + \frac{1-\tau}{2}\log\left(1-\zeta\right) - \log\left(1-\frac{1-\tau}{2}\zeta\right) + \frac{1}{4}\phi\sigma^2\theta\left[2\left(1+\tau^2\right) - \left(1-\tau\right)^2\zeta\right], \quad \text{(A115)}$$

while under integration it is

$$W_I = \log b + \log\frac{\Phi\theta\left[1-\left(1-\tau\right)\zeta/2\right]}{1+\Phi\left[1-\left(1-\tau\right)\zeta/2\right]} - \log 2 + \frac{1}{4}\phi\sigma^2\theta\left[2-\left(1-\tau\right)\zeta\right]. \quad \text{(A116)}$$

The welfare gain (or loss) from integration is

$$\Delta_W = \log\left(1-\frac{1-\tau}{2}\zeta\right) - \frac{1-\tau}{2}\log\left(1-\zeta\right) - \frac{1}{2}\left[\left(1+\tau\right)\log\left(1+\tau\right) + \left(1-\tau\right)\log\left(1-\tau\right)\right] - \frac{1}{4}\phi\sigma^2\theta\tau\left[2\tau + \left(1-\tau\right)\zeta\right], \quad \text{(A117)}$$

with limits

$$\lim_{\zeta\to 0}\Delta_W = -\frac{1}{2}\left[\left(1+\tau\right)\log\left(1+\tau\right) + \left(1-\tau\right)\log\left(1-\tau\right)\right] - \frac{1}{2}\phi\sigma^2\theta\tau^2 < 0 \text{ and } \lim_{\zeta\to 1}\Delta_W = \infty. \quad \text{(A118)}$$

The first derivative is

$$\frac{\partial\Delta_W}{\partial\zeta} = \frac{1}{2}\frac{\left(1-\tau^2\right)\zeta}{\left[2-\left(1-\tau\right)\zeta\right]\left(1-\zeta\right)} - \frac{1}{4}\phi\sigma^2\theta\tau\left(1-\tau\right), \quad \text{(A119)}$$

with limits

$$\lim_{\zeta\to 0}\frac{\partial\Delta_W}{\partial\zeta}=-\frac{1}{4}\phi\sigma^2\theta\tau\left(1-\tau\right)<0,\text{ and }\lim_{\zeta\to 1}\frac{\partial\Delta_W}{\partial\zeta}=\infty. \tag{A120}$$

The second derivative is

$$\frac{\partial^2\Delta_W}{\partial\zeta^2}=\frac{1}{2}\frac{\left(1-\tau^2\right)\left[2-\left(1-\tau\right)\zeta^2\right]}{\left[2-\left(1-\tau\right)\zeta\right]^2\left(1-\zeta\right)^2}>0. \tag{A121}$$

Thus there is a unique value $\bar{\zeta}\in(0,1)$ such that $\Delta_W\geq 0$ if and only if $\zeta\geq\bar{\zeta}$.

Comparative statics are $\partial\bar{\zeta}/\partial\sigma>0$ because

$$\frac{\partial\Delta_W}{\partial\sigma^2}=-\frac{1}{4}\phi\theta\tau\left[2\tau+\left(1-\tau\right)\zeta\right]<0, \tag{A122}$$

and $\partial\bar{\zeta}/\partial\tau>0$ because

$$\frac{\partial\Delta_W}{\partial\tau}=\frac{\zeta}{2-\left(1-\tau\right)\zeta}+\frac{1}{2}\log\left(1-\zeta\right)-\frac{1}{2}\left[\log\left(1+\tau\right)-\log\left(1-\tau\right)\right]$$
$$-\frac{1}{4}\phi\sigma^2\theta\left[4\tau+\left(1-2\tau\right)\zeta\right]<0. \tag{A123}$$

A.7. Proof of Proposition 5

If a voter i in region l has utility

$$u_t^i=\tilde{u}_t^i+\left(1-\xi\right)\log g_{l,t}+\frac{\xi}{L}\sum_{m=1}^{L}\log g_{m,t}, \tag{A124}$$

the expected ability of a local politician is

$$\mathbb{E}\hat{\eta}_l^D=\phi\sigma^2\left(1-\xi\frac{L-1}{L}\right)\theta_l \tag{A125}$$

and rent extraction under decentralization is

$$\rho_l^D=\left[1+\Phi\left(1-\xi\frac{L-1}{L}\right)\theta_l\right]^{-1}. \tag{A126}$$

The expected ability of a central politician is $\mathbb{E}\hat{\eta}^C=\phi\sigma^2\bar{\theta}$, so

$$\mathbb{E}\hat{\eta}^C-\frac{1}{L}\sum_{l=1}^{L}\mathbb{E}\hat{\eta}_l^D=\phi\sigma^2\bar{\theta}\xi\frac{L-1}{L}>0\text{ for all }\xi>0 \tag{A127}$$

with

$$\frac{\partial}{\partial\xi}\left(\mathbb{E}\hat{\eta}^C-\frac{1}{L}\sum_{l=1}^{L}\mathbb{E}\hat{\eta}_l^D\right)=\phi\sigma^2\bar{\theta}\frac{L-1}{L}>0. \tag{A128}$$

Rent extraction under centralization is $\rho^C = \left(1 + \Phi\bar{\theta}\right)^{-1}$, so

$$\frac{\partial}{\partial \xi}\left(\frac{1}{L}\sum_{l=1}^{L}\rho_l^D - \rho^C\right) = \Phi\xi\frac{L-1}{L^2}\sum_{l=1}^{L}\theta_l\left(\rho_l^D\right)^2 > 0, \tag{A129}$$

with

$$\lim_{\xi\to 0}\frac{1}{L}\sum_{l=1}^{L}\rho_l^D - \rho^C \geq 0. \tag{A130}$$

A.8. Proof of Corollary 2

Under centralization, the share of the spillover-inducing good in each region l is

$$\beta_g^C = \alpha_g \tag{A131}$$

with the welfare-maximizing uniformity constraint. Even without a uniformity constraint,

$$\beta_{g,l}^C = \alpha_g\left[\xi + (1-\xi)\frac{\theta_l}{\bar{\theta}}\right] \Rightarrow \frac{1}{L}\sum_{l=1}^{L}\beta_{g,l}^C = \alpha_g, \tag{A132}$$

so the allocation is socially optimal across goods although not across regions.

Under decentralization,

$$\beta_{g,l}^D = \frac{\left(1 - \frac{L-1}{L}\xi\right)\alpha_g}{1 - \frac{L-1}{L}\xi\alpha_g} < \alpha_g, \tag{A133}$$

such that

$$\frac{\partial\beta_{g,l}^D}{\partial\xi} = -\frac{\alpha_g\left(1-\alpha_g\right)}{\left(1 - \frac{L-1}{L}\xi\alpha_g\right)^2}\frac{L-1}{L} < 0. \tag{A134}$$

A.9. Proof of Proposition 6

Under decentralization, region l has welfare

$$\mathbb{E}u_l^D = \log b + \log\frac{\Phi\theta_l}{1+\Phi\theta_l} + \phi\sigma^2\theta_l. \tag{A135}$$

Under centralization,

$$\mathbb{E}u_l^C = \log b + \log\frac{\Phi\bar{\theta}}{1+\Phi\bar{\theta}} + \phi\sigma^2\bar{\theta} + \omega\left(\log\theta_l - \log\bar{\theta}\right). \tag{A136}$$

Thus region l prefers centralization if and only if

$$\log\left(1+\Phi\theta_l\right) - (1-\omega)\log\theta_l - \phi\sigma^2\theta_l \geq \log\left(1+\Phi\bar{\theta}\right) - (1-\omega)\log\bar{\theta} - \phi\sigma^2\bar{\theta}. \tag{A137}$$

The left-hand side is a function f_P with

$$\frac{\partial f_P}{\partial \theta_l} = \frac{\Phi}{1+\Phi\theta_l} - \frac{1-\omega}{\theta_l} - \phi\sigma^2 \tag{A138}$$

and

$$\frac{\partial^2 f_P}{\partial \theta_l^2} = \frac{1-\omega}{\theta_l^2} - \left(\frac{\Phi}{1+\Phi\theta_l}\right)^2. \tag{A139}$$

Therefore, it has a minimum at $\theta_l = \bar{\theta}$ if and only if

$$\sigma^2 < \frac{\Phi}{\phi\left(1+\Phi\bar{\theta}\right)^2} \text{ and } \omega = \frac{1}{1+\Phi\bar{\theta}} + \phi\sigma^2\bar{\theta} \equiv \tilde{\omega}, \tag{A140}$$

such that $\tilde{\omega} \in \left(\rho^C, 1-\left(1-\rho^C\right)^2\right)$ with

$$\frac{\partial \tilde{\omega}}{\partial \sigma^2} = \bar{\theta} > 0 \tag{A141}$$

and

$$\frac{\partial \tilde{\omega}}{\partial \bar{\theta}} = \phi\sigma^2 - \frac{\Phi}{\left(1+\Phi\bar{\theta}\right)^2} < 0. \tag{A142}$$

When $\omega = \tilde{\omega}$,

$$\frac{\partial f_P}{\partial \theta_l} = \frac{\Phi\phi\sigma^2}{\left(1+\Phi\theta_l\right)\theta_l}\left(\theta_l - \bar{\theta}\right)\left[\frac{1}{\phi\sigma^2\left(1+\Phi\bar{\theta}\right)} - \frac{1}{\Phi} - \theta_l\right], \tag{A143}$$

so the only other stationary point of f_P is a maximum. f_P is monotone increasing in $\theta_l \in \left(\bar{\theta}, 1\right)$ if

$$\sigma^2 \leq \frac{\Phi}{\phi\left(1+\Phi\right)\left(1+\Phi\bar{\theta}\right)} \equiv \bar{\sigma}^2. \tag{A144}$$

If (but not only if) this last condition holds, then every region with $\theta_l \neq \bar{\theta}$ strictly prefers centralization with discretionality $\tilde{\omega}$ to decentralization.