

# DISCUSSION PAPER SERIES

No. 9514

**CENTRALIZATION AND  
ACCOUNTABILITY: THEORY AND  
EVIDENCE FROM THE CLEAN AIR  
ACT**

Federico Boffa, Amedeo Piolatto and  
Giacomo AM Ponzetto

***INTERNATIONAL TRADE AND  
REGIONAL ECONOMICS and PUBLIC  
POLICY***



**Centre for Economic Policy Research**

**[www.cepr.org](http://www.cepr.org)**

Available online at:

**[www.cepr.org/pubs/dps/DP9514.asp](http://www.cepr.org/pubs/dps/DP9514.asp)**

# CENTRALIZATION AND ACCOUNTABILITY: THEORY AND EVIDENCE FROM THE CLEAN AIR ACT

Federico Boffa, Università di Macerata and IEB  
Amedeo Piolatto, IEB, Universitat de Barcelona  
Giacomo A.M. Ponzetto, CREI, Universitat Pompeu Fabra, Barcelona GSE and  
CEPR

Discussion Paper No. 9514  
June 2013  
Revised June 2014

Centre for Economic Policy Research  
77 Bastwick Street, London EC1V 3PZ, UK  
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820  
Email: [cepr@cepr.org](mailto:cepr@cepr.org), Website: [www.cepr.org](http://www.cepr.org)

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL TRADE AND REGIONAL ECONOMICS and PUBLIC POLICY**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Federico Boffa, Amedeo Piolatto and Giacomo A.M. Ponzetto

## ABSTRACT

### Centralization and Accountability: Theory and Evidence from the Clean Air Act\*

This paper studies fiscal federalism when regions differ in voters' ability to monitor public officials. We develop a model of political agency in which rent-seeking politicians provide public goods to win support from heterogeneously informed voters. In equilibrium, voter information increases government accountability but displays decreasing returns. Therefore, political centralization reduces aggregate rent extraction when voter information varies across regions. It increases welfare as long as the central government is required to provide public goods uniformly across regions. The need for uniformity implies an endogenous trade off between reducing rents through centralization and matching idiosyncratic preferences through decentralization. We find that a federal structure with overlapping levels of government can be optimal only if regional differences in accountability are sufficiently large. The model predicts that less informed regions should reap greater benefits when the central government sets a uniform policy. Consistent with our theory, we present empirical evidence that less informed states enjoyed faster declines in pollution after the 1970 Clean Air Act centralized environmental policy at the federal level.

JEL Classification: D72, D82, H41, H73, H77 and Q58

Keywords: air pollution, elections, environmental policy, government accountability, imperfect information, interregional heterogeneity and political centralization

Federico Boffa  
Università di Macerata  
Piaggia della Torre, 8  
62100 Macerata (MC)  
ITALY

Amedeo Piolatto  
Barcelona Institute of Economics  
(IEB)  
Universidad de Barcelona  
c/ Tinent Coronel Valenzuela, 1-11  
08034, Barcelona  
SPAIN

Email: federico.boffa@unimc.it

Email: piolatto@ub.edu

For further Discussion Papers by this author see:  
[www.cepr.org/pubs/new-dps/dplist.asp?authorid=177397](http://www.cepr.org/pubs/new-dps/dplist.asp?authorid=177397)

For further Discussion Papers by this author see:  
[www.cepr.org/pubs/new-dps/dplist.asp?authorid=177398](http://www.cepr.org/pubs/new-dps/dplist.asp?authorid=177398)

Giacomo A.M. Ponzetto  
CREI - Universitat Pompeu Fabra  
C/ Ramon Trias Fargas, 25-27  
08005 Barcelona  
SPAIN

Email: [gponzetto@crei.cat](mailto:gponzetto@crei.cat)

For further Discussion Papers by this author see:  
[www.cepr.org/pubs/new-dps/dplist.asp?authorid=171448](http://www.cepr.org/pubs/new-dps/dplist.asp?authorid=171448)

\*We are grateful for their helpful comments to Peter Backus, Paula Bustos, Vasco Carvalho, Edward Glaeser, Matthew Kahn, Gianmarco León, Ana Nuevo-Chiquero, Luigi Pascoli, Diego Puga, Thijs van Rens, Kurt Schmidheiny, Hannes Schwandt, Paolo Surico, Yanos Zylberberg, and seminar participants at CEMFI, CREI, Erasmus University Rotterdam, IMT Lucca, Toulouse School of Economics, UPF, and the Royal Economic Society 2013 Annual Conference. We are indebted to Mark Strazicich for sharing data with us. Piolatto acknowledges financial support from IVIE, the Spanish Ministry of Science and Innovation (grant ECO-2012-37131), and the Generalitat de Catalunya (grant 2009 SGR 102). 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, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075), the Barcelona GSE Research Network and the Generalitat de Catalunya (2009 SGR 1157).

Submitted 18 April 2013; revised 2 June 2014

# 1 Introduction

Two former governors of a U.S. state, Don Siegelman of Alabama and Rod Blagojevich of Illinois, are currently serving sentences in federal prison following conviction for corruption. Since the Second World War, eight more state governors as well as nine members of state executives have been similarly convicted of acts of official corruption and sentenced to jail. During the same period, no president of the United States nor any member of the federal cabinet has been charged with crimes investigated by the Department of Justice as part of the federal prosecution of public corruption.

History lends broader support to the view that corruption is more widespread in state and local governments than at the federal level. By introducing federal oversight of welfare spending, the New Deal eradicated the patronage and political manipulation that had hitherto characterized relief programs managed by the states and localities (Wallis 2000, 2006; Wallis, Fishback, and Kantor 2006). Thus, the most dramatic episode of centralization in U.S. history achieved a striking decrease in corruption.

International evidence highlights other instances of a positive impact of political centralization 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 modernization of European states. It proved a necessary step for the consolidation of state capacity, which was in turn a critical determinant of economic development (Dincecco 2009, 2011; Besley and Persson 2011; Gennaioli and Voth 2011; Dincecco and Katz 2013). In recent decades, Blanchard and Shleifer (2001) argue that China grew faster than Russia thanks to the greater strength of its central government vis à vis local politicians.<sup>1</sup>

In this paper, we present a theoretical model that explains how centralization can enhance government accountability. We build on the observation that voters in different regions are unequally capable of holding politicians accountable. Three out of ten state governors convicted of corruption were from Illinois. More systematically, Glaeser and Saks (2006) document that the rate of official corruption from 1976 to 2002 was five times lower in Oregon than in Alaska, Mississippi, or Louisiana. When such disparate regions are united in a single national polity, we find that the federal executive is held accountable mainly by the most capable voters. As a consequence, his incentives and performance are better than those of the average local politician.

---

<sup>1</sup>Contemporary cross-country studies of decentralization and perceived corruption have yielded conflicting results (Treisman 2007; Fan, Lin, and Treisman 2009).

In our model, self-interest politicians can extract wasteful private rents instead of providing public goods. Such rent-seeking behavior is constrained by electoral discipline. Career concerns induce the incumbent to provide public goods so that informed citizens derive a positive inference of his ability and vote for his re-election. In equilibrium, politicians extract lower rents if voters are better informed.

In a dynamic framework we prove that the link between voter information and political accountability is subject to decreasing returns. The expectation of closer scrutiny in the future reduces the appeal of re-election and therefore the effectiveness of tighter current monitoring. Decreasing returns to voter information imply that national elections provide much better incentives than local elections in the least informed regions, and not much worse than in the most informed ones. Centralization thus increases overall efficiency by reducing aggregate political rents.

Our finding of accountability gains from centralization is novel in the literature on the political economy of fiscal federalism. Prior work has mostly emphasized the advantages of decentralization (Lockwood 2006). In particular, Besley and Case (1995) and Besley and Smart (2007) have studied how yardstick competition across local jurisdictions can make decentralized government more accountable.<sup>2</sup>

Our model also accounts for the regional distribution of the efficiency gains from centralization. When the central government provides public goods uniformly throughout the union, lower-information regions enjoy a transfer of accountability from their more informed partners. We prove that a region's welfare gains from centralization are then strictly decreasing in its residents' information.

We present empirical evidence supporting this theoretical prediction in the context of U.S. environmental policy. The Clean Air Act of 1970 transferred responsibility for pollution regulation from the state and local governments to the federal Environmental Protection Agency. A difference-in-differences analysis establishes that pollutant emissions began to decline, relative to pre-existing trends, considerably faster in states with lower newspaper circulation after national air-quality standards were introduced.

The predictions of our model are antithetical if the central government can provide public goods dishomogeneously across regions. Then higher-information regions enjoy a transfer of power from their less informed peers. The central government targets discretionary spending to the informed voters who monitor it most closely. This theoretical prediction is borne out by empirical evidence that discretionary New Deal funds were disproportionately allocated

---

<sup>2</sup>Conversely, the central government could be less susceptible to capture by special interest groups. However, formal analysis of this possibility has reached ambiguous conclusions (Bardhan and Mookherjee 2000, 2006a,b).

to more informed counties within each state (Strömberg 2004).

Our theory highlights the importance of striking a balance between uniform and discretionary public goods provision at the central level. Without any uniformity, centralization would be welfare reducing despite the associated reduction in political rents because it entails regressive redistribution. With sufficient uniformity, instead, centralization increases aggregate welfare. Given the right balance between uniform and discretionary items, we find that centralization can even be a Pareto improvement over decentralization.

The distributive tensions that underpin Pareto-improving centralization chime with the pattern of interregional political discourse in the United States, and even more so in the European Union. Better informed states in the North must be transferring accountability to the South through the provision of some uniform public goods. Thus, it is natural they should complain of a drain on their resources and a pollution of their own institutional quality. At the same time, less informed states in the South must be losing power through the provision of some discretionary public goods. Hence their expected complaint that the North plays a privileged role in setting union-wide policy priorities and imposing them on southern states.

We extend our framework to consider region-specific preferences for different public goods. Our model then entails a trade off between higher preference-matching under decentralization and higher efficiency under centralization. A similar trade off has long been a focus of the economic analysis of fiscal federalism (Oates 1972, 1999). However, the classic theory posits as exogenous technological forces both economies of scale from centralization and a uniformity requirement for centralized public-good provision. Instead, our model of political agency microfounds both forces as endogenous outcomes. Centralization raises efficiency by reducing rent extraction, but it needs to be subject to a uniformity constraint or else it would reduce welfare by inducing regressive redistribution.

Our model entails economies of scope in government activity. When a politician is tasked with providing a broader set of public goods and assigned a proportionately larger budget, he dissipates a lower fraction of it as rents. Hence, we find that the division of powers between multiple levels of government is liable to increase rent extraction. This result is consistent with cross-country evidence that reported bribery increases with the number of administrative tiers (Fan, Lin, and Treisman 2009).

As a consequence of economies of scope we show that a true federal structure, with both a central and a regional level of government, can be welfare maximizing only if differences in voter information across regions are sufficiently large. Hence, our model suggests that it is far from a coincidence that government accountability should vary so widely across the United States or the European Union. On the contrary, such heterogeneity may be the very

reason why a federal structure with overlapping layers of government is desirable.

We find that the least informed regions are those that benefit the most from dividing powers between a central federal government and decentralized local ones. The best informed regions would prefer instead complete centralization, or else complete decentralization. The predicted pattern is consistent with the observation that southern U.S. states, which have lower voter information and government accountability than the average, also tend to be the most vocal in their support for federalism and opposition to further expansion of the powers of the federal government.

In a final extension, we introduce spillovers in public goods across regions. We find that externalities raise the benefits of centralization through three forces. First, they induce the selection of better politicians. Unlike voters in local elections, the national electorate internalizes all the benefits of politicians' competence. Therefore, it is keener on screening politicians for their skills. Second, this very mechanism induces a further improvement in politicians' incentives due to career concerns. Third, the central politician allocates efficiently the government budget, net of rent extraction, while local politicians underprovide public goods that generate spillovers.

The remainder of the paper is organized as follows. Section 2 sets up our model of political career concerns. Section 3 solves for its equilibrium and establishes positive but decreasing returns to voter information. Section 4 applies the model to multiple regions and compares rent extraction and government productivity under centralization and decentralization. Section 5 characterizes the welfare ranking of centralization and decentralization, both from the aggregate perspective and for each region separately. In Section 6 we introduce heterogeneous preferences and analyze the endogenous trade off between efficiency and preference-matching. Section 7 studies a federal structure that divides powers between a central government and decentralized local governments. Section 8 considers interregional externalities. Section 9 presents our novel empirical evidence on U.S. environmental policy. Section 10 concludes.

## 2 Political Agency and Public-Good Provision

The economy is populated by 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

$$u_t^i = \tilde{u}_t^i + \log g_t, \tag{1}$$



where  $g_t$  is public-good provision, while  $\tilde{u}_t^i$  is utility from private consumption. We treat  $\tilde{u}_t^i$  as an exogenous mean-zero shock and focus exclusively on public goods.<sup>3</sup>

Public goods are produced by the government with technology

$$g_t = e^{\eta_t} x_t. \quad (2)$$

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

Productivity  $\eta_t$  represents the stochastic competence of the incumbent politician in providing public goods. It follows a first-order moving average process

$$\eta_t = \varepsilon_t + \varepsilon_{t-1}. \quad (3)$$

The shocks  $\varepsilon_t$  are independent and identically distributed over time, and across politicians. They have support  $[\check{\varepsilon}, \hat{\varepsilon}]$ , mean zero and variance  $\sigma^2$ . This standard specification of the stochastic process admits an intuitive interpretation in terms of infinitely lived political dynasties or parties, composed of overlapping generations with independent competence. In each period, the government consists of older leaders approaching retirement, whose competence is  $\varepsilon_{t-1}$ , and rising young politicians with competence  $\varepsilon_t$ , who will take over the party leadership in the following period. Thus, the aggregate productivity of the ruling party is  $\eta_t$ .

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 incumbent extracts a rent

$$r_t = b - x_t. \quad (4)$$

The most immediate interpretation of rent extraction is pecuniary. The incumbent allocates a given government budget  $b$ , invariant over time and subject to a balanced-budget constraint in each period. He spends an amount  $x_t$  on public-good provision, while he devotes the remainder  $r_t$  to socially unproductive ends ranging from party finance to outright embezzlement. While this reading is perhaps the most intuitive, the model can identically represent slacking instead of stealing. The incumbent enjoys an invariant exogenous reward  $b$  from holding office, including compensation and perks as well as the “ego rent” of being in power.

---

<sup>3</sup>In the appendix we solve a more general version of the model with several distinct public goods. Results are essentially unchanged.

However, he incurs a cost  $x_t$  from exerting effort to provide public goods. Rent extraction  $r_t$  then captures politicians' failure to work diligently in their constituents' interest.

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 commitment, so the election is not based on campaign promises, but rather on retrospective evaluation of the incumbent's track record. Current rent-extraction is disciplined by career concerns. The incumbent invests in providing public goods because higher public-good provision raises voters' inference of the incumbent's ability  $\varepsilon_t$  and thereby increases his chances of re-election (Holmström [1982] 1999; Persson and Tabellini 2000; Alesina and Tabellini 2008).

In the standard model of political career concerns, all voters have the same information concerning policy outcomes, while none observes the politician's choices underpinning them. We assume that some voters fail to observe, or to understand, policy outcomes as well. Thus, political incentives are shaped by the inferences of voters who reach the election with heterogeneous information (Besley and Burgess 2002; Strömberg 2004; Glaeser, Ponzetto, and Shapiro 2005; Ponzetto 2011; Glaeser and Ponzetto 2014; Ponzetto and Troiano 2014).

The timeline within each period  $t$  is the following.

1. The incumbent politician's past competence shock  $\varepsilon_{t-1}$  becomes common knowledge.
2. The incumbent chooses investment  $x_t$ , and residually rent  $r_t$ , without knowing the realization of his period- $t$  competence shock  $\varepsilon_t$ .
3.  $\varepsilon_t$  is realized and the provision of public goods  $g_t$  is determined. Each voter  $i$  observes  $g_t$  with probability  $\theta_i$ ; with probability  $1 - \theta_i$  he remains completely uninformed. The arrival of information is independent across voters. No voter has any direct observation of  $\varepsilon_t$ ,  $r_t$ , or  $x_t$ .
4. An election is held, pitting the incumbent against a single challenger, randomly drawn from the same pool of potential office-holders.

The incumbent's and the challenger's competence shocks are known to be independent draws from a common distribution. Moreover, voters have rational expectations that any politician in every period will choose the same allocation  $\bar{x}$ , because the environment is stationary and performance is separable in effort and ability.

Since no information about the challenger is available, all voters have identical rational expectations that public-good provision if he wins the election is going to equal

$$\mathbb{E} \log g_{t+1}^C = \log \bar{x} + \mathbb{E} \eta_{t+1}^C = \log \bar{x}. \quad (5)$$

Uninformed voters also have no way of assessing the incumbent's skill innovation  $\varepsilon_t$ , and thus his future ability  $\eta_{t+1}$ . Hence they expect no difference in public-good provision whether the incumbent is reelected or the challenger defeats him:

$$\Delta_0 = \mathbb{E} \log g_{t+1}^I - \mathbb{E} \log g_{t+1}^C = 0. \quad (6)$$

As a consequence, their voting behavior is not affected by the actual extent of public-good provision. This could be because uninformed voters are simply unaware that public goods affect their utility. This form of ignorance may be particularly natural for public goods that generate long-run benefits, such as investments in education, research, or pollution abatement. Citizens could also assess correctly the benefits of public goods, but fail to recognize the extent to which public-good provision depends on the actions of the incumbent politician and on his competence (Strömberg 2004).

Informed voters, instead, can use their knowledge of public-good provision  $g_t$  to infer the incumbent's competence  $\eta_t$  and thus its long-lasting component  $\varepsilon_t$ :

$$\mathbb{E} (\eta_{t+1}^I | g_t) = \mathbb{E} (\varepsilon_t | g_t) = \log g_t - \log \bar{x} - \varepsilon_{t-1}. \quad (7)$$

In a rational-expectation equilibrium their inference turns out to be perfect ( $x_t = \bar{x}$ ), accurately revealing  $\varepsilon_t$ . Based on this inference, informed voters expect a difference in the utility provided by a reelected incumbent relative to a victorious challenger equal to

$$\Delta_1(g_t) = \mathbb{E} (\log g_{t+1}^I | g_t) - \mathbb{E} \log g_{t+1}^C = \mathbb{E} (\eta_{t+1}^I | g_t) - \mathbb{E} \eta_{t+1}^C = \log g_t - \log \bar{x} - \varepsilon_{t-1}. \quad (8)$$

The electorate consists of a continuum of atomistic voters, which can be partitioned into  $J$  groups. Group  $j$  comprises a fraction  $\lambda_j$  of voters, who have identical probability  $\theta_j$  of information acquisition. The voters' average information is denoted by  $\bar{\theta} = \sum_{j=1}^J \lambda_j \theta_j$ .

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 in the following period. These preferences are summarized by the expected difference in utility from public goods  $\Delta_i(g_t) \in \{\Delta_0, \Delta_1(g_t)\}$ , depending on voter  $i$ 's information about  $g_t$ . Second, voters have preferences for candidates' characteristics other than their competence at providing public goods: e.g., oratorical skill, personal likability, or party ideology. These preferences can be decomposed into an aggregate shock  $\Psi_t$  and an independent idiosyncratic shock  $\psi_t^i$  which is i.i.d. across voters. Then voter  $i$  votes for the incumbent if and only if  $\Delta_i(g_t) \geq \Psi_t + \psi_t^i$ .

The common shock to the incumbent’s popularity accounts for the aggregate uncertainty in the electoral outcome. The idiosyncratic shock accounts for the imperfect predictability of each agent’s voting decision. Both shocks are symmetric around zero, so idiosyncratic preferences do not lead voters to favor systematically either incumbents or challengers. Moreover, neither the outcome of the election nor any voter’s ballot is perfectly predictable on the basis of observed public-good provision  $g_t$  alone.<sup>4</sup> Finally, we assume that both  $\Psi_t$  and  $\psi_t^i$  are uniformly distributed.

The uniform density  $\phi$  of the aggregate shock  $\Psi_t$  measures voters’ keenness on politicians’ competence, relative to their other characteristics that do not affect public-good provision. When  $\phi$  is high, relatively small differences in perceived competence translate into large swings in electoral support. When  $\phi$  is low, relatively large differences in perceived competence have a minor impact on electoral outcomes, which are instead mostly determined by random tastes unrelated to public-good provision.

### 3 Government Accountability from Voter Information

The incumbent’s probability of re-election as a function of his policy choice  $x_t$  equals

$$\pi(x_t) = \frac{1}{2} + \phi\bar{\theta}(\log x_t - \log \bar{x}). \quad (9)$$

This function summarizes the politician’s career concerns. The incumbent is exposed to the uncertain realization of his own competence shock  $\varepsilon_t$  at the time of choosing the amount of investment in public goods  $x_t$ . By moderating rent extraction, he can raise informed voters’ inference of his ability regardless of its true realization. This increases raises the incumbent’s chances of being re-elected and therefore continue to extract rents in future periods. Wider voter awareness of policy outcomes makes the probability of re-election more sensitive to investment in public goods. Thus, better voter information translates directly into sharper career concerns for incumbent politicians.

Let  $R$  denote the value of re-election, which coincides with the endogenous present value of expected future rents conditional on remaining in office. Then the trade-off between current and future rent extraction leads the incumbent to extract rents equal to

$$r(R) = \arg \max_{r_t} \{r_t + R\pi(b - r_t)\} = b - \phi\bar{\theta}R. \quad (10)$$

By equation (9), the equilibrium probability that the incumbent wins is  $\pi = 1/2$  in

---

<sup>4</sup>Formally, the support of the voters’ preference shocks  $\Psi_t$  and  $\psi_t^i$  is sufficiently wide, while the support of the politicians’ competence shock  $\varepsilon_t$  is sufficiently narrow.

every election when voters have rational expectations ( $\bar{x} = x_t$ ). The politician does not have any private information at the time of the policy choice. This rules out opportunities for signaling. Voters with rational expectations cannot be fooled, and their preferences ( $\Psi_t$  and  $\psi_t^i$ ) do not display a systematic bias against incumbency nor in its favor. As a consequence, a politician who rationally expects to extract a rent  $r$  whenever he is in office estimates a net present value of re-election equal to

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

because he also rationally expects equal chances of winning or losing each subsequent election.

Let  $\rho \equiv r/b \in [0, 1]$  denote the fraction of the budget allocated to rents. Then we obtain the following characterization of the stationary rational-expectation equilibrium of our model of political career concerns.

**Proposition 1** *In equilibrium, ruling politicians extract rents*

$$\rho = \left(1 + \frac{2\delta}{2-\delta} \phi \bar{\theta}\right)^{-1}.$$

*Rent extraction is a decreasing and convex function of voter information ( $\partial\rho/\partial\bar{\theta} < 0$  and  $\partial^2\rho/\partial\bar{\theta}^2 > 0$ ).*

The proposition describes the effect of voter information on government accountability. Intuitively, better information allows voters to monitor politicians more tightly. Voters can reward the provision of public goods only if they accurately perceive and understand it. As a consequence, voter information has a beneficial impact both on the moral-hazard problem of politician incentives and on the adverse-selection problem of politician selection.

As voter awareness of public-good provision improves, career concerns sharpen politician's incentives and induce lower rent extraction ( $\partial\rho/\partial\bar{\theta} < 0$ ). This result accords with the empirical finding that government performance improves with media scrutiny (Besley and Burgess 2002; Adserà, Boix, and Payne 2003; Ferraz and Finan 2008; Snyder and Strömberg 2010; Ponzetto 2011).<sup>5</sup>

---

<sup>5</sup>The link between greater information and better governance does not suffice to create incentives for agents to acquire political information, due to the paradox of the rational voter. Since each voter has a negligible (in the model, precisely nil) chance of determining the outcome of the election, he also has vanishing incentives to improve his own monitoring ability. Thus, the decision to acquire information  $\bar{\theta}$  is not endogenous to the election game, but derives from exogenous characteristics of the electorate. In particular, Putnam (1993) argues that newspaper readership reflects an individual's civic involvement and social capital.

Furthermore, Proposition 1 establishes that the beneficial impact of voter information on rent extraction has decreasing returns ( $\partial^2 \rho / \partial \bar{\theta}^2 > 0$ ). This key result follows from the dynamic nature of the politician's problem. The immediate impact of voter information on rent extraction is linear, as shown in equation (10). For a given value of re-election  $R$ , more informed voters induce one-to-one more productive investment and lower political rents. This would be the entire effect of a purely transitory one-period increase in voter information.

However, in a dynamic setting, a permanent increase in voter information has an additional indirect effect. The politician understands he will be more closely monitored in the future if re-elected. Therefore, the expected future rents from holding office decrease. Their decline reduces the incentives to refrain from immediately extracting rents. Thus, the direct effect of improved monitoring is mitigated. The higher the level of voters' information  $\bar{\theta}$ , the more sensitive current rent-extraction to the expectation of future rent. The decrease in rent extraction ensuing from a marginal improvement in monitoring becomes smaller the more voters are informed to begin with.

The economic intuition for decreasing returns to monitoring can be grasped most immediately by comparing the extreme cases of government accountability. If no workers are aware of public-good provision, career concerns are absent and rent extraction is unchecked ( $\bar{\theta} = 0 \Rightarrow \rho = 1$ ). Introducing a modicum of monitoring induces a forceful reaction by politicians who are afraid of losing very large rents. At the opposite extreme, if all workers perfectly observe public-good provision, career concerns are at their strongest but rent-extraction cannot be reduced to zero ( $\bar{\theta} = 1 \Rightarrow \rho > 1$ ). The incumbent politician always extract some rent because it is only the appeal of future rent extraction that induces him to make productive investments in the first place. Marginally worsening perfect monitoring has but a small effect on politicians. Their career incentives are dampened by the very fact that an active electorate keeps them constantly on their toes and makes incumbency of modest value.

The description of the equilibrium is completed by the characterization of the equilibrium competence of ruling politicians.

**Corollary 1** *In equilibrium, the expected competence of ruling politicians is*

$$\mathbb{E}\hat{\eta}_t = \phi \bar{\theta} \sigma^2.$$

*An increase in voter information  $\bar{\theta}$  raises the competence of ruling politicians  $\hat{\eta}_t$  in the sense of first-order stochastic dominance.*

Elections provide not only politicians' incentives but also a screening mechanism. The

expected competence of a ruling politician is above average ( $\mathbb{E}\hat{\eta}_t > 0$ ) for all  $\bar{\theta} > 0$  because voters tend to re-elect politicians who delivered an unexpectedly high level of public goods and conversely to replace those who underperformed. As voter knowledge improves, the mapping between public-good provision and the incumbent's chances of re-election becomes tighter. Thus, the effectiveness of the electoral selection mechanism rises with the share of informed voters who cast their ballots based on inference of the incumbent's skill.

Better incentives and better screening naturally combine to increase social welfare. In equilibrium, welfare equals

$$\mathbb{E}u_t = \log b + \log(1 - \rho) + \mathbb{E}\hat{\eta}_p. \quad (12)$$

It is an increasing and concave function of voter information ( $\partial\mathbb{E}u_t/\partial\bar{\theta} > 0$  and  $\partial^2\mathbb{E}u_t/\partial\bar{\theta}^2 < 0$ ), mirroring the decreasing returns to monitoring depicted by Proposition 1.

Other determinants of government quality and social welfare are straightforward. Rent extraction declines when politicians are more patient, for then they are more willing to sacrifice current benefits for a higher probability of remaining in office in the future ( $\partial\rho/\partial\delta < 0$ ). Average competence increases with the variance of the population distribution of ability, which measures the gains available from screening ( $\partial\mathbb{E}\hat{\eta}_t/\partial\sigma^2 > 0$ ). Finally, both incentives and selection improve when voters are keener on competence and less likely to be swayed by unrelated determinants of candidate popularity ( $\partial\rho/\partial\phi < 0$  and  $\partial\mathbb{E}\hat{\eta}_t/\partial\phi > 0$ ).<sup>6</sup>

## 4 Efficiency Benefits of Centralization

The main focus of our analysis is on the difference between centralized and decentralized government control of public-good provision. Thus, we consider an economy divided into  $L$  regions. With decentralized government, in each region  $l$  a local politician with ability  $\eta_{l,t}^D$  independently invests in the provision of public goods  $x_{l,t}^D$  and extracts rent

$$r_{l,t}^D = b - x_{l,t}^D. \quad (13)$$

---

<sup>6</sup>The findings of Proposition 1, however, do not hinge on the assumption that voters have preferences for candidate characteristics other than competence at providing public goods. They would not be materially affected if we assumed instead that the representative voter cares exclusively about politicians' competence and is informed of public-good provision  $g_t$  with probability  $\bar{\theta}$ . Rent extraction is exactly unchanged as long as  $\varepsilon_t$  has a distribution that is symmetric around the origin and has density  $\phi$  at zero. Indeed, if the distribution is uniform equation (9) still describes exactly the probability of re-election. The endogenous distribution of competence  $\hat{\eta}_t$  would be quantitatively but not qualitatively different.

Following proposition 1, we measure rent extraction in region  $l$  under decentralization by  $\rho_l^D = r_l^D/b$ . This can be immediately interpreted as the fraction of the regional budget  $b$  that the politician misallocates. Identically, as discussed above, it could measure the extent to which the local politician enjoys the rewards of office  $b$  without exerting effort  $x_{l,t}^D$ .

Centralization means that a single central politician with ability  $\eta_t^C$  chooses investment in public goods  $x_{l,t}^C$  for all regions  $l$ . and extracts rents

$$r_t^C = bL - \sum_{l=1}^L x_{l,t}^C. \quad (14)$$

If we interpret rent extraction simply as embezzlement, the size of the central politician's rent follows immediately. Under centralization, he controls an aggregate budget ( $bL$ ) equal to the sum of the budgets controlled by local politicians under decentralization. With the alternative interpretation of rent extraction as slacking, this expression corresponds to the additional assumption that the perks and ego rents of office are similarly additive. In either case, rent extraction under centralization can be measured by  $\rho^C = r^C/(bL)$ , which has the same normalization as  $\rho_l^D$ .

Centralization may also require the central government to provide public goods uniformly across regions ( $x_{l,t}^C = x_t^C$  for all  $l$ , implying  $g_{l,t}^C = g_t^C$  given the common productivity  $\eta_t^C$ ). The literature has typically assumed such a uniformity constraint, which provides a simple rationale for decentralization given heterogeneous preferences across regions (Oates 1972; Alesina and Spolaore 1997; Alesina, Angeloni, and Etro 2005). Nonetheless, imposing a uniformity constraint on centralized public-good provision is not necessarily realistic in all settings. Discretionary federal spending is not required to be allocated uniformly across states, almost by definition of discretionality. Lockwood (2002) and Besley and Coate (2003) have modelled fiscal federalism under the alternative hypothesis that the central government can arbitrarily vary the provision of public goods across regions. We consider below the different distributional implications of centralized public-good provision with and without a uniformity constraint.

To begin with, however, we can establish that centralization improves overall political accountability regardless of the presence or absence of a uniformity constraint. Suppose that voters in region  $l$  have information  $\theta_l$  (on average). Then the following result obtains.

**Proposition 2** *If information is heterogeneous across regions, aggregate rent extraction is lower under centralization than decentralization ( $\rho^C < \sum_{l=1}^L \rho_l^D/L$  if  $\theta_l \neq \theta_m$  for some  $l \neq m$ ).*

If voters are heterogeneously informed ( $\theta_l \neq \theta_m$  for  $l \neq m$ ), and thus politicians are



heterogeneously accountable, centralization has beneficial aggregate effects on accountability. By joining heterogeneous regions into a single polity, centralization leads to an overall level of political information equal to the average  $\bar{\theta}$  of information across regions. Decreasing returns to monitoring then imply a decline in aggregate rents. The central government extract slightly higher rents than the local governments in regions with above-average information. Yet it extracts much lower rents than local politicians in regions with below-average information. The aggregate effect of centralization is thus an unambiguous decrease in rent extraction, consistent with historical evidence from the New Deal (Wallis 2000, 2006; Wallis, Fishback, and Kantor 2006), from precolonial Africa (Gennaioli and Rainer 2007a,b), and from the rise of modern nation-states in Europe (Dincecco 2011; Gennaioli and Voith 2011; Dincecco and Katz 2012).

Public-good provision and social welfare depend on both rent extraction and politicians' skill. Concerning the latter, Corollary 1 immediately implies an invariance result.

**Corollary 2** *The expected competence of ruling politicians is identical on average under centralization and decentralization ( $\mathbb{E}\hat{\eta}_t^C = \sum_{l=1}^L \mathbb{E}\hat{\eta}_{l,t}^D/L = \phi\bar{\theta}\sigma^2$ ).*

Through Proposition 2 and Corollary 2, our model establishes a novel political-economy mechanism yielding efficiency gains from centralization. This new driving force is additional to, and independent of, the classic technological rationales for centralization or decentralization. Accordingly, Proposition A1 in the appendix proves that if information is homogeneous across regions ( $\theta_l = \theta$  for all  $l$ ) our framework replicates exactly Oates's (1972) Decentralization Theorem.

In his classic analysis, centralization is useful to internalize cross-regional externalities. Conversely, decentralization is beneficial to avoid the cost of policy uniformity when regions have heterogeneous preferences. Thus, their welfare ranking follows a three-fold taxonomy. First, centralization and decentralization yield identical outcomes if preferences are homogeneous and there are no externalities, as we have formally assumed so far. Second, centralization yields higher welfare than decentralization if preferences are homogeneous and there are externalities because public goods provided to one region also affect welfare in others. Third, decentralization yields higher welfare than centralization if there are no externalities, centralized public-good provision is subject to a uniformity constraint, but preferences are heterogeneous because the government provides distinct public goods and different regions have different ideal baskets of these goods.

Despite imperfect political agency and ineliminable distortions due to political rent extraction, our model reproduces the same taxonomy if and only if voters in all regions are

identically informed.<sup>7</sup> With heterogeneous information, instead, Proposition 2 establishes a new force that makes centralization efficiency-enhancing.

## 5 Distributional and Welfare Consequences

Whether the increase in efficiency resulting from centralization translates into an increase in welfare depends on the distribution of the benefits from reduced rent extraction, which are generically uneven across regions. We now turn to these distributive consequences of centralization. In this analysis, the distinction between uniform and discretionary public goods provision by the central government plays a prominent role.

We begin by characterizing the welfare consequences of centralization under a uniformity constraint.

**Proposition 3** *Suppose that public-good provision under centralization is subject to a uniformity constraint ( $x_{l,t}^C = x_t^C$  for all  $l$ ). Then centralization increases aggregate social welfare ( $\sum_l \mathbb{E}u_l^C \geq \sum_l \mathbb{E}u_l^D$ ). It yields a greater increase in expected public-good provision and welfare for less informed regions (if  $\theta_l < \theta_m$  then  $\mathbb{E}(g^C - g_l^D) > \mathbb{E}(g^C - g_m^D)$  and  $\mathbb{E}(u_l^C - u_l^D) > \mathbb{E}(u_m^C - u_m^D)$ ).*

When a uniformity constraint is imposed upon centralized public-good provision, centralization yields higher aggregate welfare than decentralization for two reasons. First, as established in Proposition 2, rent extraction falls and therefore more resources are made available for public-good provision. Second, Proposition 3 highlights that centralization subject to a uniformity constraint determines progressive redistribution. The increase in productive public spending caused by reduced rent extraction is directed to the regions who need it most.

Progressivity results from an effective transfer of accountability from the regions with higher to those with lower voter information. The more informed regions are better at incentivizing and selecting local politicians. Conversely, the less informed regions are plagued

---

<sup>7</sup>In this case imperfect agency causes the same distortions under centralization or decentralization. Constituency size affects political agency through two opposing forces (Seabright 1996; Persson and Tabellini 2000). Centralization reduces the probability that voters in any one region are pivotal in the election. Hence a central politician is less responsive to each voter's preferences than a local politician is to those of his fewer constituents. Conversely, centralization increases the scale of political rent. When the politician allocates the larger central budget instead of a smaller regional budget, re-election is more valuable. A greater value of re-election sharpens the incentives for the central politician to perform well. In our framework these forces are perfectly balanced. Centralization expands the budget by a factor  $L$ , while reducing the electoral clout of each region by a factor  $1/L$ . The politician's incentives are thus invariant with respect to the scale of his constituency. Rent extraction is proportional to the government budget, as established in Proposition 1.

with rent-extracting and incompetent local governments. Centralization enables them to out-source their governance to a central government that is held accountable by better-informed voters in other regions. As a consequence, regions with below-average information benefit from centralization because the central government has much better incentives than their decentralized local governments. On the other hand, better informed regions suffer from a dilution of their government accountability under centralization.

The egalitarian inter-regional allocation mandated by the uniformity constraint directs the efficiency gains from centralization towards the least informed and worst governed regions. These are precisely those whose marginal utility from public goods is highest. The better informed and better run regions stand to suffer a loss instead. But their marginal cost of lower public-good provision is moderate because they are already well supplied under decentralization. As a result, utilitarian social welfare increases as inequality falls at the same time as aggregate efficiency rises.

The distinctive empirical prediction of Proposition 3 is that empowering the central government to set a uniform nation-wide policy should benefit each region in inverse proportion to its citizens' information. In Section 9, we provide empirical evidence of the predicted pattern for the case of environmental policy in the United States. Additional support for our theoretical prediction is also provided by suggestive evidence from Europe.

The European Union encompasses large disparities in the quality of government across regions and member states (Charron, Dijkstra, and Lapuente 2013). Consistent with our model, Fredriksson and Gaston (2000) conclude that 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.”

Italy provides a striking example of large regional disparities in information and accountability within a single country (Putnam 1993; Del Monte and Papagni 2001, 2007; Golden and Picci 2005). Durante, Labartino, and Perotti's (2011) empirical findings on decentralization in the Italian public university system correspond perfectly to our theory. A 1998 reform transferred responsibility for faculty hiring from the national level to individual universities. As a result, the quality of academic recruitment fell in provinces with lower newspaper readership. Those with higher readership experienced no decline, but no more than a marginal improvement, implying an aggregate efficiency loss from decentralization.

The distributional and welfare consequences of centralization are inverted in the absence of a uniformity constraint.

**Proposition 4** *Suppose that public-good provision under centralization is not subject to a*

*uniformity constraint. Then the central government allocates public goods across regions in proportion to their residents' information ( $x_{l,t}^C/x_{m,t}^C = g_{l,t}^C/g_{m,t}^C = \theta_l/\theta_m$ ). Centralization reduces aggregate social welfare ( $\sum_l \mathbb{E}u_l^C \leq \sum_l \mathbb{E}u_l^D$ ). It yields a greater decline in welfare for regions whose residents are less informed (if  $\theta_l < \theta_m$  then  $\mathbb{E}(u_l^C - u_l^D) < \mathbb{E}(u_m^C - u_m^D)$ ) if the variance of politicians' ability is sufficiently low ( $\sigma^2 \leq \hat{\sigma}^2$ ).*

In the absence of a uniformity constraint centralization yields lower aggregate welfare than decentralization. The positive welfare effect of rent reduction is more than offset by the detrimental effect stemming from regressive redistribution.

Regressivity results from a transfer of power from the less to the more informed regions. Discretionary policies target government spending towards more politically influential regions. In our model, political influence stems from information because more knowledgeable voters provide more of the politicians' incentives. As a consequence, the central government provides public goods disproportionately to better informed regions, unless it is constrained by a uniformity requirement. This pattern is reflected in Strömberg's (2004) evidence on the regional allocation of discretionary government spending during the New Deal. State governors directed more public funds to counties with a greater share of radio listeners, and thus with better informed voters.

Proposition 4 establishes that the geographic misallocation of expenditure by a central government without a uniformity constraint is more detrimental than higher rent extraction by decentralized local governments. On average, uninformed voters in worse-off regions suffer more when the central government channels their taxes towards public spending in better informed regions than they do when they are defrauded by rent-extracting local politicians.

Net expenditure on public-good provision in region  $l$  is  $(1 - \rho_l^D)$  under decentralization, but  $(1 - \rho^C) \theta_l/\bar{\theta}$  under centralization. Due to decreasing returns to monitoring, the former is a concave function of voter information. The latter is instead linear in it, because a region's political influence over the central government is directly proportional to residents' political awareness. Equilibrium rent extraction from Proposition 1 implies that

$$(1 - \rho^C) \frac{\theta_l}{\bar{\theta}} - (1 - \rho_l^D) = \frac{2\delta}{2 - \delta} \phi \theta_l (\rho^C - \rho_l^D) < 0 \Leftrightarrow \rho^C < \rho_l^D \Leftrightarrow \theta_l < \bar{\theta}. \quad (15)$$

Therefore, centralization increases productive government expenditure in regions with above-average voter information, and decreases it in those with below average voter information.

However, centralization without a uniformity constraint causes productivity to move in the opposite direction as inputs. It reduces the resources directed at providing public goods in less informed regions. It increases, by a larger aggregate amount, investment in public goods for more informed regions. At the same time, the central politician has average expected

competence. He is less productive than local politicians in the high-information regions but more productive than those in the low-information ones.

Hence, the comparison of welfare consequences for any two given regions depends on the distribution of politicians' ability. If its variance is sufficiently low ( $\sigma^2 \leq \hat{\sigma}^2$ ), changes in the budget allocation dominate changes in productivity for every pair of regions. If instead politicians have highly heterogeneous skills ( $\sigma^2 > \hat{\sigma}^2$ ), the most informed regions as well as the least informed ones may prefer decentralization to centralization. In spite of the larger amount of public spending targeted to those regions, the central government may provide them with a lower expected amount of public goods.

Globally, the regressive transfer of resources unambiguously dominates the progressive transfer of productivity. Thus the aggregate welfare impact of centralization without uniform public-good provision is negative. This result provides a microfounded rationale for the requirement that the central government should provide public goods uniformly across regions. Such a constraint is exogenously assumed in the classic theory of fiscal federalism (Oates 1972), and it is taken to be a technological obstacle that creates costs of centralization. On the contrary, the comparison of Propositions 3 and 4 establishes that in our setting of imperfect political agency it is a necessary condition to turn the efficiency gains from centralization into welfare gains.

However, Propositions 3 and 4 only depicted the stark binary choice between two extremes: uniformity with progressive redistribution on the one hand, and full discretionality with regressive redistribution on the other. In reality, intermediate cases are possible because the government provides many public goods. The uniformity constraint could apply to some but not all of them. Such a nuanced arrangement can make centralization politically easier to achieve by modulating the regional distribution of the efficiency gains.

Formally, suppose that there are multiple public goods  $p = 1, 2, \dots, P$  and that agents have preferences

$$u_t^i = \tilde{u}_t^i + \sum_{p=1}^P \alpha_p \log g_{p,t} \quad \text{with} \quad \sum_{p=1}^P \alpha_p = 1. \quad (16)$$

The welfare weights  $\alpha_p$  coincide with the shares of each public good in the optimal allocation of resources.

A politician has competence  $\eta_{p,t}$  at providing a specific public good  $p$ . Stochastic skills evolve as first-order moving average processes:  $\eta_{p,t} = \varepsilon_{p,t} + \varepsilon_{p,t-1}$ . The shocks  $\varepsilon_{p,t}$  have mean zero and variance  $\sigma^2$ . They are independent and identically distributed across goods as well as over time and across politicians.

Let the  $P$  public goods be partitioned into a set  $\mathcal{U}$  whose centralized provision is subject to a uniformity constraint, and a complementary set  $\mathcal{D}$  of public goods that the central

government can instead provide in different amounts to different regions. The aggregate welfare weight of public goods subject to the uniformity constraint is denoted by

$$\alpha_{\mathcal{U}} \equiv \sum_{p \in \mathcal{U}} \alpha_p \in [0, 1]. \quad (17)$$

Then we obtain the following result.

**Proposition 5** *Centralization increases aggregate social welfare if and only if the centralized provision of a sufficiently important set of public goods is subject to a uniformity constraint ( $\alpha_{\mathcal{U}} \geq \bar{\alpha}_{\mathcal{U}}$  for a threshold  $\bar{\alpha}_{\mathcal{U}} \in (0, 1 - \rho^C)$ ).*

*If the variance of politicians' ability is sufficiently low ( $\sigma^2 \leq \bar{\sigma}^2$  for a threshold  $\bar{\sigma}^2 \in (0, \hat{\sigma}^2]$ ), then for any distribution of voter information every region is better off under centralization than decentralization provided that a uniformity constraint is imposed on public goods whose importance is  $\alpha_{\mathcal{U}}^*(\bar{\theta}, \sigma^2) \in (\bar{\alpha}_{\mathcal{U}}, 1 - \rho^C)$ . The extent of uniformity that makes centralization Pareto-improving is increasing in average voter information ( $\partial \alpha_{\mathcal{U}}^* / \partial \bar{\theta} > 0$ ) and decreasing in the variance of politicians' ability ( $\partial \alpha_{\mathcal{U}}^* / \partial \sigma^2 < 0$ ).*

The relative importance ( $\alpha_{\mathcal{U}}$ ) of public goods whose centralized provision is subject to a uniformity constraint measures the balance between the two opposite distributional forces described by Propositions 3 and 4. The first part of Proposition 5 simply shows that the transition between the two extremes is gradual. Aggregate social welfare is the higher, the more the central government is constrained to provide public goods uniformly. There is an interior threshold  $\bar{\alpha}_{\mathcal{U}}$  at which the uniformity requirement is widespread enough to guarantee that centralization provides higher overall welfare than decentralization.

Most important, the proposition also establishes that a partial uniformity constraint can be designed to fine-tune the distribution of the efficiency gains from centralization. Better incentives for central politicians reduce aggregate rent extraction and thus create an overall surplus that can be shared across regions. Propositions 3 and 4 established that less informed regions are the beneficiaries when public goods are uniformly provided, and more informed ones when their allocation is discretionary (as long as politicians' competence is not too heterogeneous). An appropriate mix between the two modes of public-good provision can succeed at distributing the gains to all regions, regardless of their level of information.

The better-informed regions benefit from the disproportionate influence over the allocation of discretionary public goods, which represent a share  $1 - \alpha_{\mathcal{U}}$  of central-government resources. If this share is larger than rent-extraction by the central government ( $1 - \alpha_{\mathcal{U}} \geq \rho^C$ ) then every region with above-average information ( $\theta_i > \bar{\theta}$ ) strictly prefers the resource allocation induced by centralization. Control of centrally provided goods not subject to the

uniformity constraint is more valuable than the decrease in rent extraction that decentralization would entail (to  $\rho_i^D < \rho^C$ ).

At the same time, if the share of non-uniform public goods is lower than rent-extraction by the central government ( $1 - \alpha_U \geq \rho^C$ ) then every region with below-average information ( $\theta_i < \bar{\theta}$ ) strictly prefers the resource allocation induced by centralization. Losing control of centrally provided discretionary goods is less costly than the increase in rent extraction that decentralization would entail (to  $\rho_i^D > \rho^C$ ).

When rent extraction and unconstrained budgeting by the central government are exactly matched ( $1 - \alpha_U = \rho^C$ ), the endogenous allocation of resources under centralization is a Pareto improvement over decentralization. The key is that political rent extraction has convex costs. As a consequence, more informed regions are happy to accept an increase in rent extraction up to  $\rho^C$  in order to gain control of a share  $1 - \alpha_U$  of the aggregate budget, while less informed regions are not willing to tolerate a further increase in rents above  $\rho^C$  in order to avoid relinquishing control of non-uniform spending.

Proposition 5 establishes that universally beneficial centralization is achieved with a degree of uniformity  $\alpha_U^* < 1 - \rho^C$ . The extent of uniformity required for a Pareto improvement is increasing in voter knowledge ( $\partial \alpha_U^* / \partial \bar{\theta} > 0$ ) because it inversely tracks rent extraction  $\rho^C$ , which in turn decreases with information. In addition, better-informed regions have to be compensated for the deterioration of their politicians' expected competence under centralization. When politicians' skills are more dispersed, high-information regions suffer a larger loss from declining expected competence. Accordingly, the Pareto-improving extent of uniformity is decreasing in the variance of politicians' ability ( $\partial \alpha_U^* / \partial \sigma^2 < 0$ ).

In the limit, Pareto-improving centralization becomes impossible if ability is so heterogeneous that disciplining politicians' rent extraction becomes a lesser concern relative to selecting more skilled politicians, because then better-informed regions can never be compensated for the dilution of their screening ability. However, we view as a natural benchmark the case in which moral hazard is an even greater problem in political agency than adverse selection ( $0 < \sigma^2 \leq \bar{\sigma}^2$ ). Then it is possible to constrain the operation of a central government so that all regions unanimously prefer centralization to decentralization.

We do not formally model a process of bargaining at the constitutional table that would enable regions to agree on a binding uniformity constraint with a domain  $\alpha_U^*$ . On the one hand, such an arrangement is merely second-best: if regions could agree instead to efficient lump-sum transfers, joint surplus maximization would require instead the completely uniform allocation of government spending described by Proposition 3. On the other hand, agreement on a Pareto improvement might not be practically feasible or enforceable.

Nonetheless, the debate within the European Union, whose founding treaties are adopted

by unanimity of the member states, appears consistent with the patterns described by our model. “Core” countries such as Austria, Finland, Germany and the Netherlands complain about low institutional quality and ineffective or corrupt politicians in “peripheral” countries such as Greece, Italy, Portugal, and Spain. Such complaints chime with our model 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 the core and—e.g., in the case of monetary policy—disproportionately takes into account the needs and interests of core countries. Again, this accords with our model of declining policy-making power for the less informed regions. Proposition 5 thus suggests that intra-European frictions may be simple manifestations of a Pareto-improving agreement that actually works to make the Union suboptimal for each member, but crucially advantageous for them all.

## **6 Heterogeneous Preferences vs. Heterogeneous Accountability**

Our key novel contribution lies in establishing the accountability benefits of centralization when voter information varies across regions, as evidenced by Propositions 2, 3 and 5. Yet, Propositions 4 and 5 also proved that centralization is welfare-reducing if it lacks a uniformity constraint.

A microfounded requirement for uniform public good provision by the central government brings to the forefront a trade-off that has been emphasized since Oates (1972). If different regions have different preferences over the allocation of public spending, a uniformity requirement imposes a cost by making it impossible to provide each region with its preferred bundle of public goods. Thus, a trade-off emerges between greater efficiency on the one hand and higher preference matching on the other.

In our framework, both the need for uniformity and the accountability gains from centralization stem from the same source: differences in voter information and thus government accountability across regions. Yet, there is a key difference in the way such heterogeneity affects preference-matching and government efficiency. Proposition 4 has established that centralization without uniformity is welfare-reducing no matter how small the differences in voter information. Conversely, the reduction in rent extraction caused by centralization is monotone increasing in the extent of interregional differences in voter information. As a consequence, our model yields a novel take on the classic trade off between efficiency and preference-matching. Accountability considerations prevail, prompting efficient centralization, if and only if political accountability is more heterogeneous across regions than



preferences over public goods.

In order to provide a formal statement of this result, we consider a large number  $P$  of distinct public goods. The representative resident of region  $l$  has utility

$$u_t^l = \tilde{u}_t^l + \sum_{p=1}^P \alpha_p^l \log g_{l,p,t} \quad \text{with} \quad \sum_{p=1}^P \alpha_p^l = 1. \quad (18)$$

The utility weights  $\alpha_p^l$  coincide with the shares of every public good in the optimal allocation of resources within region  $l$ . Every region's preference vector  $\boldsymbol{\alpha}^l$  is an independent draw from an identical distribution that is symmetric across goods. Thus, the marginal distribution  $\alpha_p^l$  is the same for all  $p$  and has mean  $\mathbb{E}\alpha_p^l = 1/P$ .

The homogeneity of preferences across regions is summarized by a parameter  $\nu > 0$ . In the limit as  $\nu \rightarrow 0$ , information is maximally heterogeneous given the independence of preferences. 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  $\nu \rightarrow \infty$  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  $\nu$  increases, in the sense that any decrease in  $\nu$  entails a mean-preserving spread of  $\alpha_p^l$ .<sup>8</sup>

Voter information  $\theta_l$  is independent of preferences, and is also independently drawn for each region from an identical distribution with mean  $\mathbb{E}\theta_l = \bar{\theta} \in (0, 1)$ . The homogeneity of information across regions is summarized by a parameter  $\iota > 0$ . In the limit as  $\iota \rightarrow 0$ , information is maximally heterogeneous. Each region is either perfectly informed ( $\theta_l = 1$  with probability  $\bar{\theta}$ ) or completely uninformed ( $\theta_l = 0$  with probability  $1 - \bar{\theta}$ ). Conversely, in the limit as  $\iota \rightarrow \infty$ , information is perfectly homogeneous. Every region has the same level of information  $\bar{\theta}$ . The distribution of information contracts smoothly as  $\iota$  increases, in the sense that any decrease in  $\iota$  entails a mean-preserving spread of  $\theta_l$ .<sup>9</sup>

Assuming a continuum of locations to abstract from differences between sample distributions and population distributions, we can establish the following result.

**Proposition 6** *Suppose that preferences and information are independent of each other and across regions. Aggregate social welfare is always higher under decentralization than under*

---

<sup>8</sup>These properties are satisfied for instance if the preference vector  $\boldsymbol{\alpha}^l$  has a symmetric Dirichlet distribution on the regular  $(P - 1)$ -simplex with concentration parameter  $\nu > 0$ . Then  $\text{Var}(\alpha_p^l) = (P - 1) / [P^2 (1 + \nu P)]$ . Our findings are qualitatively independent of this particular functional specification.

<sup>9</sup>These properties are satisfied for instance if information has a beta distribution  $\theta_l \sim B(\bar{\theta}\iota, (1 - \bar{\theta})\iota)$ . Then  $\text{Var}(\theta_l) = \bar{\theta}(1 - \bar{\theta}) / (1 + \nu)$ . Again, our findings do not require assuming this particular distribution.

*centralization without a uniformity constraint. It is higher under centralization with a uniformity constraint than under decentralization if and only if preference homogeneity is above a positive finite threshold ( $v \geq \bar{v}(\iota, \sigma) > 0$ ) that is increasing in information homogeneity ( $\partial \bar{v} / \partial \iota > 0$ ) and in the variance of politicians' ability ( $\partial \bar{v} / \partial \sigma > 0$ ).*

The proposition highlights the irreducible tension between differences in preferences across regions and the need for uniform public-good provision. As Proposition 4 already established for uniform preferences, Proposition 6 confirms more generally that centralization is welfare-reducing if the central government is allowed to operate without a uniformity constraint. The welfare loss is even greater with preference heterogeneity because, as we show in the appendix, centralization also distorts the allocation of talent. 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.

On the other hand, it remains true that centralization yields the efficiency gains described by Proposition 2. Following Proposition 3, if a uniformity constraint is imposed centralization also provides an egalitarian distribution of resources, which is appealing from the point of view of aggregate utilitarian welfare-maximization.

The question is whether these benefits can overcome the cost of imposing a homogeneous basket of public goods onto different regions with different ideal allocations. Intuitively, Proposition 6 establishes that uniform centralized public-good provision is welfare maximizing when preference heterogeneity is low ( $v \geq \bar{v}$ ) and information heterogeneity is high ( $\partial \bar{v} / \partial \iota > 0$ ). The former implies that the costs of uniformity are moderate, and the latter that the efficiency gains from centralization are large. Higher heterogeneity in politicians' ability also makes centralization less attractive ( $\partial \bar{v} / \partial \sigma > 0$ ) because it increases the costs of preference mismatch, which are reflected in the allocation of talent ( $\eta$ ) as well as of resources ( $b$ ).

On the one hand, the proposition captures a departure from the classic Decentralization Theorem when accountability varies across regions. When preference heterogeneity is moderate but not nil, centralization remains optimal, unlike in a setting without political-agency frictions (Oates 1972). On the other hand, we find that it may be impossible to reap the efficiency gains from centralization when they are small relative to differences in preferences across public goods. Centralization is always efficiency-enhancing in the sense of reducing aggregate rent-extraction. Yet political frictions imply that centralization is welfare-reducing if preference heterogeneity is high while information heterogeneity is low.

## 7 Federalism

So far our analysis has compared centralization and decentralization. We have not considered a federal structure that divides powers between the central government and decentralized local governments. Such a division of powers lacks an intuitive justification if preferences are homogeneous, but also if preference heterogeneity is fully symmetric across different goods, as in Proposition 6. Instead, proper federalism becomes intuitively appealing when preferences are homogenous across regions for some public goods, and heterogenous for some others. Then the first set of goods could be provided by the central government, and the second by local governments. This federal arrangement could provide a solution to the trade-off between centralized efficiency and decentralized preference-matching.

To consider this scenario analytically, assume there is a public good that all regions have identical preferences for. We will denote this homogeneously desired public good as good 0. Furthermore, there is another public good over which regions have antithetical preferences. This idiosyncratically preferred public good can be provided in  $L$  varieties. Each region  $l$  benefits exclusively from its own ideal variety  $l$ , and derives no utility at all from any of the  $L - 1$  alternatives. The representative resident of region  $l$  has utility

$$u_t^l = \tilde{u}_t^l + \alpha_0 \log g_{l,0,t} + (1 - \alpha_0) \log g_{l,l,t}, \quad (19)$$

where  $\alpha_0 \in (0, 1)$  is the optimal budget share of the homogeneously desired public good. This parameter also provides our measure of preference homogeneity in this setting. We retain the same distribution of information as in Proposition 6, parametrized by mean  $\bar{\theta}$  and homogeneity  $\iota > 0$ .

The structure of government is described by an allocation of budgets and powers to the two levels of government, local and central. The overall budget is exogenously fixed at  $Lb$ . Since our focus is on distributional tensions on the expenditure side, we assume that all regions contribute equally to the central government budget. Then the budget allocation is entirely defined by the size of the central-government budget  $b^C$ , which determines residually the local-government budgets  $b_l^D = b - b^C/L$  for every  $l$ .

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

Centralized provision of each good may or may not be subject to a uniformity constraint. In the case of the homogeneously desired good, the constraint binds with respect to quantity, forcing the government to spend the same amount of resources on each region. For the idiosyncratically preferred good, the constraint is also binding with respect to variety. Intu-

itively, we disallow the provision of “separate but equal” varieties. If the central government is allowed to differentiate varieties across regions, it must also gain the ability to differentiate in the effective quantity provided. Different varieties are, so to speak, apples and oranges whose quantities cannot be properly compared by an impartial auditor. Thus, a uniformity constraint requires the central government to provide the same idiosyncratically preferred public goods  $g_{l,j,t} = g_{m,j,t}$  to all regions  $l \neq m$  in the same variety  $j$ . To provide to each region its ideal variety, the central government then has to provide all regions with a fully differentiated bundle comprising all  $L$  varieties.

As before, government structure is fully decentralized if each local government provides the residents of its region  $l$  with both the homogeneously desired good  $g_{l,0}$  and their ideal variety of the idiosyncratically preferred good,  $g_{l,l}$  ( $\chi_0 = \chi_1 = 0$ ). Conversely, the government is fully centralized if the central government is tasked with providing both public goods ( $\chi_0 = \chi_1 = 1$ ). The richer structure of this section allows us to consider a more complicated federal system. The central government provides the uniformly preferred public good  $g_{l,0}$  to all regions ( $\chi_0 = 1$ ). At the same time, however, every region has its own local government provide the locally desired variety of the idiosyncratically preferred good,  $g_{l,l}$  ( $\chi_1 = 0$ ).<sup>10</sup>

We begin by characterizing the unique optimal allocation of public funds to the two levels of government in a federal structure.

**Lemma 1** *In a federal system, the unique Pareto efficient budget allocation grants resources to each level of government in exact proportion to the ideal budget share of the goods it is providing ( $b^C = \alpha_0 bL$  and  $b_l^D = (1 - \alpha_0) b$ ).*

When all regions contribute identically to the central government budget, they all agree on a unique optimal budget allocation. This allocation intuitively reflects the desired budget allocation across different public goods. It is independent of differences in government accountability and efficiency because the elasticity of substitution across public goods equals one. Thus, there is perfect balance between the desire to reduce the budget of unaccountable politicians in order to avoid waste (the price effect), and the opposite desire to increase their budget in order to ensure sufficient provision of the respective public goods in spite of inefficiencies (the income effect).

We can then proceed to establish our core findings on federalism and political accountability.

---

<sup>10</sup>It is also theoretically possible in theory to structure a federal system with the opposite allocation of powers ( $\chi_0 = 0$  and  $\chi_1 = 1$ ). Such a government structure is intuitively undesirable, and we prove in the appendix it is unambiguously welfare-reducing.

**Proposition 7** *The fraction of each government's budget that is dissipated in the form of politicians' rents is decreasing in the scope of the government's powers ( $\partial \rho_l^D / \partial \chi_g > 0 > \partial \rho^C / \partial \chi_g$  for  $g = 0, 1$ ).*

*Full centralization minimizes aggregate rent extraction. Rent extraction is lower under a federal system than under full decentralization if and only if preferences are sufficiently homogeneous ( $\alpha_0 \geq \bar{\alpha}_\rho$ ). A federal system is more likely to reduce rent extraction relative to full decentralization when information is more heterogenous ( $\partial \bar{\alpha}_\rho / \partial \iota > 0$ ).*

Political-agency frictions imply endogenous economies of scope in government accountability. Rent extraction, as a fraction of the public budget, declines as the set of public goods the government is responsible for providing increases. These economies of scope, like the decreasing returns to information in Proposition 1, follow from the recursive structure of incentives in the dynamic problem facing incumbent politicians.

Consider the local government of region  $l$ . Incumbent politicians who perceive a value  $R$  of re-election invest  $x_{l,l} = (1 - \alpha_0) \theta_l \phi R$  in providing the locally desired variety of the idiosyncratically preferred good. They also invest  $x_{l,0} = \alpha_0 \theta_l \phi R$  in providing the homogeneously desired good, if its provision falls within their remit. Under complete decentralization the local government has a budget  $b$  and extracts rents  $r_l = b - \theta_l \phi R$ . In a federal structure the local government has budget  $(1 - \alpha_0) b$  and extract rents  $r_l = (1 - \alpha_0) (b - \theta_l \phi R)$ . In both cases, equilibrium rents capture a fraction  $1 - \theta_l \phi R / b$  of the government budget.

Rent extraction would be independent of the scope of public-good provision in a one-shot model with an exogenously given value of re-election  $R$ . In a dynamic environment, however, the appeal of incumbency is endogenous: it varies with expected future rent extraction (equation 11) and thus indirectly with the budget whose allocation the politician administers. If local politicians are responsible for providing both public goods and control the undivided budget  $b$ , they extract as rent a fraction

$$\rho_l^D(0, 0) = \left( 1 + \frac{2\delta\phi}{2-\delta} \theta_l \right)^{-1} \quad (20)$$

of their budget in the stationary equilibrium. If instead they are tasked with providing only the idiosyncratically preferred good, they are accordingly given a lower budget  $(1 - \alpha_0) b$ . The lower budget makes re-election a less powerful incentive, and rent-extraction rises to a fraction

$$\rho_l^D(1, 0) = \left[ 1 + \frac{2\delta\phi}{2-\delta} (1 - \alpha_0) \theta_l \right]^{-1} > \rho_l^D(0, 0) \quad (21)$$

of the local-government budget.

The argument applies identically to the central government. Thus, political accountabil-

ity tends to decline with the creation of multiple layers of government. As in Proposition 2, in this richer setting too rent extraction is unambiguously minimized by full centralization, which exploits both economies of scope and the efficiency benefits of delegating government monitoring to the best monitors.

The comparison between full decentralization and a federal system is instead ambiguous. On the one hand, federalism forfeits economies of scope, which tends to increase rent extraction. On the other hand, it transfers provision of at least the homogeneously preferred good to the central government, which tends to extract lower rents. Federalism tends to be efficiency-enhancing if information is highly heterogeneous across regions ( $\partial \bar{\alpha}_\rho / \partial \iota > 0$ ) because the accountability benefits of creating a federal government are higher. These benefits are also greater when a greater fraction of public spending is controlled by the federal government, with its lower rent-extraction. Thus, federalism also tends to be efficiency-enhancing when preferences are relatively homogeneous across regions ( $\alpha_0 \geq \bar{\alpha}_\rho$ ).

The findings of Proposition 7 are consistent with empirical evidence that bribery is more prevalent in countries with a larger number of administrative tiers (Fan, Lin, and Treisman 2009). In our model, federalism always increases rent extraction compared to centralization, and sometimes also relative to decentralization.

As in the previous sections, we proceed from aggregate rent extraction to social welfare by considering first the distribution of the central government budget.

**Lemma 2** *Centralized provision of the homogeneously preferred public good is proportional to a region's information ( $g_{l,0}^C/g_{m,0}^C = \theta_l/\theta_m$ ) unless a uniformity constraint is imposed. Imposing uniform provision of this good raises aggregate social welfare. It raises welfare in region  $l$  if and only if its voters are less informed than average ( $\theta_l < \bar{\theta}$ ).*

*If the central government provides the idiosyncratically preferred public good, each region receives its ideal variety in proportion to its voters's information ( $g_{l,t}^C/g_{m,t}^C = \theta_l/\theta_m$ ), whether or not a uniformity constraint is imposed. Imposing uniform provision of this good reduces welfare in every region.*

The first result mirrors Proposition 3. When all regions desire the same public good, mandating uniformity is beneficial because it induces a progressive transfer from the informed to the uninformed. The crucial result in the lemma concerns centralization with preference heterogeneity.

The central government always reflects the priorities of different regions according to their level of political knowledge. Hence, as in Proposition 4, public good provision is precisely proportional to voter information. With fully idiosyncratic preferences, however, this result obtains irrespective of a uniformity constraint. When different regions demand different

varieties of public goods, the central government will provide disproportionately those that are being demanded by the most informed. A requirement of uniform provision has no impact on this distortion, but merely adds another one. Every region is wastefully provided with varieties of public goods that the politicians are unable to target at their intended well-informed recipients. As a consequence, uniform provision of idiosyncratically preferred public goods is not only welfare-reducing but starkly Pareto inefficient. It makes every single region worse off than discretionary provision by the central government.

The welfare ranking of different political structures is the following.

**Proposition 8** *A federal system is welfare maximizing if voter information is sufficiently heterogeneous across regions ( $\iota < \bar{\iota}$ ), and furthermore preference heterogeneity lies in an intermediate range ( $\alpha_0 \in (\bar{\alpha}_{D\sim F}, \bar{\alpha}_{F\sim C})$ ), expanding with information heterogeneity ( $\partial \bar{\alpha}_{D\sim F} / \partial \iota > 0$  and  $\partial \bar{\alpha}_{F\sim C} / \partial \iota < 0$ )*

*Full centralization is welfare maximizing if preferences are relatively homogeneous ( $\iota < \bar{\iota}$  and  $\alpha \geq \bar{\alpha}_{F\sim C}$ , or  $\iota \geq \bar{\iota}$  and  $\alpha \geq \bar{\alpha}_{D\sim C}$ ). Full decentralization is welfare maximizing if preferences are strongly heterogeneous ( $\iota < \bar{\iota}$  and  $\alpha \leq \bar{\alpha}_{D\sim F}$ , or  $\iota \geq \bar{\iota}$  and  $\alpha < \bar{\alpha}_{D\sim C}$ ).*

The main result in the proposition concerns the pivotal role of differences in accountability across regions in making a federal system desirable. Welfare can be maximized by a federal structure with two layers of government only if voter information is sufficiently heterogeneous ( $\iota < \bar{\iota}$ ). Instead, when accountability is relatively homogeneous across regions ( $\iota \geq \bar{\iota}$ ) a single tier of government is optimal, and a federal structure is dominated.

As information becomes more homogeneous, less is gained by centralizing provision of the homogeneously desired good. The difference in rent extraction by the central and the local government shrinks, and it disappears in the limit as information becomes homogeneous ( $\iota \rightarrow \infty$ ). Economies of scope in government accountability, instead, do not decline. Thus, complete decentralization is increasingly likely to be preferred to two-tier federalism ( $\partial \bar{\alpha}_{D\sim F} / \partial \iota > 0$ ).

It might seem surprising that, at the same time, complete centralization also becomes more attractive ( $\partial \bar{\alpha}_{F\sim C} / \partial \iota < 0$ ). The intuition behind this result is that the allocative distortions arising from centralization without uniformity also decline when regions are more homogeneously informed. They disappear in the limit as information becomes homogeneous ( $\iota \rightarrow \infty$ ).

The role of preference heterogeneity is intuitive, following Proposition 6. When preferences over public good are almost completely unrelated across regions ( $\alpha_0 \rightarrow 0$ ) full decentralization is optimal. Since idiosyncratically preferred public goods are all-important, welfare is maximized when they are provided by local governments, who are best at preference-

matching. Then these local governments should be granted all powers so as to reduce their rent extraction. Conversely, if voter preferences are almost completely identical across regions ( $\alpha_0 \rightarrow 1$ ), complete centralization is welfare maximizing. When idiosyncratically preferred public goods are unimportant, so is preference-matching, and only rent-minimization matters. Hence, even when government accountability varies sharply across regions, welfare is maximized by a federal structure only if preference heterogeneity is intermediate.

Furthermore, we can characterize unambiguous comparative statics on the variance of the distribution of politicians' competence.

**Corollary 3** *As politicians' ability becomes heterogeneous, welfare becomes less likely to be maximized by full centralization. If accountability is sufficiently heterogeneous, a federal system becomes more likely to be welfare maximizing ( $\partial \bar{\alpha}_{F \sim C} / \partial \sigma > 0 = \partial \bar{\alpha}_{D \sim F} / \partial \sigma$  and  $\partial \bar{v} / \partial \sigma > 0$ ). If accountability is relatively homogeneous, full decentralization becomes more likely to be welfare maximizing ( $\partial \bar{\alpha}_{D \sim C} / \partial \sigma > 0$ ).*

Key to these results is the distortion of electoral screening when the central government provides idiosyncratically preferred public goods. When these are provided by local governments, each region selects, to the best of its imperfect ability, politicians that are skilled at providing its preferred variety, ignoring their skill for provision of any other locally useless variety. Under full centralization, instead, the central politician has to be a jack of all trades and master of none. He is selected to some extent for all skills, and thus his talents please on average all regions—and endogenously more so the more informed ones. Yet, his competence profile fails to fit any region's preferences as precisely as a specialized local politician. When there is greater variation in the pool of political talent, centralization is therefore less likely to be welfare-maximizing ( $\partial \bar{\alpha}_{D \sim C} / \partial \sigma > 0$  and  $\partial \bar{\alpha}_{F \sim C} / \partial \sigma > 0$ ). Indirectly, then, high variance of politicians' skill also makes a federal structure optimal for a wider range of levels of information heterogeneity ( $\partial \bar{v} / \partial \sigma > 0$ ).

We conclude this section by describing the distributional tensions underlying the overall welfare ranking.

**Corollary 4** *Suppose that centralized provision of the homogeneously desired public good is subject to a uniformity constraint. Then the welfare benefits of a federal system relative to full decentralization are greater for regions whose residents are less informed ( $\partial^2 \mathbb{E}u^l / (\partial \theta_i \partial \chi_0) < 0$ ). The welfare benefits of a federal system relative to full centralization are also greater for less informed regions ( $\partial^2 \mathbb{E}u^l / (\partial \theta_i \partial \chi_1) > 0$ ) if the variance of politicians' ability is sufficiently low ( $\sigma^2 \leq \tilde{\sigma}^2$ ).*

This corollary is the intuitive analogue of Propositions 3 and 4. Centralized provision of



the homogeneously desired good, subject to the welfare-maximizing uniformity constraint, implies a transfer of accountability from the informed to the uninformed. Hence, in a move from full decentralization to a federal structure, the benefit accrue to the least informed regions—or identically, the costs are borne disproportionately by the most informed ones.

Conversely, centralized provision of the idiosyncratically preferred good, which cannot be subject to a uniformity constraint, implies a transfer of power from the uninformed to the informed. A move from federalism to full centralization also induces a beneficial increase in the accountability of the central government. This efficiency gain is enjoyed both by all regions. Nonetheless, the overall benefits accrue disproportionately to the most informed one. Identically, when a federal structure is optimal, the costs of excessive centralization fall mostly on the least informed regions.

The model thus implies that the regions with the lowest voter information and the least accountable local governments should be the keenest on proper federalism, with a balanced division of powers and spending between the states and the federal government. The pattern of political discourse in the United States provides suggestive support for this prediction. Appeals to limit the expansion of federal power and protect states' rights tend to be more popular in southern states. The South also displays a higher average prevalence of political corruption (Glaeser and Saks 2006). Its level of voter information is also lower, as shown for instance by data on newspaper circulation in 1970 depicted by Figure 1.

## 8 Externalities and Government Efficiency

The main focus of our analysis has been the effect of centralization in shaping politicians' incentives and thus their rent extraction. Yet, elections also have value as a means for selecting more capable politicians. In our baseline analysis centralization has a lesser impact on this screening function. When preferences are homogeneous, Corollary 1 showed that expected government productivity is identical under centralization and decentralization. The only difference is that a central government provides the same productivity to all regions, while decentralized government are more productive in more informed regions—hence the distributional consequences discussed in Proposition 5. When preferences are heterogeneous, Proposition 6 and Corollary 3 showed that centralization tends to worsen the selection of politicians. However, the opposite is true when public goods generate spillovers across regions, so that a region is not concerned uniquely with local public-good provision.

To establish this result, we introduce externalities with a simple symmetric specification. Returning for simplicity to a single composite public good, the representative resident of

region  $l$  has utility

$$u_t^l = \tilde{u}_t^l + (1 - \xi) \log g_{l,t} + \frac{\xi}{L} \sum_{m=1}^L \log g_{m,t}, \quad (22)$$

where  $\xi \in [0, 1]$  is an index of externalities.

An intuitive interpretation is that there is free mobility of citizens within a federation such as the United States or the European Union. Within a period  $t$ , an individual has probability  $\xi/L$  of moving to any region  $m$  in the union. Alternatively,  $\xi$  could capture altruism, or technological spillovers—the parametrization is such that public-good provision is characterized by constant returns to scale regardless of  $\xi$ .

Then we can prove that externalities introduce unambiguous systematic differences between the productivity of the central government and the average local government.

**Proposition 9** *Suppose that public goods are characterized by interregional spillovers ( $\xi > 0$ ). Then average government productivity is higher under centralization than decentralization ( $\mathbb{E}\hat{\eta}^C > \sum_{l=1}^L \mathbb{E}\hat{\eta}_l^D / L$ ). The productivity advantage of the central government is increasing in the extent of spillovers ( $\partial \left( \mathbb{E}\hat{\eta}^C - \sum_{l=1}^L \mathbb{E}\hat{\eta}_l^D / L \right) / \partial \xi > 0$ ) and more sharply so the higher voter information ( $\partial^2 \left( \mathbb{E}\hat{\eta}^C - \sum_{l=1}^L \mathbb{E}\hat{\eta}_l^D / L \right) / \partial \xi \partial \theta_l > 0$  for all  $l$ ).*

*Aggregate rent extraction is lower under centralization than decentralization ( $\rho^C < \sum_{l=1}^L \rho_l^D / L$ ). The decline in rent upon centralization is increasing in the extent of spillovers ( $\partial \left( \sum_{l=1}^L \rho_l^D / L - \rho^C \right) / \partial \xi > 0$ ) and more sharply so the higher voter information ( $\partial^2 \left( \sum_{l=1}^L \rho_l^D / L - \rho^C \right) / \partial \xi \partial \theta_l > 0$  for all  $l$ ).*

The key new result in the proposition is that internalizing spillovers through centralization increases the screening value of elections, and therefore the expected productivity of politicians who gain re-election. In our model, informed citizens are more likely to support the incumbent if he has proved to be more capable than average. The intensity of popular support, however, depends not only on the extent of ability, but also on its importance. A voter who is informed of the incumbent's poor skills may nonetheless vote for him because of his personal likability or ideological affinity. Yet, voters are less likely to be swayed by such factors when the politician's competence is more important.

If public goods create interregional spillovers, the economic stakes are higher in a union-wide election than in a local election. The ability of a local politician influences only local public goods. The ability of a central politician influences both local public goods and spillovers from other regions. Therefore, a voter who cares about spillovers is keener on electing highly competent politicians at the central than at the regional level. Hence, centralization reduces the influence of random popularity shocks on electoral outcomes.

Just as a sharper voter focus on competence improves the screening value of elections, so does it improve their monitoring values. Thus, externalities not only determine a novel productivity-enhancing impact of centralization, but they also reinforce the finding of efficiency gains in Proposition 2. In the presence of externalities, Proposition 9 establishes that rent extraction is lower under centralization even when information is identical across regions. Moreover, both efficiency advantages of centralization—reduced rent extraction and higher government productivity—are monotone increasing in the extent of spillovers.

The productivity gains from centralized provision of externality-inducing public goods are enjoyed by all regions. Spillovers modify the distributional results of Propositions 3 and 4 by increasing the appeal of centralization for all regions. Moreover, externalities make Pareto-improving centralization more likely than in Proposition 5. In particular, if the central government provides public goods uniformly to all regions, centralization is no longer necessarily costly for informed regions. First, they directly internalize the spillovers from higher public-good provision in uninformed region. Second, a region with above-average information may also enjoy an increased provision of local public goods, because voters in all region monitor more attentively the central government than their local one.

Finally, Proposition 9 shows that externalities and information are complements in generating the efficiency benefits of centralization. This result is intuitive because the electoral mechanism is driven by voters who are simultaneously informed of government activity and highly keen on government productivity. As a consequence, centralizing decision-making for externality-inducing policies becomes more important the higher the level of institutional quality. In the presence of technological costs of centralization—or of preference heterogeneity as in Proposition 6—centralization would be suboptimal when government accountability is too low, and become optimal once it rises above a threshold.

Thus, Proposition 9 can account for the patterns observed in the emergence of modern European states. A key element in the building of state capacity was the centralization of tax-setting and revenue collection, overcoming pre-modern fiscal fragmentation. This creation of central fiscal capacity went hand in hand with the development of political institutions that promoted government accountability and the pursuit of the common interest (Dincecco 2009, 2011; Besley and Persson 2011). In the United States, too, the power of the federal government relative to the states has steadily increased since the founding of the country, while the franchise and democratic institutions have gradually grown more inclusive.

The improvement in politicians' selection and incentives described by Proposition 9 is distinct from the classic benefits of policy coordination (Oates 1972), which would be reflected in an improvement in the allocation of resources, rather than in government productivity. This traditional force is absent from our model when we consider a single public-good bundle,

but re-emerges when we distinguish between different public goods.

Specifically, suppose the government provides a public good  $g$  that generates inter-regional spillovers  $\xi > 0$  and another public good  $h$  whose benefits are purely local. The representative resident of region  $l$  has utility

$$u_t^l = \tilde{u}_t^l + \alpha \left[ (1 - \xi) \log g_{l,t} + \frac{\xi}{L} \sum_{m=1}^L \log g_{m,t} \right] + (1 - \alpha) \log h_{l,t}, \quad (23)$$

where  $\alpha \in (0, 1)$  is the share of resources that would be allocated to the externality-generating public good by a benevolent planner. Then we can prove that following result.

**Proposition 10** *Suppose there are public goods that generate spillovers and others whose benefits are purely local. The share of resources allocated to externality-generating public goods by the central government is socially optimal ( $\beta^C = \alpha$ ). The share resources allocated to externality-generating public goods by decentralized local governments is suboptimal ( $\beta_l^D < \alpha$  for all  $l$ ) and decreasing in the extent of spillovers ( $\partial \beta_l^D / \partial \xi < 0$ ).*

The proposition provides a microfoundation for gains from policy coordination under centralization. The standard theory of fiscal federalism assumes that each local government maximizes welfare in its own region, but that local politicians are exogenously incapable of cooperating to reach Pareto-improving bargains. In our model, instead, even if local politicians could cooperate fully across regional borders, they would have no incentives to maximize welfare.

The only goal of each incumbent is to signal his own ability to his own constituents. The most effective signaling is achieved by ignoring all spillovers, under-providing externality-inducing public goods and over-providing purely local ones. A cooperative reallocation making voters in all regions better off wouldn't convey any useful signal of ability. Each government would wastefully signal its ability at generating welfare for regions that do not vote for its re-election.

Centralization, instead, endogenously aligns politicians' incentives with the optimal allocation of government resources (net of rent extraction) to all public goods. The key difference is that in a union-wide election all beneficiaries of public-good provision vote for the incumbent's re-election. Thus his career concerns take into account all inter-regional spillovers. Only then does he find it optimal to allocate resources in proportion to the full social value of each investment (and each skill).

## 9 Evidence from the Clean Air Act

We test the main empirical prediction of Proposition 3 based on a clear discontinuity in U.S. environmental policy. Up to the 1960s, air pollution had been regulated primarily by state and local governments. The year 1970 marked a dramatic centralizing intervention by the federal government. Federal involvement rested on two pillars: the establishment of the Environmental Protection Agency (EPA), and the passage of the Clean Air Act of 1970 and subsequent amendments, which phased in national air quality standards for a set of criterion pollutants. National standardization stood in sharp contrast with the previous state-based regulations, which had been adopted only by a few states, imposing very heterogeneous standards (U.S. Senate, 1970). We use the Clean Air Act, and the sharp regulatory shift it entailed, to test the distributional predictions of our model for a uniform nation-wide policy.

We consider emissions of sulfur dioxide a very significant and harmful pollutant.<sup>11</sup> At relatively high concentrations,  $SO_2$  pollution has serious adverse health effects. It harms respiratory and cardiovascular functions and is a cause of premature death. Even at much lower levels, it severely damages crops and contributes to acid rain.  $SO_2$  was among the pollutants immediately targeted by the Clean Air Act through the National Ambient Air Quality Standards, starting in 1971. After 1970, emissions of sulfur dioxides and nitrogen oxides started to abate.<sup>12</sup> Nonetheless, the extent of the causal link between the legislation and the subsequent abatement trend is still debated (Greenstone 2004; List and Sturm 2006).

Our empirical investigation focuses on the differential impact of the Clean Air Act across states. We do not aim at estimating the aggregate effect of the policy shift. Instead, we test whether the decline in emissions after the federal takeover in 1970 was faster in states with less informed voters, as predicted by our theory.

The effect of the Clean Air Act on  $SO_2$  emissions is an appropriate natural experiment to test the predictions of Proposition 3. The Act introduced uniform regulation for the entire country, with national air quality standards and a single federal regulator. At the same time, the consequences of  $SO_2$  pollution are largely localized, in contrast to the case, for instance, of  $CO_2$  and global warming. Therefore, emissions abatement generated benefits primarily at the local level, with smaller spillovers across states. In addition, the benefits

---

<sup>11</sup>Our results are qualitatively unchanged if we consider, instead,  $NO_x$  as our dependent variable. The Appendix reports the baseline table. All robustness checks are available on request.

<sup>12</sup>In absolute terms, from 1970 to 1980 aggregate sulfur dioxide emissions in the United States dropped from 31.161 to 25.905 million short tons. Relative to income, emissions fell from 9.11 to 5.40 short tons per million of real 2005 dollars. The primary contributors to this reduction were metals processing and industrial fuel combustion, whose emissions declined respectively from 4.775 to 1.842 and from 4.568 to 2.951 million short tons (U.S. EPA, 2000). In relative terms, emissions from industrial fuel combustion fell from 4.50 to 2.44 short tons per real 2005 \$ million of value added in manufacturing. Emissions from metals processing fell from 63.48 to 20.23 short tons per real 2005 \$ million of value added in primary metal industries.

from  $SO_2$  reduction accrue in the long run only. It is therefore reasonable to assume that some voters were not immediately aware of the positive impact of pollution regulation on their own welfare.

Hence, our model predicts that until 1970 states with uninformed voters suffered from bad environmental regulation. They applied ineffective standards, or no standards whatsoever, because the local government in charge failed to invest money and regulatory effort on air quality control. The introduction of uniform federal requirements starting in 1971 is predicted to have yielded differential benefits that are monotone decreasing in voter information. Our testable hypothesis is that the lower the level of information in a state, the more rapid the decline in pollutant emissions in that state after the enactment of the Clean Air Act, relative to the pre-1970 baseline.

In the robustness analysis, we also test that this faster reduction in pollution corresponds to an improvement in technique, rather than a change in the composition or scale of economic activity. Thus, we control for the concern that reduced emissions could be accompanied by related developments—and particularly a reduction in employment or income—that could make their welfare consequences ambiguous.

## 9.1 Empirical Specification

Our econometric analysis consists of a difference-in-differences estimate of the effect of information on the reduction in emissions following the imposition of national air quality standards in 1970.

We use a balanced panel of the contiguous United States from 1960 to 1981. We choose this time horizon because in 1981 Ronald Reagan took office. In the following years, his environmental policy choices weakened the EPA and curtailed its budget and staff. Moreover, the Reagan administration championed devolution and a general expansion of the role of the states. Therefore, developments after 1981 may have entailed a roll-back of the centralization episode that is the focus of our analysis.

Our baseline regression specification is the following:

$$p_{i,t} = \alpha + \delta_t + \zeta_i + \theta_i t + d_t x'_i \beta + d_t x'_i \gamma (t - 1970) + \varepsilon_{i,t}. \quad (24)$$

We include year fixed effects  $\delta_t$ , state fixed effects  $\zeta_i$ , and state-specific linear time trends  $\theta_i$ . We cluster the standard errors  $\varepsilon_{i,t}$  by state to account for serial correlation of state-specific shocks. Since spatial correlation is also likely to be present, we allow for two-way clustering by year as well as by state (Cameron, Gelbach, and Miller 2011).<sup>13</sup>

---

<sup>13</sup>A concern with this specification is that the number of clusters in the time dimension is relatively small

As the first difference, we compare pollution  $p_{i,t}$  before the Clean Air Act ( $d_t = 0$  from 1960 to 1970) and after its enactment ( $d_t = 1$  from 1971 to 1981). The difference in differences explores differential changes depending on a vector of state characteristics  $x_i$ , including both our key explanatory variable and additional controls.<sup>14</sup> We consider two interactions. First, the interaction with the indicator variable  $d_t$  would capture a level break ( $\beta$ ) in the series upon the introduction of federal emission standards. Second, we add an interaction with the time elapsed since the creation of the EPA ( $d_t(t - 1970)$ ). This would capture a break ( $\gamma$ ) in the trend of emissions after the reform.

The switch to federal regulation should have an impact on the trend rather than the level of emissions. The effect of regulatory changes is necessarily gradual because the object of regulation is durable capital that is only gradually scrapped and replaced. In fact, the standards introduced by the Clean Air Act and its subsequent amendments stipulated more stringent regulation on new pollution sources than on pre-existing ones. Motor vehicles provide an intuitive example: increasingly strict requirements were mandated for successive model-years, so over time tighter standards applied to a steadily growing share of the U.S. vehicle fleet (Kahn 1996). The same pattern holds for stationary sources: a particularly significant case is differential regulation of old and newly built power plants, the main source of  $SO_2$  emissions (Nelson, Tietenberg, and Donihue 1993). In addition to the technological constraint of natural capital turnover, the implementation of the Clean Air Act was also somewhat gradual. The National Ambient Air Quality Standards were defined in 1970, but compliance was expected to be achieved by 1975. The number of operating monitors reading the concentration of air pollutants, a key factor in enforcement, increased steadily throughout the 1970s (Greenstone, 2004).

Thus, the main testable predictions of Proposition 3 is that the coefficient  $\gamma_1$  on newspaper circulation should be positive. Yet, the finding of a differential decline in emissions on the basis of newspaper circulation cannot be taken immediately as an empirical validation of the model. The remaining concern is that information is not randomly assigned across states. It might spuriously correlate with omitted determinants of a differential response to the Clean Air Act. In particular, Proposition 3 predicts that uninformed regions benefit from centralization by receiving a public good their residents desired but were incapable of inducing local politicians to provide. Instead, a uniform national policy could also impose emissions abatements on recalcitrant states whose residents do not share the environmental

---

(22 years). Therefore, replicated all our results with one-way clustering by state only. Since the standard errors are very close in the two cases, we are confident in the validity of two-way clustered standard errors in our application.

<sup>14</sup>We measure all state characteristics taking 1970 as the reference year

preferences of federal policymakers.<sup>15</sup>

To address these concerns, we introduce a battery of controls that capture possible alternative determinants of regional preferences for pollution. Each control is interacted with the structural breaks in the same manner as newspaper circulation, so it should absorb its significance if it is a better proxy for the true cause of a differential impact of the centralization of environmental policy. Our controls belong to two main categories.

First, we introduce economic variables that ought to capture the respective costs of pollution and pollution abatement. We begin with income and population, reflecting the presumption that poorer states can ill afford environmentalism, while more sparsely populated states may incur lower costs from pollution. We also consider measures of industrial specialization. States that specialize in particularly polluting sectors can be expected to have a harder time reducing emissions. Finally, since  $SO_2$  is primarily released by the combustion of fossil fuels, we control for measures of energy consumption.

Our second set of controls captures the political environment of each state, beyond our main focus on voter information. We measure partisan and ideological leaning on the grounds that conservative voters and Republican politician are presumably less invested in environmental protection and more skeptical of government regulation of polluting emissions.

## 9.2 Data

Our outcome of interest is sulfur dioxide emissions, summarized in Table 1. We use the same data as Bulte, List, Strazicich (2007), who obtained from the EPA panel data on emissions in the contiguous United States from 1929 to 1999. The availability of pollution data dictates the level of disaggregation of our analysis. States are the finest geographic unit for which we have emissions data both before and after the Clean Air Act. At the county level, pollution data are completely unavailable before 1969; even after the creation of the EPA, emissions were monitored in a very small subset of counties until the second half of the 1970s.<sup>16</sup>

Our preferred measure of pollution is the emission intensity of income, measured in tons per real dollar. This choice reflects our focus on improvements in technique, and represents our first step in controlling for a potential reduction in the size of economic activity due to a tightening of environmental regulation. We show in the appendix that our results are robust to alternative scalings of the dependent variable, including emissions per capita and the density of emission per square mile.

---

<sup>15</sup>Proposition 6 analyzes theoretically the costs of imposing a homogeneous policy on regions with heterogeneous preferences

<sup>16</sup>Sulfur dioxide monitors were initially operating in 16 counties, and the sample did not grow to 100 until 1974.



Table 2 lists summary statistics for our dataset. We proxy citizens’ information with average daily newspaper circulation per capita. This is a particularly apt measure of voters’ ability to acquire information about government policy. Up to the 1980s, newspapers were Americans’ main source of political news. Moreover, newspaper readers are better informed and more involved in politics than consumers of other media (Graber 1984; Putnam 1993, 2000; Gentzkow 2006; Gentzkow, Shapiro, and Sinkinson 2011). Even in recent years, newspapers remain an influential source of information on politics and environmental issues. When they are subject to greater newspaper coverage, members of Congress more actively support their constituents’ interests (Snyder and Strömberg 2010). We obtain circulation data from Gentzkow, Shapiro, and Sinkinson’s (2011) dataset. Since their data are only available for presidential election years, we select as our measure the average of newspaper circulation per capita in 1968 and 1972.

Figure 1 shows the pattern of newspaper circulation per capita across the United States around 1970. The colors depict four bands whose boundaries are at the cross-state mean and one standard deviation above and below it. Some geographic clustering of high- and low-information states is apparent. Beyond allowing for arbitrary spatial correlation of the residuals through our two-way clustering strategy, we also check that the results are not driven by region-specific common factors. We can focus on the role of information within each Census Region by adding to the vector of explanatory variables  $x_i$  a full set of dummies for the four regions.

Standard economic variables are from the BEA Regional Economic Accounts. We use average personal income both to construct the dependent variable and as a control variable.<sup>17</sup> We use Census population estimates to compute both newspaper circulation per capita and population density.<sup>18</sup> Moreover, we exploit the decomposition of state value added by major industry groups (2-digit SIC sector).

Following Greenstone’s (2002) methodology, we categorize a sector as a heavy polluter on the basis of EPA estimates of its contribution to total emissions by industry as a whole. This procedure clearly identifies five polluting manufacturing industries: Paper and allied products (SIC 26), Chemicals and allied products (SIC 28), Petroleum and coal products (SIC 29), Stone, clay, glass, and concrete (SIC 32), and Primary metal industries (SIC 33). Each of these industry groups accounts for more than 10% of the aggregate industrial emissions. Instead, every other 2-digit SIC sector contributes less than 5% of emissions.

In addition to the share of value added contributed by the sum of these five polluting

---

<sup>17</sup>We transformed amounts originally in current dollars into constant real dollars by using the U.S. GDP deflator.

<sup>18</sup>Land area for each state is taken from the 2000 Census.

industries, our controls include the share of manufacturing and the share of Electric, gas, and sanitary services (SIC 49). The latter is relevant because it includes power generation. Electric utilities were responsible for almost 56% of anthropogenic  $SO_2$  emissions in the United States in 1970 (90% of which from coal-fueled power plants), while the entire industrial sector accounted for 29%—half due to metals processing and the other by the combustion of high-sulfur fuels in a variety of industrial processes.

Data on state energy consumption are from the State Energy Data System (SEDS) database of the Energy Information Administration (EIA). We compute the fossil-fuel intensity of income as the ratio of aggregate consumption of all fossil fuels (in Btu) to aggregate personal income (in real dollars). Additional controls are the shares of coal and of motor gasoline in total fossil-fuel consumption.

We measure ideology by the average DW-Nominate score of the state’s two U.S. senators. The score ranks each member of Congress according to his ideology (left to right), based on legislative roll-call voting behavior (Poole and Rosenthal, 1985). It proxies for the ideological orientation of the voters the senators represent.

We consider two measures of partisanship. First, we compute the share of political offices controlled by Republicans in mid-1970. We consider a total of six positions: the governorship, the two U.S. Senate seats, the majority leaderships in the state senate and in the state house of representatives, and finally the majority in the state delegation to the U.S. House of Representatives. We collected data on state legislatures from the *Book of the States*, and on all other institutions from *Wikipedia*.

Our second measure of partisanship is the share of Republican votes in gubernatorial elections. Gentzkow, Shapiro, and Sinkinson (2011) provide the data for elections that took place in the presidential years 1968 and 1972. We collected data on gubernatorial elections from 1969 to 1971 from *Wikipedia* and the website [www.ourcampaigns.com](http://www.ourcampaigns.com). Our measure of partisanship around 1970 is the average share of the vote won by the Republican party in all gubernatorial elections in the state from 1968 to 1972. We use the same data to construct a standard measure of electoral evenness. For every gubernatorial election, the index is defined by  $1 - |\% \text{ Republican} - 0.5|$ . Hence, it equals zero if either party wins 100% of the votes, and one if the two parties are exactly tied. Again, our overall measure is the average of the indices for all gubernatorial elections in the state from 1968 to 1972.

We also control for proximity to the election of the state’s U.S. senators. The U.S. Senate is divided into three classes of senators serving overlapping six-year terms, so that a third of the seats are scheduled for re-election every two years. In the period of interest, Class 2 senators were up for re-election in 1972, Class 3 senators in 1974, and Class 1 senators in 1976. We include among our controls a full set of dummies for the three possible cases,

i.e., senators of Classes 1 and 2, 1 and 3, or 2 and 3. Finally, we include as a control variable the share of white population (from the U.S. Census), to check for the possibility of environmental racism.

### 9.3 Results

Before turning to our regression analysis, we can starkly visualize the main result in Figure 2. The graph plots average sulfur dioxide emissions for two group of states: those with above-average newspaper circulation in 1970, and those with below-average newspaper circulation. The difference-in-differences emerges clearly: uninformed states have considerably higher average emissions before 1970, and start decreasing them faster than the informed states as soon as national emission standards are introduced by the federal government. The convergence is gradual, but the break in the trend is dramatic.

Table 3 confirms this result in our full regression specification. Column (1), with no controls, shows a differential break in the trend of  $SO_2$  emissions after 1970, which is significant at the 10% level. The impact of the Clean Air Act on pollution is a function of the level of information in each state. As expected, the break in levels is not significant, but the trends change differently as a consequence of the phasing in of federal air quality standards. Consistent with our theoretical prediction, the speed of emissions abatement following the Clean Air Act is inversely proportional to newspaper circulation. If newspaper circulation in 1970 differed across two states by an amount equal to the standard deviation of the cross-state distribution (.05 copies per person), the rate of decline of  $SO_2$  emissions in the 1970s compared to the 1960s was higher in the less informed state by 1.7 percentage points per year.

The following columns sequentially add a set of controls. Their inclusion never has a material impact on our key finding. The point estimate for the coefficient on newspaper circulation is remarkably stable. Moreover, its precision increases as we add controls, reaching the 5% and even 1% significance threshold. These results strengthen the empirical support for the prediction of Proposition 3, and indicate that the effect of newspaper circulation is unlikely to result from its correlation with an omitted variable.

Column (2) adds the basic controls for income and population density. Controlling for income addresses concerns related to the environmental Kuznets curve. According to this line of reasoning, air quality is a luxury good, so poorer states desire a higher level of pollution. The Clean Air Act might then have imposed an inefficient emissions reduction on poor states, compelling them to reduce emissions to a level suitable only for richer regions. However, the coefficient on income is insignificant and has a negative point estimate, which seems to rule

out this interpretation.

Controlling for population density speaks to a similar concern. Since pollution is more harmful in more densely populated areas, due to its localized adverse health effects, the Clean Air Act might have suboptimally compelled a greater reduction of emissions in low-density areas. This does not seem to be the case, since there is no significant difference in differences based on population density—even if the point estimate is positive in this case, and marginally significant in column (3).

Column (3) also finds some evidence that the introduction of federal regulation had differential effects depending on the concentration of polluting manufacturing industries. Surprisingly, the regression suggests that the Clean Air Act immediately induced a greater decline in pollution in states with a lower GDP share of polluting manufacturing. We should not overemphasize this result: the coefficient becomes insignificant when further controls are added in column (7). Moreover, it is paired with a large albeit imprecisely estimated coefficient on the trend, so that any differential effect would seem to disappear before the end of the decade. In any case, the absence of a significant negative coefficient dispels the concern that the federal regulations might have forced overly stringent regulations on regions with a specialization in polluting manufacturing industries.

The result also suggests that the effects of the Clean Air Act, and local policy before its enactment, were not determined primarily by industry lobbying. Under this interpretation, polluting industries would have pressured state politicians into adopting overly lax regulations. This alternative scenario would also have implied a faster emissions reduction after 1970 in states with a greater specialization in polluting manufacturing, albeit with associated welfare gains. The data, however, do not point to a sharper incidence of federal regulation on states with a greater concentration of polluting industries.

Column (4) introduces ideology. We might be concerned that the states most affected by federal environmental regulation are those that ideologically oppose it. We proxy voters' ideology with the ideology of their elected representatives, and precisely with the average DW-Nominate score of each state's U.S. senators. We find no statistically discernible effect. This suggests that differences in the impact of the Clean Air Act are driven by differences in accountability, consistent with our model, rather than in ideology.

Column (5) includes region fixed effects, addressing the concern that the groups of more and less informed states have an imperfect geographic balance, as seen in Figure 1. We find that newspaper readership is an even stronger and more significant determinant of the differential impacts of the Clean Air Act within than across regions.

Finally, Figure 2 might suggest that our findings are purely driven by a correlation between information and the level of pollution before 1970. While this pattern would be consistent

with Proposition 3, our econometric results are much stronger than a mere cross-section for the period before 1970. The difference-in-differences specification allows us to test for state fixed effects and constant state-specific trends. We provide yet stronger supporting evidence in column (6), which controls for  $SO_2$  intensity in 1970.

As expected, this control is highly significant and hardly affected by the inclusion of additional controls in column (7). States that started with higher emissions have to reduce them more rapidly as they converge to a uniform national level of pollution. Nonetheless, the effect of information persists almost unchanged, and is more precisely estimated. This combined evidence supports our reading that newspaper circulation explains not simply 1970 pollution as a whole, but more precisely a specific inefficient component of the emissions level. Consistent with Proposition 3, lower voter information implied greater inefficiency of state regulation, and thus a faster improvement after the switch to federal regulation.

## 9.4 Robustness

Table 4 explores the robustness of our results to different controls for the industrial structure of each state and for its energy consumption. Both the point estimates and the significance of the coefficient on newspaper circulation hardly change across specifications.

Column (1) in this table replicates column (3) of the baseline Table 3. Just as with polluting manufacturing industries, we might be concerned that the Clean Air Act affected primarily states that specialized in manufacturing as a whole, and that as a consequence had optimally adopted non-restrictive standards before its passage. Column (2) finds no evidence of such a pattern, which would imply a significant negative coefficient on the share of manufacturing. Analogously, column (3) shows that there is no differential effect due to specialization in power generation, the single most polluting sector (accounting for 56% of  $SO_2$  emissions, as opposed to 29% for manufacturing).

Columns (4) to (6) control for total consumption of fossil fuels, relative to income, and for its breakdown by primary energy source. As expected, the Clean Air Act induced a differentially faster reduction of  $SO_2$  emissions in states with a greater reliance of coal, the principal source of sulfur dioxide. Coal consumption is essentially a proxy for emissions in 1970, and accordingly there is a tight correspondence between Table 4, column (5), and Table 3, column (6).

Table 5 expands the set of political control beyond the measure of ideology included in column (4) of the baseline Table 3, which is replicated here as column (1). Our baseline results are again remarkably robust. Column (2) controls for the share of white population. The concern arising from environmental justice model is that, absent uniform prescriptive

standards, non-whites face disproportionately more pollution due to their lower political clout. If so, the Clean Air Act should disproportionately reduce emissions in states with a lower percentage of white residents. We find instead that  $SO_2$  emissions declined more rapidly after 1970 in states with a higher share of white population.

Columns (3) and (4) control for partisanship.<sup>19</sup> The Clean Air Act might have a disproportionate impact in more Republican states due to stronger political opposition at the local level. Overall, we find no evidence of a differential effect in Democratic- and Republican-leaning states, whether measuring partisanship by vote shares in the gubernatorial election or by the number of state-wide elections won by the Republican party. It is worth recalling in this context that the Clean Air Act was passed by a Democratic majority in Congress, but the creation of the EPA was proposed by a Republican president, Richard Nixon.

Column (5) controls for a differential effect on states whose elections are more evenly split between the two parties. The absence of a significant effect is consistent with our theoretical model, in which the margin of victory in the election is determined by random shocks rather than by the underlying level of political accountability. Column (6) considers the proximity of U.S. Senate elections, finding no effect.

## 9.5 Discussion

Overall, our empirical evidence supports the theoretical prediction that the Clean Air Act induced a faster decline of pollution in states with lower newspaper circulation. This pattern can be plausibly interpreted according to Proposition 3. The differential emissions abatement represents the welfare gains that the uninformed reap from centralization. Federal intervention solves a regulatory failure in states with low newspaper readership, whose local governments failed to invest efficiently in environmental regulation.

Suggestive evidence of such a regulatory failure in uninformed states can be gleaned from direct measures of local government activity. We look at two different regulatory inputs: expenditures by state and local governments for air quality control, and actual regulations implemented by states and local governments before 1970.

We obtain expenditure data from the U.S. Department of Commerce yearly report: “Environmental Quality Control” (U.S. Bureau of the Census 1971, 1980). The report publishes the yearly expenditure for air quality control by states, counties, and cities. Since the first published report is for the fiscal year 1969, this source allows a rough comparison between one year before the 1970 Clean Air Act to one after: we consider a decade and compare data

---

<sup>19</sup>In our sample, partisanship is distinct from ideology because party affiliation around 1970 masks substantial ideological differences between the very conservative Southern Democrats and the more liberal Democrats in the rest of the country.

for the fiscal year 1969 and 1978. By combining state, county and city data, we construct a measure total expenditure for air quality control regulation for each year and state. Splitting the sample into two groups of states, based on average newspaper circulation in 1970, we find that in 1969 spending relative to GDP in uninformed states was on average 71% as much as in informed states. In 1978, spending in uninformed states rose to 86% of spending in informed states. After the introduction of uniform standards, uninformed states closed half the gap with informed states. Hence, the 1970 policy shift may have fostered convergence of state and local government expenditures in air pollution regulation, with less informed states gradually catching up to more informed ones.

Our second measure of regulatory input is the total number of standards implemented at the state level before 1970. Our source are the hearings of the subcommittee on air and water pollution of the U.S. Senate (U.S. Senate, 1970). The document reports the standards adopted by states and local governments before 1970 on ten pollutants.<sup>20</sup> Counting the number of pollutants that each state had regulated prior to 1970, we find that states with above-average newspaper circulation had adopted four standards on average, while uninformed states had adopted three standards only. The hearings also report the number of states that had proposed or adopted emission standards for sulfur dioxide and for total suspended particulate as a result of the provision of the 1967 Air Quality Act.<sup>21</sup> 50% of the states with high newspaper circulation had proposed or adopted emission standards for the two pollutants, as opposed to only 25% of the states with low newspaper circulation. This suggests that informed states were putting more effort in the development of air quality standards, while uninformed states were in greater need of federal intervention.

The benefits associated with environmental regulation are large and well documented. Chay and Greenstone (2005) estimate the hedonic value of the improvements in air quality induced by the Clean Air Act through their impact on housing prices. By using data on total suspended particulates, they find that better air quality caused a substantial increase in house prices, which amounted in the aggregate to \$45 billion (in 2001 dollars). Furthermore, air pollution is shown to be positively related to infant mortality. Chay and Greenstone (2003) estimate that a reduction in suspended particulates by  $1 \mu\text{g}/\text{m}^3$  is associated to approximately 200 additional infants per year surviving to one year of age in the United States.

---

<sup>20</sup>The ten pollutants are carbon monoxides, beryllium, fluorides, hydrogen sulfide, lead, nitrogen oxides, sulfur dioxides, sulfuric acid, suspended particulates and total oxidants.

<sup>21</sup>The 1967 Air Quality Act required that states establish air quality control regions and that the Department of Health, Education, and Welfare promulgate criteria to serve as the basis for setting emission standards. States would then use the HEW information to set air quality standards. Under the Air Quality Act, states retained autonomy in their decision of setting the criteria.

On the other hand, a legitimate concern is that clean air may come at the cost of a deterioration in local economic conditions. Plants may choose to relocate away from tightly regulated areas, implying an inefficiency at least for a subset of states. At the local level, air quality regulation does affect industrial location and causes reductions in employment, investment and shipments (Henderson 1996). Tight environmental standards affect most sharply heavily polluting industries, which experience a reduction in employment, output, capital stock, and total factor productivity (Greenstone, 2002).

Our finding that the differential effects of the Clean Air Act are not determined by industrial composition suggests a limited role for displacement. We test for displacement more directly by examining differences in differences for outcomes related to economic activity and industrial composition. In Table 6 we run the difference-in-differences specification from equation 24, replacing emissions on the left-hand side with eight variables that may have been affected, directly or indirectly, by abatement efforts mandated by the Clean Air Act. On the right-hand side, in addition to newspaper circulation, we include our fundamental controls: income and population density.

We find no significant differential effect of information on income or population density after 1970. After the passage of the Clean Air Act, states with lower newspaper circulation experienced a faster reduction in pollution, but neither slower income growth nor slower population growth. This finding lends support to the view that federal intervention helped the uninformed without harming them, neither by reducing their income in the short run, nor by reducing the appeal of the state and inducing outmigration in the longer run. The differential reduction in pollution was derived without any impact on the scale of economic activity.

There is limited evidence that more informed states had a slower decline in the share of polluting manufacturing industries after 1970. This suggests that at least some displacement may have occurred, although the significant coefficient is on a break in the level of the series, rather than on its trend, where it would be expected due to the gradual phase-in of environmental regulation.<sup>22</sup> In any case, the economic impact of such displacement is quantitatively minimal. Considering two states whose 1970 newspaper circulation differed by one standard deviation (.05 copies per person), the rate of decline of pollutant emissions in the 1970s compared to the 1960s was higher in the less informed state by 1-2 percentage points per year. At the same time, its share of value added in polluting manufacturing industries would suffer a one-time decline by 0.1 percentage points.

---

<sup>22</sup>Moreover, the statistical significance of this coefficient uniquely depends on the choice of two-way or one-way clustering. If we rely, conservatively, on the larger of the two standard errors, the null hypothesis of no differential impact cannot be rejected.



There is no evidence of displacement for manufacturing as a whole, nor for the single most polluting sector, Electric, gas, and sanitary services (SIC 49, which is not a manufacturing industry). If the differential reduction in pollution in uninformed states after 1970 was accompanied by a change in the composition of economic activity, this change was of very limited size, and restricted in scope to highly polluting manufacturing industries.

Having ruled out changes in scale and substantial changes in composition, we conclude that the differential impact of the Clean Air Act operated through differential improvement in technique. Our baseline analysis showed that uninformed states achieved faster reductions in the emission intensity of income after 1970. Table 6 explores the channel in greater detail and finds evidence in particular of an impact on the fuel intensity of income. Aggregate consumption of fossil fuels fell more quickly in states with lower newspaper circulation. In this case, too, the impact appears to be entirely on technique and not on composition, since there is no differential effect on the share of coal or of motor gasoline in total fuel consumption.

## 10 Conclusions

Political accountability and the quality of government vary across regions within a country like the United States, and across member states of international organizations like the European Union. In this paper, we have shown that such regional differences imply that centralization increases political accountability.

Our model emphasizes the role of differences in voters' information. Rent-seeking politicians have better incentives when their constituents are more informed about the provision of public goods. We have shown that electoral discipline has decreasing returns. Therefore, a central politician answerable to the whole national electorate extracts lower rents than a collection of local politicians, some monitored tightly by well-informed voters and some loosely by poorly-informed constituents. Hence, we have found that centralization reduces rent extraction whenever voter information is heterogeneous across regions. This result can help to explain the steady growth of the federal government over the history of the United States, and the sharp increase in the scope and extent of the powers of the European Union since the 1970s (Alesina, Angeloni, and Schuknecht 2005).

Our model also predicts distributional consequences of centralization when regions have different levels of information. When the central government provides public goods uniformly across the union, the benefits of centralization are monotone decreasing in voter information. We have tested this prediction by analyzing the differential impact of the 1970 Clean Air Act across the United States. Beforehand, environmental regulation was in the hands of state

and local governments. In 1970, the federal government took charge and started introducing uniform national standards. We have found significant and robust evidence of differences-in-differences. Consistent with our theoretical model, centralization of environmental policy induced a faster decline in pollution in states with lower newspaper readership.

The finding that centralization benefits the least informed regions hinges on uniform public goods provision, as in the case of national air quality standards for pollutant emissions. In our model, we have shown conversely that if the central politician can differentiate local public goods provision across regions, he targets the most informed. While uniform policy entails a transfer of accountability from the informed to the uninformed, discretionary spending reflects a transfer of power from the uninformed to the informed. Thus we have identified the balancing role of a uniformity requirement for central-government policies. Some uniformity is necessary for centralization to be welfare increasing. A carefully calibrated constraint can ensure that centralization is a Pareto improvement over decentralization.

Our framework encompasses differences in preferences across regions and externalities from public goods. We have shown that our model provides political-economy microfoundations for the building blocks of the classic theory of fiscal federalism. Centralized public-good provision must be subject to a uniformity constraint, or else it will determine welfare-reducing regressive redistribution. An endogenous trade-off emerges between lower rent extraction under centralization and preference-matching under decentralization. Furthermore, public-good spillovers imply that centralization is more efficient than decentralization because it reduces rent extraction, raises government productivity, and improves the budget allocation across public goods.

Finally, our model sheds light on the characteristics of a federal system with overlapping levels of government: a central federal government and decentralized local governments. Such a structure is meaningful when regions have identical preferences over some public goods but different preferences over some others. We have shown that political-agency frictions determine endogenous economies of scope in government activity. The division of powers in a federal structure then raises rent extraction relative to complete centralization, and possibly also relative to complete centralization. As a result, we have found that federalism can be welfare-maximizing only if regional differences in accountability are sufficiently large.

# A Derivations and Proofs

## A.1. The Model with Multiple Public Goods

Individual  $i$  in period  $t$  derives instantaneous utility

$$u_t^i = \tilde{u}_t^i + \sum_{p=1}^P \alpha_p^i \log g_{p,t}, \quad (\text{A1})$$

where  $\tilde{u}_t^i$  is exogenous utility from private consumption, and  $g_{p,t}$  the provision of public good  $p$ . The relative importance of each good for individual  $i$  is described by the shares  $\alpha_p^i \geq 0$  such that  $\sum_{p=1}^P \alpha_p^i = 1$ .

Public goods are produced by the government with technology

$$g_{p,t} = e^{\eta_{p,t}} x_{p,t}, \quad (\text{A2})$$

where  $x_{p,t}$  measures per capita investment in each public good  $p$ . Productivity  $\eta_{p,t}$  represents the stochastic competence of the incumbent politician in providing good  $p$ . It is independent across public goods, and it follows a first-order moving average process

$$\eta_{p,t} = \varepsilon_{p,t} + \varepsilon_{p,t-1}. \quad (\text{A3})$$

The shocks  $\varepsilon_{p,t}$  are independent and identically distributed across goods, over time, and across politicians. They have support  $[\hat{\varepsilon}, \bar{\varepsilon}]$ , mean zero and variance  $\sigma^2$ .

Each period, the incumbent politician extracts a rent

$$r_t = b - \sum_{p=1}^P x_{p,t}. \quad (\text{A4})$$

All voters have identical rational expectations that public-good provision if the challenger wins the election is going to provide utility

$$\mathbb{E} \left( \sum_{p=1}^P \alpha_p^i \log g_{p,t+1}^C \right) = \sum_{p=1}^P \alpha_p^i \log \bar{x}_p. \quad (\text{A5})$$

Uninformed voters expect no difference in the utility value public-good provision whether the incumbent is reelected or the challenger defeats him:

$$\Delta_0 \equiv \mathbb{E} \left( \sum_{p=1}^P \alpha_p^i \log g_{p,t+1}^I - \sum_{p=1}^P \alpha_p^i \log g_{p,t+1}^C \right) = 0. \quad (\text{A6})$$

Informed voters, instead, expect a difference in the utility of public goods provided by a

relected incumbent relative to a victorious challenger equal to

$$\begin{aligned}\Delta_1^i(\mathbf{g}_t) &\equiv \mathbb{E} \left[ \sum_{p=1}^P \alpha_p^i (\log g_{p,t+1}^I | \mathbf{g}_t) - \sum_{p=1}^P \alpha_p^i \log g_{p,t+1}^C \right] \\ &= \sum_{p=1}^P \alpha_p^i \mathbb{E}(\varepsilon_{p,t} | g_{p,t}) = \sum_{p=1}^P \alpha_p^i (\log g_{p,t} - \log \bar{x}_p - \varepsilon_{p,t-1}).\end{aligned}\quad (\text{A7})$$

In a rational expectations equilibrium their inference turns out to be perfect ( $x_t = \bar{x}$ ), accurately revealing  $\varepsilon_t$ .

The electorate consists of a continuum of atomistic voters, which can be partitioned into  $J$  groups. Group  $j$  comprises a fraction  $\lambda_j$  of voters, who have identical preferences  $\alpha^j$  and an identical probability  $\theta_j$  of information acquisition. Each group  $j$  comprises a continuum of agents and the arrival of information is independent across agents, so a share  $\theta_j$  of its members have observed public goods provision  $\mathbf{g}_t$ , while the remainder  $1 - \theta_j$  have not.

The preference shocks  $\Psi_t \sim U[-1/(2\phi), 1/(2\phi)]$  and  $\psi_t^i \sim U[-\bar{\psi}, \bar{\psi}]$  have a sufficiently wide support, and the competence shocks  $\varepsilon_{p,t}$  have sufficiently narrow support, that

$$\frac{1}{2\phi} - \bar{\psi} \leq \check{\varepsilon} < \hat{\varepsilon} \leq \bar{\psi} - \frac{1}{2\phi} \quad \text{and} \quad -\frac{1}{2\phi} \leq \check{\bar{\theta}} < \hat{\bar{\theta}} \leq \frac{1}{2\phi}.\quad (\text{A8})$$

Given the independent realizations of the uniform idiosyncratic shock  $\psi^i$ , the share of members of group  $j$  who vote for the incumbent is

$$\begin{aligned}v_t^j(\mathbf{g}_t, \Psi_t) &= \theta_j \Pr(\psi_t^i \leq \Delta_1^j(\mathbf{g}_t) - \Psi_t) + (1 - \theta_j) \Pr(\psi_t^i \leq -\Psi_t) \\ &= \frac{1}{2} + \frac{1}{2\bar{\psi}} [\theta_j \Delta_1^j(\mathbf{g}_t) - \Psi_t],\end{aligned}\quad (\text{A9})$$

conditional on the realizations of  $\mathbf{g}_t$  and  $\Psi_t$ . Taking into account the uniform aggregate shock  $\Psi_t$ , the incumbent's probability of re-election is

$$\begin{aligned}\pi(\mathbf{g}_t) &= \Pr\left(\sum_{j=1}^J \lambda_j v_t^j(\mathbf{g}_t, \Psi_t) \geq \frac{1}{2}\right) = \Pr\left(\Psi_t \leq \sum_{j=1}^J \theta_j \lambda_j \Delta_1^j(\mathbf{g}_t)\right) \\ &= \frac{1}{2} + \phi \sum_{j=1}^J \theta_j \lambda_j \Delta_1^j(\mathbf{g}_t)\end{aligned}\quad (\text{A10})$$

conditional on the realizations of public-good provision  $\mathbf{g}_t$ . Thus, taking into account the mean-zero competence shocks  $\varepsilon_{p,t}$ , the incumbent's probability of re-election is

$$\pi(\mathbf{x}_t) = \mathbb{E}[\pi(\mathbf{g}_t) | \mathbf{x}_t] = \frac{1}{2} + \phi \sum_{j=1}^J \theta_j \lambda_j \sum_{p=1}^P \alpha_p^j (\log x_{p,t} - \log \bar{x}_p)\quad (\text{A11})$$

as a function of his policy choices  $\mathbf{x}_t$  (and residually  $r_t$ ).

## A.2. Proof of Proposition 1 and Corollary 1

The trade-off between current and future rent extraction leads to policy choices

$$\mathbf{x}(R) = \arg \max_{\mathbf{x}_t} \left\{ b - \sum_{p=1}^P x_{p,t} + R\pi(\mathbf{x}_t) \right\}, \quad (\text{A12})$$

namely

$$x_p(R) = \phi R \sum_{j=1}^J \theta_j \lambda_j \alpha_p^j \text{ for all } p = 1, \dots, P, \quad (\text{A13})$$

and thus current rent extraction

$$r(R) = b - \phi \bar{\theta} R. \quad (\text{A14})$$

By equation (11), equilibrium rent-extraction is

$$r = b \left( 1 + \frac{2\delta}{2 - \delta} \phi \bar{\theta} \right)^{-1}, \quad (\text{A15})$$

which is decreasing and convex in  $\bar{\theta}$ .

Investment in the provision of public good  $p$  is  $x_p = \beta_p (1 - \rho) b$ , with relative shares

$$\beta_p \equiv \sum_{j=1}^J \frac{\theta_j}{\bar{\theta}} \lambda_j \alpha_p^j. \quad (\text{A16})$$

The incumbent is re-elected if and only if

$$\Psi_t \leq \sum_{j=1}^J \theta_j \lambda_j \sum_{p=1}^P \alpha_p^j \varepsilon_{p,t}. \quad (\text{A17})$$

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

$$\hat{\eta}_t = \chi_{t-1} (\varepsilon_{t-1}^I + \varepsilon_t^I) + (1 - \chi_{t-1}) (\varepsilon_{t-1}^C + \varepsilon_t^C), \quad (\text{A18})$$

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

$$\begin{aligned}
\Pr(\hat{\eta}_{p,t} \leq \eta) &= \Pr[\chi_{t-1}(\varepsilon_{p,t-1}^I + \varepsilon_{p,t}^I) + (1 - \chi_{t-1})(\varepsilon_{p,t-1}^C + \varepsilon_{p,t}^C) \leq \eta] \\
&= \Pr(\chi_{t-1} = 1 \wedge \varepsilon_{p,t-1}^I + \varepsilon_{p,t}^I \leq \eta) + \Pr(\chi_{t-1} = 0 \wedge \varepsilon_{p,t-1}^C + \varepsilon_{p,t}^C \leq \eta) \\
&= \Pr\left(\Psi_{t-1} \leq \sum_{j=1}^J \lambda_j \theta_j \sum_{q=1}^P \alpha_q^j \varepsilon_{q,t-1} \wedge \varepsilon_{p,t-1}^I + \varepsilon_{p,t}^I \leq \eta\right) + \frac{1}{2} \Pr(\varepsilon_{p,t-1}^C + \varepsilon_{p,t}^C \leq \eta) \\
&= \int_{-\infty}^{\infty} \left(1 + \varepsilon \phi \sum_{j=1}^J \lambda_j \theta_j \alpha_p^j\right) F_\varepsilon(\eta - \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon, \quad (\text{A19})
\end{aligned}$$

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

$$\int_{-\infty}^{\infty} \varepsilon F_\varepsilon(\eta - \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon = \mathbb{E}[\varepsilon F_\varepsilon(\eta - \varepsilon)] < \mathbb{E}\varepsilon \mathbb{E}[F_\varepsilon(\eta - \varepsilon)] = 0, \quad (\text{A20})$$

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

$$\begin{aligned}
\mathbb{E}\hat{\eta}_{p,t} &= \mathbb{E}(\chi_{t-1} \varepsilon_{p,t-1}) = \int_{-\infty}^{\infty} \left(\frac{1}{2} + \phi \sum_{j=1}^J \lambda_j \theta_j \sum_{q=1}^P \alpha_q^j \varepsilon_q\right) \varepsilon_p f_\varepsilon(\varepsilon_p) d\varepsilon_p \\
&= \phi \sigma^2 \sum_{j=1}^J \lambda_j \theta_j \alpha_p^j. \quad (\text{A21})
\end{aligned}$$

The equilibrium utility of each member of group  $j$  equals

$$\mathbb{E}u_j = \sum_{p=1}^P \alpha_p^j \mathbb{E} \log g_{p,t} = \log b + \log(1 - \rho) + \sum_{p=1}^P \alpha_p^j (\mathbb{E}\hat{\eta}_p + \log \beta_p). \quad (\text{A22})$$

If voters have identical preferences  $\alpha^j = \alpha$  for all  $j$ , social welfare is simply

$$\mathbb{E}u = \log b + \log(1 - \rho) + \sum_{p=1}^P \alpha_p (\alpha_p \phi \bar{\theta} \sigma^2 + \log \alpha_p), \quad (\text{A23})$$

which is increasing and concave in  $\bar{\theta}$ .

### A.3. Equilibrium with Many Regions

When the economy is divided into  $L$  regions, there are  $LP$  public goods:  $g_{l,p,t}$  is the provision of public good  $p$  in location  $l$  at time  $t$ .

We allow for the externalities in public-good provision, measured by an index  $\xi_p \in [0, 1]$ .

A resident of region  $l$  derives utility

$$\alpha_{l,p}^l = \left(1 - \frac{L-1}{L}\xi_p\right) \alpha_p^l > 0 \quad (\text{A24})$$

from the amount of public good  $p$  provided to his own region, but he may also derive additional utility

$$\alpha_{m,p}^l = \frac{1}{L}\xi_p \alpha_p^l \text{ for } l \neq m \quad (\text{A25})$$

from the amount provided in each other region. Thus we can write the utility of individual  $i$  in region  $l$  as

$$u_t^i = \tilde{u}_t^i + \sum_{p=1}^P \alpha_p^l \left[ (1 - \xi_p) \log g_{l,p,t} + \frac{\xi_p}{L} \sum_{m=1}^L \log g_{m,p,t} \right]. \quad (\text{A26})$$

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

$$r_{l,t}^D = b - \sum_{p=1}^P x_{l,p,t}^D. \quad (\text{A27})$$

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

$$r_t^C = bL - \sum_{l=1}^L \sum_{p=1}^P x_{l,p,t}^C. \quad (\text{A28})$$

### A.3.1 Decentralization

Under decentralization equilibrium rent extraction is

$$\rho_t^D = \left[ 1 + \frac{2\delta}{2-\delta} \phi \left( 1 - \frac{L-1}{L} \sum_{p=1}^P \xi_p \alpha_p^l \right) \theta_l \right]^{-1}, \quad (\text{A29})$$

the expected ability of a local politician is

$$\mathbb{E} \hat{\eta}_{l,p}^D = \phi \sigma^2 \left( 1 - \frac{L-1}{L} \xi_p \right) \alpha_p^l \theta_l, \quad (\text{A30})$$

and the relative shares of each local public good are

$$\beta_{l,p}^D = \frac{\left(1 - \frac{L-1}{L}\xi_p\right) \alpha_p^l}{1 - \frac{L-1}{L} \sum_{q=1}^P \xi_q \alpha_q^l}. \quad (\text{A31})$$

Welfare in region  $l$  is

$$\mathbb{E}u_l^D = \log b + \sum_{p=1}^P \alpha_p^l \left\{ \begin{array}{l} (1 - \xi_p) [\log(1 - \rho_l^D) + \mathbb{E}\hat{\eta}_{l,p}^D + \log \beta_{l,p}^D] \\ + \frac{\xi_p}{L} \sum_{m=1}^L [\log(1 - \rho_m^D) + \mathbb{E}\hat{\eta}_{m,p}^D + \log \beta_{m,p}^D] \end{array} \right\}, \quad (\text{A32})$$

and aggregate welfare is

$$W^D = \log b + \frac{1}{L} \sum_{l=1}^L \sum_{p=1}^P [(1 - \xi_p) \alpha_p^l + \xi_p \bar{\alpha}_p] [\log(1 - \rho_l^D) + \mathbb{E}\hat{\eta}_{l,p}^D + \log \beta_{l,p}^D]. \quad (\text{A33})$$

for

$$\bar{\alpha}_p = \frac{1}{L} \sum_{l=1}^L \alpha_p^l. \quad (\text{A34})$$

### A.3.2 Centralization

Under centralization equilibrium rent extraction is

$$\rho^C = \left(1 + \frac{2\delta}{2 - \delta} \phi \bar{\theta}\right)^{-1} \text{ for } \bar{\theta} = \frac{1}{L} \sum_{l=1}^L \theta_l, \quad (\text{A35})$$

and the expected ability of a central politician is

$$\mathbb{E}\hat{\eta}_p^C = \frac{\phi \sigma^2}{L} \sum_{l=1}^L \theta_l \alpha_p^l. \quad (\text{A36})$$

We partition the  $P$  public goods into a set  $\mathcal{U}$  of public goods whose centralized provision is subject to a uniformity constraint, and a complementary set  $\mathcal{D}$  of public goods that the central government can instead provide in different amounts to different regions. Then the relative shares of each local public good are

$$\beta_p^C = \frac{1}{L^2} \sum_{l=1}^L \frac{\theta_l}{\bar{\theta}} \alpha_p^l \text{ for } p \in \mathcal{U} \quad (\text{A37})$$

and

$$\beta_{l,p}^C = \frac{1}{L} \left[ (1 - \xi_p) \frac{\theta_l}{\bar{\theta}} \alpha_p^l + \frac{\xi_p}{L} \sum_{m=1}^L \frac{\theta_m}{\bar{\theta}} \alpha_p^m \right] \text{ for } p \in \mathcal{D}. \quad (\text{A38})$$



Welfare in region  $l$  is

$$\begin{aligned} \mathbb{E}u_l = \log(Lb) + \log(1 - \rho^C) + \sum_{p=1}^P \alpha_p^l \mathbb{E}\hat{\eta}_p^C + \sum_{p \in \mathcal{U}} \alpha_p^l \log \beta_p^C \\ + \sum_{p \in \mathcal{D}} \alpha_p^l \left[ (1 - \xi_p) \log \beta_{l,p}^C + \frac{\xi_p}{L} \sum_{m=1}^L \log \beta_{m,p}^C \right] \end{aligned} \quad (\text{A39})$$

and aggregate welfare is

$$\begin{aligned} W^C = \log(Lb) + \log(1 - \rho^C) + \sum_{p=1}^P \bar{\alpha}_p \mathbb{E}\hat{\eta}_p^C + \sum_{p \in \mathcal{U}} \bar{\alpha}_p \log \beta_p^C \\ + \frac{1}{L} \sum_{l=1}^L \sum_{p \in \mathcal{D}} \alpha_p^l \left[ (1 - \xi_p) \log \beta_{l,p}^C + \frac{\xi_p}{L} \sum_{m=1}^L \log \beta_{m,p}^C \right]. \end{aligned} \quad (\text{A40})$$

#### A.4. Proof of Proposition 2

Aggregate rent extraction is lower under centralization if and only if

$$\rho^C \leq \frac{1}{L} \sum_{l=1}^L \rho_l^D, \quad (\text{A41})$$

which can be written

$$f\left(\frac{1}{L} \sum_{l=1}^L \theta_l\right) \leq \frac{1}{L} \sum_{l=1}^L f\left(\left(1 - \frac{L-1}{L} \sum_{p=1}^P \xi_p \alpha_p^l\right) \theta_l\right) \quad (\text{A42})$$

for

$$f(x) \equiv \left(1 + \frac{2\delta}{2-\delta} \phi x\right)^{-1} \quad (\text{A43})$$

a strictly decreasing and strictly convex function of  $x > 0$ :

$$f'(x) = -\frac{2\delta}{2-\delta} \phi \left(1 + \frac{2\delta}{2-\delta} \phi x\right)^{-2} < 0 \quad (\text{A44})$$

and

$$f''(x) = \left(\frac{2\delta}{2-\delta} \phi\right)^2 \left(1 + \frac{2\delta}{2-\delta} \phi x\right)^{-3} > 0.$$

Thus

$$f\left(\frac{1}{L} \sum_{l=1}^L \theta_l\right) \leq \frac{1}{L} \sum_{l=1}^L f(\theta_l) \leq \frac{1}{L} \sum_{l=1}^L f\left(\left(1 - \frac{L-1}{L} \sum_{p=1}^P \xi_p \alpha_p^l\right) \theta_l\right). \quad (\text{A45})$$

The first inequality follows from convexity by Jensen's inequality and holds strictly if  $\theta_l$  is heterogeneous across regions. The second inequality follows from monotonicity and holds strictly if  $\xi_p > 0$  for some  $p$ .

## A.5. The Decentralization Theorem

**Proposition A1** *Suppose that all public goods provided by the central government are subject to the uniformity constraint ( $\mathcal{D} = \emptyset$ ) and that information is homogeneous across regions ( $\theta_l = \theta$  for all  $l$ ).*

1. *If there are no externalities and preferences are homogeneous ( $\alpha_p^l = \alpha_p$  for all  $l$  and  $\xi_p = 0$  for all  $p$ ), then centralization and decentralization yield identical outcomes.*
2. *If there are externalities and preferences are homogeneous ( $\alpha_p^l = \alpha_p$  for all  $l$ , while  $\xi_p > 0$  for some  $p$ ), then centralization yields higher welfare than decentralization.*
3. *If there are no externalities and preferences are heterogeneous ( $\xi_p = 0$  for all  $p$ , while  $\alpha_p^l \neq \alpha_p^m$  for some  $l \neq m$  and  $p$ ), then decentralization yields higher welfare than centralization.*

Suppose that all public goods provided by the central government are subject to the uniformity constraint ( $\mathcal{D} = \emptyset$ ) and that information is homogeneous across regions ( $\theta_l = \theta$  for all  $l$ ). Then under centralization

$$\rho^C = \left(1 + \frac{2\delta}{2-\delta}\phi\theta\right)^{-1}, \quad \mathbb{E}\hat{\eta}_p^C = \phi\sigma^2\theta\bar{\alpha}_p, \quad \text{and} \quad \beta_p^C = \frac{1}{L}\bar{\alpha}_p. \quad (\text{A46})$$

while under decentralization

$$\rho_l^D = \left[1 + \frac{2\delta}{2-\delta}\phi\theta \left(1 - \frac{L-1}{L} \sum_{p=1}^P \xi_p \alpha_p^l\right)\right]^{-1}, \quad (\text{A47})$$

$$\mathbb{E}\hat{\eta}_{l,p}^D = \phi\sigma^2\theta \left(1 - \frac{L-1}{L}\xi_p\right) \alpha_p^l, \quad (\text{A48})$$

and

$$\beta_{l,p}^D = \left(1 - \frac{L-1}{L}\xi_p\right) \alpha_p^l. \quad (\text{A49})$$

1. *If there are no externalities and preferences are homogeneous ( $\alpha_p^l = \alpha_p$  for all  $l$  and  $\xi_p = 0$  for all  $p$ ), then*

$$\rho^C = \rho_l^D = \left(1 + \frac{2\delta}{2-\delta}\phi\theta\right)^{-1}, \quad \mathbb{E}\hat{\eta}_p^C = \mathbb{E}\hat{\eta}_{l,p}^D = \phi\sigma^2\theta\alpha_p \quad \text{and} \quad L\beta_p^C = \beta_{l,p}^D = \alpha_p. \quad (\text{A50})$$

2. If there are externalities and preferences are homogeneous ( $\alpha_p^l = \alpha_p$  for all  $l$ , while  $\xi_p > 0$  for some  $p$ ), then under centralization

$$\rho^C = \left(1 + \frac{2\delta}{2 - \delta} \phi \theta\right)^{-1}, \quad \mathbb{E}\hat{\eta}_p^C = \phi \sigma^2 \theta \alpha_p, \quad \text{and } L\beta_p^C = \alpha_p, \quad (\text{A51})$$

and

$$W^C = \log b + \log(1 - \rho^C) + \sum_{p=1}^P \alpha_p (\mathbb{E}\hat{\eta}_p^C + \log \alpha_p). \quad (\text{A52})$$

Under decentralization

$$\rho_l^D = \left[1 + \frac{2\delta}{2 - \delta} \phi \theta \left(1 - \frac{L-1}{L} \sum_{p=1}^P \xi_p \alpha_p\right)\right]^{-1} > \rho^C, \quad (\text{A53})$$

$$\mathbb{E}\hat{\eta}_{l,p}^D = \phi \sigma^2 \theta \left(1 - \frac{L-1}{L} \xi_p\right) \alpha_p < \mathbb{E}\hat{\eta}_p^C, \quad (\text{A54})$$

$$\beta_{l,p}^D = \frac{\left(1 - \frac{L-1}{L} \xi_p\right) \alpha_p}{1 - \frac{L-1}{L} \sum_{q=1}^P \xi_q \alpha_q}, \quad (\text{A55})$$

and

$$W^D = \log b + \log(1 - \rho_l^D) + \sum_{p=1}^P \alpha_p (\mathbb{E}\hat{\eta}_{l,p}^D + \log \beta_{l,p}^D). \quad (\text{A56})$$

Centralization achieves the optimal allocation of productive expenditures, while decentralization does not unless  $\xi_p$  is homogeneous across goods.

Thus welfare is lower under decentralization due to increased rent extraction, lower government efficiency, and also misallocation of expenditures across public goods unless  $\xi_p$  is homogeneous across goods.

3. If there are no externalities and preferences are heterogeneous ( $\xi_p = 0$  for all  $p$ , while  $\alpha_p^l \neq \alpha_p^m$  for some  $l \neq m$  and  $p$ ), then under centralization

$$W^C = \log b + \log(1 - \rho^C) + \frac{1}{L} \sum_{l=1}^L \sum_{p=1}^P \alpha_p^l [\mathbb{E}\hat{\eta}_p^C + \log(L\beta_p^C)], \quad (\text{A57})$$

while under decentralization

$$\rho_l^D = \left(1 + \frac{2\delta}{2 - \delta} \phi \theta\right)^{-1} = \rho^C, \quad \mathbb{E}\hat{\eta}_{l,p}^D = \phi \sigma^2 \theta \alpha_p^l, \quad \beta_{l,p}^D = \alpha_p^l \quad (\text{A58})$$

and

$$W^D = \log b + \log(1 - \rho_l^D) + \frac{1}{L} \sum_{l=1}^L \sum_{p=1}^P \alpha_p^l (\mathbb{E}\hat{\eta}_{l,p}^D + \log \beta_{l,p}^D). \quad (\text{A59})$$

Decentralization achieves the optimal allocation of productive expenditures, while centralization does not. Moreover, decentralization achieves a better screening of politicians

$$\frac{1}{L} \sum_{l=1}^L \sum_{p=1}^P \alpha_p^l \mathbb{E} \hat{\eta}_{l,p}^D > \frac{1}{L} \sum_{l=1}^L \sum_{p=1}^P \alpha_p^l \mathbb{E} \hat{\eta}_p^C \quad (\text{A60})$$

because for each public good  $p$

$$\frac{1}{L} \sum_{l=1}^L (\alpha_p^l)^2 > \left( \frac{1}{L} \sum_{l=1}^L \alpha_p^l \right)^2 \quad (\text{A61})$$

unless  $\alpha_p^l = \alpha_p$  for all  $l$ .

## A.6. Proof of Proposition 3

Suppose that all public goods provided by the central government are subject to the uniformity constraint ( $\mathcal{D} = \emptyset$ ), and that preferences are homogeneous across regions ( $\alpha_p^l = \alpha_p$  for all  $l$ ).

Then under centralization

$$\rho^C = \left( 1 + \frac{2\delta}{2-\delta} \phi \bar{\theta} \right)^{-1}, \quad \mathbb{E} \hat{\eta}_p^C = \phi \sigma^2 \alpha_p \bar{\theta}, \quad \text{and} \quad \beta_p^C = \alpha_p, \quad (\text{A62})$$

so

$$g_{l,p,t}^C = \frac{1}{L} x_{p,t}^C \exp(\hat{\eta}_p^C) = \beta_p^C b^D (1 - \rho^C) \exp(\hat{\eta}_p^C) \quad (\text{A63})$$

and

$$W^C = \log b + \log(1 - \rho^C) + \sum_{p=1}^P \alpha_p (\mathbb{E} \hat{\eta}_p^C + \log \beta_p^C). \quad (\text{A64})$$

Under decentralization

$$\rho_l^D = \left[ 1 + \frac{2\delta}{2-\delta} \phi \left( 1 - \frac{L-1}{L} \sum_{p=1}^P \xi_p \alpha_p \right) \theta_l \right]^{-1}, \quad (\text{A65})$$

$$\mathbb{E} \hat{\eta}_{l,p}^D = \phi \sigma^2 \left( 1 - \frac{L-1}{L} \xi_p \right) \alpha_p \theta_l, \quad (\text{A66})$$

and

$$\beta_{l,p}^D = \frac{(1 - \frac{L-1}{L} \xi_p) \alpha_p}{1 - \frac{L-1}{L} \sum_{q=1}^P \xi_q \alpha_q}, \quad (\text{A67})$$

so

$$g_{l,p,t}^D = x_{p,t}^D \exp(\hat{\eta}_{l,p}^D) = \beta_{l,p}^D b^D (1 - \rho_l^D) \exp(\hat{\eta}_{l,p}^D) \quad (\text{A68})$$

and

$$W^D = \frac{1}{L} \sum_{l=1}^L \left[ \log b + \log(1 - \rho_l^D) + \sum_{p=1}^P \alpha_p (\mathbb{E} \hat{\eta}_{l,p}^D + \log \beta_{l,p}^D) \right]. \quad (\text{A69})$$

Note that  $\partial \beta_{l,p}^D / \partial \theta_l = 0$  and  $\partial \rho_l^D / \partial \theta_l < 0$ . Recalling the proof of Corollary 1,  $\hat{\eta}_{l,p}^D$  is increasing in  $\theta_l$  in the sense of first-order stochastic dominance, so  $\partial \mathbb{E} \exp(\hat{\eta}_{l,p}^D) / \partial \theta_l > 0$ . Since  $g_{l,p}^C$  is identical for all regions, it follows that

$$\mathbb{E}(g_{l,p}^C - g_{l,p}^D) > \mathbb{E}(g_{m,p}^C - g_{m,p}^D) \Leftrightarrow \mathbb{E}g_{l,p}^D < \mathbb{E}g_{m,p}^D \Leftrightarrow \theta_l < \theta_m. \quad (\text{A70})$$

Similarly for residents' welfare

$$\begin{aligned} \mathbb{E}(u_l^C - u_l^D) > \mathbb{E}(u_m^C - u_m^D) &\Leftrightarrow \mathbb{E}u_l^D < \mathbb{E}u_m^D \Leftrightarrow \\ &\sum_{p=1}^P \alpha_p (1 - \xi_p) [\log(1 - \rho_l^D) + \mathbb{E} \hat{\eta}_{l,p}^D] < \sum_{p=1}^P \alpha_p (1 - \xi_p) [\log(1 - \rho_l^D) + \mathbb{E} \hat{\eta}_{l,p}^D] \Leftrightarrow \\ &\theta_l < \theta_m \end{aligned} \quad (\text{A71})$$

For aggregate social welfare  $W^C$  and  $W^D$ :

1. The welfare cost of rent extraction falls with centralization:

$$\log(1 - \rho^C) > \frac{1}{L} \sum_{l=1}^L \log(1 - \rho_l^D) \quad (\text{A72})$$

which can be written

$$f\left(\frac{1}{L} \sum_{l=1}^L \theta_l\right) > \frac{1}{L} \sum_{l=1}^L f\left(\left(1 - \frac{L-1}{L} \sum_{p=1}^P \xi_p \alpha_p\right) \theta_l\right) \quad (\text{A73})$$

for

$$f(x) \equiv \log x - \log\left(1 + 2\tilde{\delta}\phi x\right) \quad (\text{A74})$$

a strictly increasing and strictly concave function of  $x > 0$ :

$$f'(x) = \left[x \left(1 + \frac{2\tilde{\delta}}{2 - \tilde{\delta}} \phi x\right)\right]^{-1} > 0 \quad (\text{A75})$$

and

$$f''(x) = - \left[1 + \frac{4\tilde{\delta}}{2 - \tilde{\delta}} \phi x\right] \left[x \left(1 + \frac{2\tilde{\delta}}{2 - \tilde{\delta}} \phi x\right)\right]^{-2} < 0. \quad (\text{A76})$$

Thus

$$f\left(\frac{1}{L} \sum_{l=1}^L \theta_l\right) \geq \frac{1}{L} \sum_{l=1}^L f(\theta_l) \geq \frac{1}{L} \sum_{l=1}^L f\left(\left(1 - \frac{L-1}{L} \sum_{p=1}^P \xi_p \alpha_p\right) \theta_l\right). \quad (\text{A77})$$

The first inequality follows from convexity by Jensen's inequality and holds strictly if  $\theta_l$  is heterogeneous across regions. The second inequality follows from monotonicity and holds strictly if  $\xi_p > 0$  for some  $p$ .

2. Average efficiency in providing each public good  $p$  is weakly higher under centralization

$$\mathbb{E}\hat{\eta}_p^C \geq \frac{1}{L} \sum_{l=1}^L \mathbb{E}\hat{\eta}_{l,p}^D \iff \xi_p \geq 0, \quad (\text{A78})$$

with joint strict inequality.

3. Centralization achieves the optimal allocation of productive expenditures, while decentralization does not unless  $\xi_p$  is homogeneous across public goods.

Thus centralization cannot decrease welfare if preferences are homogeneous across regions ( $\alpha_p^l = \alpha_p$  for all  $l$ ), and it strictly increases welfare if information is heterogeneous ( $\theta_l \neq \theta_m$  for some  $l \neq m$ ) or there are externalities in the provision of public goods ( $\xi_p > 0$  for some  $p$ ).

## A.7. Proof of Proposition 4

Suppose that preferences are homogeneous, information is heterogeneous, and there are no externalities ( $\alpha_p^l = \alpha_p$  for all  $l$ ,  $\theta_l \neq \theta_m$  for some  $l \neq m$ , and  $\xi_p = 0$  for all  $p$ ). Under decentralization, welfare in region  $l$  is

$$\mathbb{E}u_l^D = \log b + \log(1 - \rho_l^D) + \sum_{p=1}^P \alpha_p (\mathbb{E}\hat{\eta}_{l,p}^D + \log \beta_{l,p}^D), \quad (\text{A79})$$

for

$$\rho_l^D = \left(1 + \frac{2\delta}{2-\delta} \phi \theta_l\right)^{-1}, \quad \mathbb{E}\hat{\eta}_{l,p}^D = \phi \sigma^2 \alpha_p \theta_l, \quad \text{and } \beta_{l,p}^D = \alpha_p. \quad (\text{A80})$$

Aggregate welfare is

$$W^D = \log b + \sum_{p=1}^P \alpha_p \log \alpha_p + \phi \sigma^2 \bar{\theta} \sum_{p=1}^P \alpha_p^2 + \frac{1}{L} \sum_{l=1}^L \log \frac{\frac{2\delta}{2-\delta} \phi \theta_l}{1 + \frac{2\delta}{2-\delta} \phi \theta_l} \quad (\text{A81})$$

Under centralization, welfare in region  $l$  is

$$\mathbb{E}u_l^C = \log(Lb) + \log(1 - \rho^C) + \sum_{p=1}^P \alpha_p \mathbb{E}\hat{\eta}_p^C + \sum_{p \in \mathcal{U}} \alpha_p \log \beta_p^C + \sum_{p \in \mathcal{D}} \alpha_p \log \beta_{l,p}^C \quad (\text{A82})$$

for

$$\rho^C = \left(1 + \frac{2\delta}{2-\delta} \phi \bar{\theta}\right)^{-1}, \quad \mathbb{E}\hat{\eta}_p^C = \phi \sigma^2 \alpha_p \bar{\theta}, \quad \beta_p^C = \frac{\alpha_p}{L} \text{ for } p \in \mathcal{U}, \quad \beta_{l,p}^C = \frac{\alpha_p \theta_l}{L \bar{\theta}} \text{ for } p \in \mathcal{D}. \quad (\text{A83})$$

Aggregate welfare is

$$W^C = \log b + \sum_{p=1}^P \alpha_p \log \alpha_p + \phi \sigma^2 \bar{\theta} \sum_{p=1}^P \alpha_p^2 + \log \frac{\frac{2\delta}{2-\delta} \phi \bar{\theta}}{1 + \frac{2\delta}{2-\delta} \phi \bar{\theta}} + (1 - \alpha_U) \left( \frac{1}{L} \sum_{l=1}^L \log \theta_l - \log \bar{\theta} \right). \quad (\text{A84})$$

Then aggregate welfare is greater under centralization if

$$\alpha_U \geq \frac{\log \left( 1 + \frac{2\delta}{2-\delta} \phi \bar{\theta} \right) - \frac{1}{L} \sum_{l=1}^L \log \left( 1 + \frac{2\delta}{2-\delta} \phi \theta_l \right)}{\log \bar{\theta} - \frac{1}{L} \sum_{l=1}^L \log \theta_l} \equiv \bar{\alpha}_U \in (0, 1). \quad (\text{A85})$$

Centralization is welfare-reducing if there is no uniformity constraint ( $\alpha_U = 0$ ).

The gains from centralization for region  $l$  are

$$\mathbb{E} (u_l^C - u_l^D) = \log \frac{1 + \frac{2\delta}{2-\delta} \phi \theta_l}{1 + \frac{2\delta}{2-\delta} \phi \bar{\theta}} - \alpha_U \log \frac{\theta_l}{\bar{\theta}} - (\theta_l - \bar{\theta}) \phi \sigma^2 \sum_{p=1}^P \alpha_p^2. \quad (\text{A86})$$

If there is no uniformity constraint ( $\alpha_U = 0$ ) then

$$\frac{\partial \mathbb{E} (u_l^C - u_l^D)}{\partial \theta_l} = \left( \frac{2 - \delta}{2\delta\phi} + \theta_l \right)^{-1} - \phi \sigma^2 \sum_{p=1}^P \alpha_p^2 \quad (\text{A87})$$

and

$$\frac{\partial^2 \mathbb{E} (u_l^C - u_l^D)}{\partial \theta_l^2} = - \left( \frac{2 - \delta}{2\delta\phi} + \theta_l \right)^{-2} < 0. \quad (\text{A88})$$

Therefore, if

$$\sigma^2 \leq \hat{\sigma}^2 \equiv \left[ \left( \frac{2 - \delta}{2\delta} + \phi \right) \sum_{p=1}^P \alpha_p^2 \right]^{-1} \quad (\text{A89})$$

then  $\mathbb{E} (u_l^C - u_l^D)$  is monotone increasing in  $\theta_l \in [0, 1]$ , so  $\theta_l < \theta_m$  implies  $\mathbb{E} (u_l^C - u_l^D) < \mathbb{E} (u_m^C - u_m^D)$ .

## A.8. Proof of Proposition 5

Suppose that preferences are homogeneous and there are no externalities:  $\alpha_p^l = \alpha_p$  for all  $l$  and  $\xi_p = 0$  for all  $p$ . In the proof of Proposition 4 above we established that centralization increases aggregate total welfare if and only if  $\alpha_U \geq \bar{\alpha}_U > 0$ .

Welfare in region  $l$  is

$$\mathbb{E} u_l^D = \log b + \log \frac{\frac{2\delta}{2-\delta} \phi \theta_l}{1 + \frac{2\delta}{2-\delta} \phi \theta_l} + \theta_l \phi \sigma^2 \sum_{p=1}^P \alpha_p^2 + \sum_{p=1}^P \alpha_p \log \alpha_p \quad (\text{A90})$$

under decentralization and

$$\mathbb{E}u_l^C = \log b + \log \frac{\frac{2\delta}{2-\delta}\phi\bar{\theta}}{1 + \frac{2\delta}{2-\delta}\phi\bar{\theta}} + \bar{\theta}\phi\sigma^2 \sum_{p=1}^P \alpha_p^2 + \sum_{p=1}^P \alpha_p \log \alpha_p + (1 - \alpha_U) \log \frac{\theta_l}{\bar{\theta}} \quad (\text{A91})$$

under centralization. Thus region  $l$  benefits from centralization if and only if

$$\begin{aligned} \log \left( 1 + \frac{2\delta}{2-\delta}\phi\theta_l \right) - \theta_l\phi\sigma^2 \sum_{p=1}^P \alpha_p^2 - \alpha_U \log \theta_l \\ \geq \log \left( 1 + \frac{2\delta}{2-\delta}\phi\bar{\theta} \right) - \bar{\theta}\phi\sigma^2 \sum_{p=1}^P \alpha_p^2 - \alpha_U \log \bar{\theta}. \end{aligned} \quad (\text{A92})$$

For ease of notation, let

$$\Delta \equiv \frac{2\delta}{2-\delta}\phi \text{ and } \Sigma \equiv \phi\sigma^2 \sum_{p=1}^P \alpha_p^2. \quad (\text{A93})$$

The gains from centralization are described by the function

$$g(\theta_l) \equiv \log(1 + \Delta\theta_l) - \Sigma\theta_l - \alpha_U \log \theta_l, \quad (\text{A94})$$

with derivative

$$g'(\theta_l) = \frac{\Delta}{1 + \Delta\theta_l} - \Sigma - \frac{\alpha_U}{\theta_l} = -\frac{\alpha_U - [(1 - \alpha_U)\Delta - \Sigma]\theta_l + \Sigma\Delta\theta_l^2}{(1 + \Delta\theta_l)\theta_l}. \quad (\text{A95})$$

The numerator is negative and increasing at  $\theta_l = 0$  if and only if  $\Delta > \Sigma$  and  $0 < \alpha_U < 1 - \Sigma/\Delta$ . The quadratic has a positive determinant if moreover  $\alpha_U < \left(1 - \sqrt{\Sigma/\Delta}\right)^2$ . Thus, if and only if

$$\Sigma < \Delta \text{ and } 0 < \alpha_U < \left(1 - \sqrt{\frac{\Sigma}{\Delta}}\right)^2, \quad (\text{A96})$$

the function  $g(\theta_l)$  has a unique interior minimum at

$$\check{\theta} = \frac{\Delta - \Sigma - \Delta\alpha_U - \sqrt{(\Delta - \Sigma)^2 - 2\Delta(\Delta + \Sigma)\alpha_U + \Delta^2\alpha_U^2}}{2\Delta\Sigma}. \quad (\text{A97})$$

If and only if

$$\Sigma \leq \frac{\Delta}{(1 + \Delta\bar{\theta})^2} \text{ and } \alpha_U = \alpha_U^* \equiv \left( \frac{\Delta}{1 + \Delta\bar{\theta}} - \Sigma \right) \bar{\theta} \quad (\text{A98})$$

the function  $g(\theta_l)$  has a unique interior minimum at  $\bar{\theta}$  and is monotone decreasing for



$\theta_l \in (0, \bar{\theta})$ . It is monotone increasing for  $\theta_l \in (\bar{\theta}, \hat{\theta})$  until it reaches a maximum

$$\hat{\theta} = \frac{1}{\Sigma (1 + \Delta \bar{\theta})} - \frac{1}{\Delta}. \quad (\text{A99})$$

A perfectly informed region ( $\theta_l = 1$ ) prefers centralization with uniformity  $\alpha_{\mathcal{U}}^*$  to decentralization if (but not only if)  $\hat{\theta} \geq 1$ , namely if

$$\Sigma \leq \frac{\Delta}{(1 + \Delta)(1 + \Delta \bar{\theta})}. \quad (\text{A100})$$

Thus,

$$\sigma^2 \leq \bar{\sigma}^2 \equiv \left[ \left( 1 + \frac{2\delta}{2-\delta} \phi \right) \left( 1 + \frac{2\delta}{2-\delta} \phi \bar{\theta} \right) \sum_{p=1}^P \alpha_p^2 \right]^{-1} \leq \hat{\sigma}^2 \quad (\text{A101})$$

is a sufficient but not necessary condition for centralization to be a Pareto-improvement over decentralization provided that  $\alpha_{\mathcal{U}} = \alpha_{\mathcal{U}}^*(\bar{\theta}, \sigma^2)$ , such that

$$\frac{\partial \alpha_{\mathcal{U}}^*}{\partial \bar{\theta}} = \frac{\Delta}{1 + \Delta \bar{\theta}} - \Sigma \geq 0 \text{ and } \frac{\partial \alpha_{\mathcal{U}}^*}{\partial \sigma^2} = -\bar{\theta} < 0. \quad (\text{A102})$$

Finally,  $\alpha_{\mathcal{U}}^*(\bar{\theta}, \sigma^2) \leq \rho^C$  implies  $\bar{\alpha}_{\mathcal{U}} < \alpha_{\mathcal{U}}^*(\bar{\theta}, \sigma^2) \leq \rho^C$  since a Pareto improvement implies an increase in aggregate welfare.

## A.9. Proof of Proposition 6

If there are no externalities and preferences are heterogeneous ( $\xi_p = 0$  for all  $p$ , while  $\alpha_p^l \neq \alpha_p^m$  for some  $l \neq m$  and  $p$ ), then under decentralization aggregate welfare is

$$W^D = \log b + \mathbb{E} \log \frac{\theta_l}{\frac{2-\delta}{2\delta\phi} + \theta_l} + \phi \sigma^2 \sum_{p=1}^P \mathbb{E} \left[ \theta_l (\alpha_p^l)^2 \right] + \sum_{p=1}^P \mathbb{E} (\alpha_p^l \log \alpha_p^l), \quad (\text{A103})$$

while under centralization it is

$$\begin{aligned} W^C = \log b + \log \frac{\mathbb{E} \theta_l}{\frac{2-\delta}{2\delta\phi} + \mathbb{E} \theta_l} + \phi \sigma^2 \sum_{p=1}^P \mathbb{E} (\theta_l \alpha_p^l) \mathbb{E} \alpha_p^l \\ + \sum_{p \in \mathcal{U}} \mathbb{E} \alpha_p^l \log \mathbb{E} (\theta_l \alpha_p^l) + \sum_{p \in \mathcal{D}} \mathbb{E} [\alpha_p^l \log (\theta_l \alpha_p^l)] - \log \mathbb{E} \theta_l. \end{aligned} \quad (\text{A104})$$

The welfare comparison can be decomposed into three elements.

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

$$\log (1 - \rho^C) = \log \frac{\mathbb{E} \theta_l}{\frac{2-\delta}{2\delta\phi} + \mathbb{E} \theta_l} > \mathbb{E} \log (1 - \rho_l^D) = \mathbb{E} \log \frac{\theta_l}{\frac{2-\delta}{2\delta\phi} + \theta_l}. \quad (\text{A105})$$

2. Centralization with heterogeneous preferences induces a misallocation of ability: since information  $\theta_l$  and preferences  $\alpha^l$  are independent,

$$\mathbb{E}(\theta_l \alpha_p^l) \mathbb{E} \alpha_p^l = \mathbb{E} \theta_l (\mathbb{E} \alpha_p^l)^2 < \mathbb{E} \theta_l \mathbb{E} \left[ (\alpha_p^l)^2 \right] = \mathbb{E} \left[ \theta_l (\alpha_p^l)^2 \right] \text{ for all } p. \quad (\text{A106})$$

3. Centralization with heterogeneous preferences and information induces a misallocation of resources: since information  $\theta_l$  and preferences  $\alpha^l$  are independent,

$$\begin{aligned} & \sum_{p \in \mathcal{U}} \mathbb{E} \alpha_p^l \log \mathbb{E}(\theta_l \alpha_p^l) + \sum_{p \in \mathcal{D}} \mathbb{E} [\alpha_p^l \log(\theta_l \alpha_p^l)] - \log \mathbb{E} \theta_l = \\ & \sum_{p=1}^P \mathbb{E}(\alpha_p^l \log \alpha_p^l) - \sum_{p \in \mathcal{U}} [\mathbb{E}(\alpha_p^l \log \alpha_p^l) - \mathbb{E} \alpha_p^l \log \mathbb{E} \alpha_p^l] - \sum_{p \in \mathcal{D}} \mathbb{E} \alpha_p^l (\log \mathbb{E} \theta_l - \mathbb{E} \log \theta_l) \\ & < \sum_{p=1}^P \mathbb{E}(\alpha_p^l \log \alpha_p^l). \end{aligned} \quad (\text{A107})$$

If no uniformity constraint is applied ( $\mathcal{U} = \emptyset$ ) 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:

$$\lim_{\sigma^2 \rightarrow 0} (W^D - W^C) = \log \left( \frac{2 - \delta}{2\delta\phi} + \mathbb{E} \theta_l \right) - \mathbb{E} \log \left( \frac{2 - \delta}{2\delta\phi} + \theta_l \right) \geq 0. \quad (\text{A108})$$

When the distribution of preferences is symmetric across goods, it is welfare-maximizing to apply the uniformity constraint either to all or to none. Uniformity is preferable if and only if

$$\log \mathbb{E} \theta_l - \mathbb{E} \log \theta_l \geq \frac{\mathbb{E}(\alpha_p^l \log \alpha_p^l) - \mathbb{E} \alpha_p^l \log \mathbb{E} \alpha_p^l}{\mathbb{E} \alpha_p^l} = \log P + P \mathbb{E}(\alpha_p^l \log \alpha_p^l) \quad (\text{A109})$$

Centralization with uniformity is preferable to decentralization ( $W^C \geq W^D$ ) if and only if the stricter condition

$$\begin{aligned} & \mathbb{E} \log \left( 1 + \frac{2 - \delta}{2\delta\phi\theta_l} \right) - \log \left( 1 + \frac{2 - \delta}{2\delta\phi\mathbb{E}\theta_l} \right) \\ & \geq \log P + P \mathbb{E}(\alpha_p^l \log \alpha_p^l) + \mathbb{E} \theta_l \phi \sigma^2 P \text{Var}(\alpha_p^l) \end{aligned} \quad (\text{A110})$$

hold.

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

$$f_L(\theta_l; \bar{\theta}) \equiv \log \left( 1 + \frac{2 - \delta}{2\delta\phi\theta_l} \right) - \log \left( 1 + \frac{2 - \delta}{2\delta\phi\bar{\theta}} \right) \quad (\text{A111})$$

such that

$$\frac{\partial^2 f_L}{\partial \theta_l^2} = \left[ \left( 1 + \frac{2\delta\phi}{2-\delta}\theta_l \right) \theta_l \right]^{-2} \left( 1 + \frac{4\delta\phi}{2-\delta}\theta_l \right) > 0. \quad (\text{A112})$$

Therefore, a mean-preserving spread of  $\theta_l$  increases the left-hand side of equation A110 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 A110 can be written as  $\mathbb{E}f_R(\alpha_p^l; \bar{\theta})$  for a function

$$f_R(\alpha_p^l; \bar{\theta}) \equiv P \left[ \alpha_p^l \log \alpha_p^l + \bar{\theta} \phi \sigma^2 (\alpha_p^l)^2 \right] - \frac{\bar{\theta} \phi \sigma^2}{P} + \log P \quad (\text{A113})$$

such that

$$\frac{\partial^2}{\partial (\alpha_p^l)^2} f_R(\alpha_p^l; \bar{\theta}) \equiv P \left( \frac{1}{\alpha_p^l} + 2\bar{\theta} \phi \sigma^2 \right) > 0. \quad (\text{A114})$$

Therefore, a mean-preserving spread of  $\alpha_p^l$  increases the right-hand side of equation A110 while leaving the left-hand side unchanged: decentralization is then more likely to be welfare-maximizing.

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

$$f(x; \mu, \nu) = \frac{1}{B(\mu\nu, (1-\mu)\nu)} x^{\mu\nu-1} (1-x)^{(1-\mu)\nu-1} \text{ for } x \in [0, 1]. \quad (\text{A115})$$

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

$$\frac{f(x; \mu, \nu_X)}{f(x; \mu, \nu_Y)} = \frac{B(\mu\nu_Y, (1-\mu)\nu_Y)}{B(\mu\nu_X, (1-\mu)\nu_X)} \left[ x^\mu (1-x)^{1-\mu} \right]^{\nu_X - \nu_Y}, \quad (\text{A116})$$

a log-concave function of  $x$ :

$$\frac{\partial^2}{\partial x^2} \log \frac{f(x; \mu, \nu_X)}{f(x; \mu, \nu_Y)} = -(\nu_X - \nu_Y) \left[ \frac{\mu}{x^2} + \frac{1-\mu}{(1-x)^2} \right] < 0 \quad (\text{A117})$$

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

In the limit as  $\iota \rightarrow 0$ , the distribution of  $\theta_l$  converges to a Bernoulli distribution with  $\Pr(\theta_l = 1) = \bar{\theta}$ . In the limit as  $\iota \rightarrow \infty$ ,  $\theta_l$  converges to the deterministic value  $\bar{\theta}$ . Thus the left-hand side of equation A110 is monotone decreasing in  $\iota$  from infinity to zero.

In the limit as  $\nu \rightarrow 0$ , the distribution of  $\alpha_p^l$  converges to a Bernoulli distribution with  $\Pr(\alpha_p^l = 1) = 1/P$ . In the limit as  $\nu \rightarrow \infty$ ,  $\alpha_p^l$  converges to the deterministic

value  $1/P$ . Thus the right-hand side of equation A110 is monotone decreasing in  $v$  from  $\log P + \bar{\theta}\phi\sigma^2(1 - 1/P)$  to zero

Thus, there exists a finite threshold  $\bar{v}(\iota, \sigma) > 0$  such that centralization with uniformity is preferable to decentralization if and only if  $v \geq \bar{v}$ . The threshold is increasing in  $\iota$ . It is increasing in  $\sigma^2$  because so is the right-hand side of equation A110.

## A.10. Federalism

Suppose there are no externalities:  $\xi_p = 0$  for all  $p$ .

From equations (A13) and (11), equilibrium rent extraction by a local politician in region  $l$  is

$$\rho_l^D = \left\{ 1 + \frac{2\delta\phi}{2-\delta} [(1-\chi_0)\alpha_0 + (1-\chi_1)(1-\alpha_0)]\theta_l \right\}^{-1}. \quad (\text{A118})$$

The politician's expected abilities are

$$\mathbb{E}\hat{\eta}_{l,0}^D = (1-\chi_0)\alpha_0\phi\sigma^2\theta_l \text{ and } \mathbb{E}\hat{\eta}_{l,l}^D = (1-\chi_1)(1-\alpha_0)\phi\sigma^2\theta_l, \quad (\text{A119})$$

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

$$\beta_{l,0}^D = \frac{(1-\chi_0)\alpha_0}{(1-\chi_0)\alpha_0 + (1-\chi_1)(1-\alpha_0)} \quad (\text{A120})$$

and

$$\beta_{l,l}^D = \frac{(1-\chi_1)(1-\alpha_0)}{(1-\chi_0)\alpha_0 + (1-\chi_1)(1-\alpha_0)}, \beta_{l,m}^D = 0 \text{ for all } m \neq l \quad (\text{A121})$$

for the allocation of his budget  $b^D = b - b^C/L$ .

Equilibrium rent extraction by a central politician is

$$\rho^C = \left\{ 1 + \frac{2\delta\phi}{2-\delta} [\chi_0\alpha_0 + \chi_1(1-\alpha_0)]\bar{\theta} \right\}^{-1}. \quad (\text{A122})$$

His expected abilities are

$$\mathbb{E}\hat{\eta}_0^C = \chi_0\alpha_0\phi\sigma^2\bar{\theta} \text{ and } \mathbb{E}\hat{\eta}_l^C = \chi_1(1-\alpha_0)\phi\sigma^2\frac{\theta_l}{L} \text{ for } l = 1, 2, \dots, L. \quad (\text{A123})$$

Given his budget  $b^C$ , if he is entrusted with providing the homogeneously desired good he chooses a budget share

$$\beta_0^C = \frac{1}{L} \frac{\chi_0\alpha_0}{\chi_0\alpha_0 + \chi_1(1-\alpha_0)} \text{ if } 0 \in \mathcal{U}, \quad (\text{A124})$$

or

$$\beta_{l,0}^C = \frac{1}{L} \frac{\chi_0\alpha_0}{\chi_0\alpha_0 + \chi_1(1-\alpha_0)} \frac{\theta_l}{\bar{\theta}} \text{ if } 0 \in \mathcal{D}. \quad (\text{A125})$$

If the central politician is entrusted with providing the idiosyncratically preferred good,

he sets a budget share

$$\beta_l^C = \frac{1}{L} \frac{\chi_1 (1 - \alpha_0)}{\chi_0 \alpha_0 + \chi_1 (1 - \alpha_0)} \frac{\theta_l}{\bar{\theta} L} \text{ if } l \in \mathcal{U}, \quad (\text{A126})$$

or

$$\beta_{l,l}^C = \frac{1}{L} \frac{\chi_1 (1 - \alpha_0)}{\chi_0 \alpha_0 + \chi_1 (1 - \alpha_0)} \frac{\theta_l}{\bar{\theta}} \text{ and } \beta_{m,l}^C = 0 \text{ for all } m \neq l \text{ if } l \in \mathcal{D}. \quad (\text{A127})$$

Welfare in region  $l$  can be decomposed into four components

$$\mathbb{E}u^l = u_b^l + u_\beta^l + u_\rho^l + \mathbb{E}u_\eta^l. \quad (\text{A128})$$

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

$$u_b^l = [(1 - \chi_0) \alpha_0 + (1 - \chi_1) (1 - \alpha_0)] \log (b - b^C/L) + [\chi_0 \alpha_0 + \chi_1 (1 - \alpha_0)] \log b^C. \quad (\text{A129})$$

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

$$u_\beta^l = (1 - \chi_0) \alpha_0 \log \beta_{l,0}^D + (1 - \chi_1) (1 - \alpha_0) \log \beta_{l,l}^D + \chi_0 \alpha_0 \log \beta_0^C + \chi_1 (1 - \alpha_0) \log \beta_{l,l}^C. \quad (\text{A130})$$

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

$$u_\rho^l = [(1 - \chi_0) \alpha_0 + (1 - \chi_1) (1 - \alpha_0)] \log (1 - \rho_l^D) + [\chi_0 \alpha_0 + \chi_1 (1 - \alpha_0)] \log (1 - \rho^C). \quad (\text{A131})$$

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

$$u_\eta^l = (1 - \chi_0) \alpha_0 \mathbb{E}\hat{\eta}_{l,0}^D + (1 - \chi_1) (1 - \alpha_0) \mathbb{E}\hat{\eta}_{l,l}^D + \chi_0 \alpha_0 \mathbb{E}\hat{\eta}_0^C + \chi_1 (1 - \alpha_0) \mathbb{E}\hat{\eta}_l^C. \quad (\text{A132})$$

## A.11. Proof of Lemma 1

The allocation of the budget between the two levels of government only affects welfare through the term  $\mathbb{E}u_b^l$ . Every region agrees that the optimal allocation solves

$$\max_{b^C} \{ [(1 - \chi_0) \alpha_0 + (1 - \chi_1) (1 - \alpha_0)] \log (b - b^C/L) + [\chi_0 \alpha_0 + \chi_1 (1 - \alpha_0)] \log b^C \} \quad (\text{A133})$$

and therefore is

$$b^C = [\chi_0 \alpha_0 + \chi_1 (1 - \alpha_0)] bL \quad (\text{A134})$$

and

$$b^D = [(1 - \chi_0) \alpha_0 + (1 - \chi_1) (1 - \alpha_0)] b. \quad (\text{A135})$$

## A.12. Proof of Proposition 7

For local politicians,

$$\frac{\partial \rho_l^D}{\partial \chi_0} = \frac{2\delta\phi}{2-\delta} \alpha_0 \theta_l (\rho_l^D)^2 > 0 \text{ and } \frac{\partial \rho_l^D}{\partial \chi} = \frac{2\delta\phi}{2-\delta} (1-\alpha_0) \theta_l (\rho_l^D)^2 > 0. \quad (\text{A136})$$

For central politicians

$$\frac{\partial \rho^C}{\partial \chi} = -\frac{2\delta\phi}{2-\delta} \alpha_0 \bar{\theta} (\rho^C)^2 < 0 \text{ and } \frac{\partial \rho^C}{\partial \chi_1} = -\frac{2\delta\phi}{2-\delta} (1-\alpha_0) \bar{\theta} (\rho^C)^2 < 0. \quad (\text{A137})$$

In a fully decentralized state, aggregate rent extraction is

$$\bar{\rho}^D = \frac{1}{L} \sum_{l=1}^L \left( 1 + \frac{2\delta\phi}{2-\delta} \theta_l \right)^{-1}. \quad (\text{A138})$$

In a fully centralized state, it is

$$\bar{\rho}^C = \left( 1 + \frac{2\delta\phi}{2-\delta} \bar{\theta} \right)^{-1} \leq \bar{\rho}^D \quad (\text{A139})$$

as in Proposition 2.

In a federal state, aggregate rent extraction is

$$\begin{aligned} \bar{\rho}^F &= \frac{b_C}{bL} \left( 1 + \frac{2\delta\phi}{2-\delta} \alpha_0 \bar{\theta} \right)^{-1} + \left( 1 - \frac{b_C}{bL} \right) \frac{1}{L} \sum_{l=1}^L \left[ 1 + \frac{2\delta\phi}{2-\delta} (1-\alpha_0) \theta_l \right]^{-1} \\ &\geq \frac{b_C}{bL} \bar{\rho}^C + \left( 1 - \frac{b_C}{bL} \right) \bar{\rho}^D \geq \bar{\rho}^C. \end{aligned} \quad (\text{A140})$$

With the optimal budget allocation across layers of government, aggregate rent extraction is

$$\bar{\rho}^F = \alpha_0 \left( 1 + \frac{2\delta\phi}{2-\delta} \alpha_0 \bar{\theta} \right)^{-1} + \frac{1-\alpha_0}{L} \sum_{l=1}^L \left[ 1 + \frac{2\delta\phi}{2-\delta} (1-\alpha_0) \theta_l \right]^{-1}. \quad (\text{A141})$$

such that  $\bar{\rho}^F \leq \bar{\rho}^D$  if and only if

$$\left( 1 + \frac{2\delta\phi}{2-\delta} \alpha_0 \bar{\theta} \right)^{-1} \leq \frac{1}{L} \sum_{l=1}^L \left\{ \left( 1 + \frac{2\delta\phi}{2-\delta} \theta_l \right) \left[ 1 + \frac{2\delta\phi}{2-\delta} (1-\alpha_0) \theta_l \right] \right\}^{-1}, \quad (\text{A142})$$

namely if and only if  $\alpha_0 \geq \bar{\alpha}_\rho \in (0, 1)$  such that

$$\left( 1 + \frac{2\delta\phi}{2-\delta} \bar{\alpha}_\rho \bar{\theta} \right)^{-1} = \frac{1}{L} \sum_{l=1}^L \left\{ \left( 1 + \frac{2\delta\phi}{2-\delta} \theta_l \right) \left[ 1 + \frac{2\delta\phi}{2-\delta} (1-\bar{\alpha}_\rho) \theta_l \right] \right\}^{-1}. \quad (\text{A143})$$

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

$$f_\rho(\theta_l, \alpha; \bar{\theta}) \equiv \{(1 + \Delta\theta_l)[1 + \Delta(1 - \bar{\alpha}_\rho)\theta_l]\}^{-1} - (1 + \Delta\bar{\alpha}_\rho\bar{\theta})^{-1}, \quad (\text{A144})$$

such that

$$\frac{\partial^2 f_\rho}{\partial \theta_l^2} = 2\Delta^2 \frac{\bar{\alpha}_\rho^2 + 3(1 - \bar{\alpha}_\rho) + 3\Delta(2 - \bar{\alpha}_\rho)(1 - \bar{\alpha}_\rho)\theta_l + 3\Delta^2(1 - \bar{\alpha}_\rho)^2\theta_l^2}{\{(1 + \Delta\theta_l)[1 + \Delta(1 - \bar{\alpha}_\rho)\theta_l]\}^3} > 0. \quad (\text{A145})$$

Therefore, a mean-preserving spread of  $\theta_l$  increases  $\mathbb{E}f_\rho(\theta_l, \bar{\alpha}_\rho; \bar{\theta}) = 0$ . At the same time,  $\partial \mathbb{E}f_\rho(\theta_l, \bar{\alpha}_\rho; \bar{\theta}) / \partial \alpha > 0$ . Hence  $\partial \bar{\alpha}_\rho / \partial \iota > 0$ .

### A.13. Proof of Lemma 2

Imposing the uniformity constraint on centralized provision of the idiosyncratically preferred public good reduces welfare in every region because the utility-generating variety  $l$  is then provided in region  $l$  in the amount  $\beta_{l,l}^C = \beta_{l,l}^C / L$ .

Let  $\omega$  be an indicator variable denoting if the centralized provision of the homogeneously desired good is subject to the uniformity constraint. Then with the optimal budget rule  $b^C$

$$u_b^l + u_\beta^l = \log b + \alpha_0 \log \alpha_0 + (1 - \alpha_0) \log(1 - \alpha_0) + [(1 - \omega)\chi_0\alpha_0 + \chi_1(1 - \alpha_0)](\log \theta_l - \log \bar{\theta}). \quad (\text{A146})$$

Uniformity constraints do not appear in any other part of the welfare function, so

$$u^l|_{\omega=1} > u^l|_{\omega=0} \Leftrightarrow \log \theta_l < \log \bar{\theta}. \quad (\text{A147})$$

This uniformity constraint increases aggregate social welfare, as in Proposition 3.

### A.14. Proof of Proposition 8 and Corollary 3

Given optimal central-government budget and the welfare-maximizing uniformity constraints,

$$u_b^l + u_\beta^l = \log b + \alpha_0 \log \alpha_0 + (1 - \alpha_0) \log(1 - \alpha_0) + \chi_1(1 - \alpha_0)(\log \theta_l - \log \bar{\theta}). \quad (\text{A148})$$

Moreover, given the equilibrium levels of rent extraction,

$$\begin{aligned} u_b^l + u_\beta^l + u_\rho^l &= \log b + \alpha_0 \log \alpha_0 + (1 - \alpha_0) \log(1 - \alpha_0) + \log \Delta \\ &\quad + \alpha_0 [(1 - \chi_0) \log \theta_l + \chi_0 \log \bar{\theta}] + (1 - \alpha_0) \log \theta_l \\ &\quad + [(1 - \chi_0)\alpha_0 + (1 - \chi_1)(1 - \alpha_0)] \log \frac{(1 - \chi_0)\alpha_0 + (1 - \chi_1)(1 - \alpha_0)}{1 + \Delta [(1 - \chi_0)\alpha_0 + (1 - \chi_1)(1 - \alpha_0)] \theta_l} \\ &\quad + [\chi_0\alpha_0 + \chi_1(1 - \alpha_0)] \log \frac{\chi_0\alpha_0 + \chi_1(1 - \alpha_0)}{1 + \Delta [\chi_0\alpha_0 + \chi_1(1 - \alpha_0)] \bar{\theta}}. \end{aligned} \quad (\text{A149})$$

Given the equilibrium skill of incumbent politicians,

$$\mathbb{E}u_\eta^l = \phi\sigma^2 \left\{ \alpha_0^2 [(1 - \chi_0)\theta_l + \chi_0\bar{\theta}] + (1 - \alpha_0)^2 \left(1 - \chi_1 + \frac{\chi_1}{L}\right) \theta_l \right\}. \quad (\text{A150})$$

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

$$\begin{aligned} W = & \log b + \alpha_0 \log \alpha_0 + (1 - \alpha_0) \log (1 - \alpha_0) + \log \Delta \\ & + \alpha_0 [(1 - \chi_0) \mathbb{E} \log \theta_l + \chi_0 \log \mathbb{E} \theta_l] + (1 - \alpha_0) \mathbb{E} \log \theta_l \\ & + [(1 - \chi_0) \alpha_0 + (1 - \chi_1) (1 - \alpha_0)] \mathbb{E} \log \frac{(1 - \chi_0) \alpha_0 + (1 - \chi_1) (1 - \alpha_0)}{1 + \Delta [(1 - \chi_0) \alpha_0 + (1 - \chi_1) (1 - \alpha_0)]} \theta_l \\ & + [\chi_0 \alpha_0 + \chi_1 (1 - \alpha_0)] \log \frac{\chi_0 \alpha_0 + \chi_1 (1 - \alpha_0)}{1 + \Delta [\chi_0 \alpha_0 + \chi_1 (1 - \alpha_0)]} \mathbb{E} \theta_l \\ & + \phi\sigma^2 [\alpha_0^2 + (1 - \alpha_0)^2 (1 - \chi_1)] \mathbb{E} \theta_l. \end{aligned} \quad (\text{A151})$$

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

$$\begin{aligned} W_D = & \log b + \alpha_0 \log \alpha_0 + (1 - \alpha_0) \log (1 - \alpha_0) + \phi\sigma^2 [\alpha_0^2 + (1 - \alpha_0)^2] \mathbb{E} \theta_l \\ & + \mathbb{E} \log \frac{\Delta \theta_l}{1 + \Delta \theta_l}. \end{aligned} \quad (\text{A152})$$

Under a federal structure ( $\chi_0 = 1$  and  $\chi_1 = 0$ ) social welfare equals

$$\begin{aligned} W_F = & \log b + \alpha_0 \log \alpha_0 + (1 - \alpha_0) \log (1 - \alpha_0) + \phi\sigma^2 [\alpha_0^2 + (1 - \alpha_0)^2] \mathbb{E} \theta_l \\ & + \alpha_0 \log \frac{\Delta \alpha_0 \mathbb{E} \theta_l}{1 + \Delta \alpha_0 \mathbb{E} \theta_l} + (1 - \alpha_0) \mathbb{E} \log \frac{\Delta (1 - \alpha_0) \theta_l}{1 + \Delta (1 - \alpha_0) \theta_l}. \end{aligned} \quad (\text{A153})$$

Under full centralization ( $\chi_0 = \chi_1 = 1$ ) social welfare equals

$$\begin{aligned} W_C = & \log b + \alpha_0 \log \alpha_0 + (1 - \alpha_0) \log (1 - \alpha_0) + \phi\sigma^2 [\alpha_0^2 + (1 - \alpha_0)^2] \mathbb{E} \theta_l \\ & - \phi\sigma^2 (1 - \alpha_0)^2 \mathbb{E} \theta_l + \log \frac{\Delta \mathbb{E} \theta_l}{1 + \Delta \mathbb{E} \theta_l} + (1 - \alpha_0) (\mathbb{E} \log \theta_l - \log \mathbb{E} \theta_l). \end{aligned} \quad (\text{A154})$$

Since the first line of all three expression is identical, for ease of notation we can rescale social welfare measure by an additive constant and write instead

$$W_D = -\mathbb{E} \log \left( 1 + \frac{1}{\Delta \theta_l} \right), \quad (\text{A155})$$

$$W_F = -(1 - \alpha_0) \mathbb{E} \log \left[ 1 + \frac{1}{\Delta (1 - \alpha_0) \theta_l} \right] - \alpha_0 \log \left( 1 + \frac{1}{\Delta \alpha_0 \mathbb{E} \theta_l} \right), \quad (\text{A156})$$



and

$$W_C = -(1 - \alpha_0) (\log \mathbb{E}\theta_l - \mathbb{E} \log \theta_l) - (1 - \alpha_0)^2 \phi \sigma^2 \mathbb{E}\theta_l - \log \left( 1 + \frac{1}{\Delta \mathbb{E}\theta_l} \right). \quad (\text{A157})$$

Welfare under a federal structure has limits

$$\lim_{\alpha_0 \rightarrow 0} W_F = -\mathbb{E} \log \left( 1 + \frac{1}{\Delta \theta_l} \right) < \lim_{\alpha_0 \rightarrow 1} W_F = -\log \left( 1 + \frac{1}{\Delta \mathbb{E}\theta_l} \right). \quad (\text{A158})$$

Its first derivative is

$$\begin{aligned} \frac{\partial W_F}{\partial \alpha_0} = \mathbb{E} \left\{ \log \left[ 1 + \frac{1}{\Delta (1 - \alpha_0) \theta_l} \right] - \frac{1}{1 + \Delta (1 - \alpha_0) \theta_l} \right\} \\ - \left[ \log \left( 1 + \frac{1}{\Delta \alpha_0 \mathbb{E}\theta_l} \right) - \frac{1}{1 + \Delta \alpha_0 \mathbb{E}\theta_l} \right] \end{aligned} \quad (\text{A159})$$

with limits

$$\lim_{\alpha_0 \rightarrow 0} \frac{\partial W_F}{\partial \alpha_0} = -\infty \text{ and } \lim_{\alpha_0 \rightarrow 1} \frac{\partial W_F}{\partial \alpha_0} = +\infty. \quad (\text{A160})$$

Its second derivative is

$$\frac{\partial^2 W_F}{\partial \alpha_0^2} = \frac{1}{1 - \alpha_0} \mathbb{E} [1 + \Delta (1 - \alpha_0) \theta_l]^{-2} + \frac{1}{\alpha_0} (1 + \Delta \alpha_0 \mathbb{E}\theta_l)^{-2} > 0. \quad (\text{A161})$$

Welfare under full centralization has limits

$$\begin{aligned} \lim_{\alpha_0 \rightarrow 0} W_C = -(\log \mathbb{E}\theta_l - \mathbb{E} \log \theta_l) - \phi \sigma^2 \mathbb{E}\theta_l - \log \left( 1 + \frac{1}{\Delta \mathbb{E}\theta_l} \right) \\ < \lim_{\alpha_0 \rightarrow 1} W_C = -\log \left( 1 + \frac{1}{\Delta \mathbb{E}\theta_l} \right). \end{aligned} \quad (\text{A162})$$

Its first derivative is

$$\frac{\partial W_C}{\partial \alpha_0} = \log \mathbb{E}\theta_l - \mathbb{E} \log \theta_l + 2(1 - \alpha_0) \phi \sigma^2 \mathbb{E}\theta_l > 0 \quad (\text{A163})$$

with limits

$$\lim_{\alpha_0 \rightarrow 0} \frac{\partial W_C}{\partial \alpha_0} = \log \mathbb{E}\theta_l - \mathbb{E} \log \theta_l + 2\phi \sigma^2 \mathbb{E}\theta_l > \lim_{\alpha_0 \rightarrow 1} \frac{\partial W_C}{\partial \alpha_0} = \log \mathbb{E}\theta_l - \mathbb{E} \log \theta_l. \quad (\text{A164})$$

Its second derivative is

$$\frac{\partial^2 W_C}{\partial \alpha_0^2} = -2\phi \sigma^2 \mathbb{E}\theta_l < 0 \quad (\text{A165})$$

It is never welfare-maximizing to assign powers so that the uniformly preferred public good is decentralized and all idiosyncratically preferred public goods are centralized, because

this would yield welfare

$$W_{-F} = -(1 - \alpha_0) (\log \mathbb{E}\theta_l - \mathbb{E} \log \theta_l) - (1 - \alpha_0)^2 \phi \sigma^2 \mathbb{E}\theta_l \\ - \mathbb{E}\alpha_0 \log \left( 1 + \frac{1}{\Delta \alpha_0 \theta_l} \right) - (1 - \alpha_0) \log \left[ 1 + \frac{1}{\Delta (1 - \alpha_0) \mathbb{E}\theta_l} \right] < W_C. \quad (\text{A166})$$

There is a threshold  $\bar{\alpha}_{D \sim C} \in (0, 1)$  defined by  $W_C(\bar{\alpha}_{D \sim C}) = W_D$  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  $W_F(\bar{\alpha}_{D \sim F}) = W_D$  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  $\alpha_0$ ,  $W_F(\alpha_0)$  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  $\mathbb{E}f_{D \sim F}(\theta_l, \bar{\alpha}_{D \sim F}; \bar{\theta}) = 0$ , where

$$f_{D \sim F}(\theta_l, \alpha; \bar{\theta}) \equiv \log \left( 1 + \frac{1}{\Delta \theta_l} \right) - (1 - \alpha) \log \left[ 1 + \frac{1}{\Delta (1 - \alpha) \theta_l} \right] \\ - \alpha \log \left( 1 + \frac{1}{\Delta \alpha \bar{\theta}} \right), \quad (\text{A167})$$

such that

$$\frac{\partial^2 f_{D \sim F}}{\partial \theta_l^2} = \alpha \frac{1 + 2(2 - \alpha) \Delta \theta_l + 3(1 - \alpha) (\Delta \theta_l)^2}{\{\theta_l (1 + \Delta \theta_l) [1 + (1 - \alpha) \Delta \theta_l]\}^2} > 0. \quad (\text{A168})$$

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

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

$$f_{F \sim C}(\theta_l, \bar{\alpha}_{F \sim C}; \bar{\theta}, \sigma) \equiv (1 - \alpha) \log \left[ \frac{1}{\Delta (1 - \alpha)} + \theta_l \right] \\ + \alpha \log \left( 1 + \frac{1}{\Delta \alpha \bar{\theta}} \right) - (1 - \alpha) \log \bar{\theta} - (1 - \alpha)^2 \phi \sigma^2 \bar{\theta} - \log \left( 1 + \frac{1}{\Delta \bar{\theta}} \right), \quad (\text{A169})$$

such that

$$\frac{\partial^2 f_{F \sim C}}{\partial \theta_l^2} = -\frac{\Delta^2 (1 - \alpha)^3}{[1 + \Delta (1 - \alpha) \theta_l]^2} < 0. \quad (\text{A170})$$

Therefore, a mean-preserving spread of  $\theta_l$  decreases  $\mathbb{E}f_{F \sim C}(\theta_l, \bar{\alpha}_{F \sim C}; \bar{\theta}, \sigma)$ . At the same time,  $\partial \mathbb{E}f_{F \sim C}(\theta_l, \bar{\alpha}_{F \sim C}; \bar{\theta}, \sigma) / \partial \alpha > 0$  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  $\iota \rightarrow \infty$  information becomes perfectly homogeneous ( $\theta_l = \bar{\theta}$ ), so

$$\lim_{\iota \rightarrow \infty} W_D = -\log \left( 1 + \frac{1}{\Delta \bar{\theta}} \right), \quad (\text{A171})$$

while

$$\lim_{\iota \rightarrow \infty} W_F = -(1 - \alpha_0) \log \left[ 1 + \frac{1}{\Delta (1 - \alpha_0) \bar{\theta}} \right] - \alpha_0 \log \left( 1 + \frac{1}{\Delta \alpha_0 \bar{\theta}} \right), \quad (\text{A172})$$

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

$$\lim_{\iota \rightarrow \infty} W_C = -(1 - \alpha_0)^2 \phi \sigma^2 \bar{\theta} - \log \left( 1 + \frac{1}{\Delta \bar{\theta}} \right). \quad (\text{A173})$$

Thus

$$\lim_{\iota \rightarrow \infty} \bar{\alpha}_{D \sim C} = \lim_{\iota \rightarrow \infty} \bar{\alpha}_{D \sim F} = 1 > \lim_{\iota \rightarrow \infty} \bar{\alpha}_{F \sim C}. \quad (\text{A174})$$

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

$$\lim_{\iota \rightarrow 0} \frac{W_F}{W_D} = \lim_{\iota \rightarrow 0} \frac{W_C}{W_D} = 1 - \alpha < \lim_{\iota \rightarrow 0} \frac{W_C}{W_F} = 1. \quad (\text{A175})$$

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

$$\lim_{\iota \rightarrow 0} \bar{\alpha}_{D \sim F} = \lim_{\iota \rightarrow 0} \bar{\alpha}_{D \sim C} = 0 < \lim_{\iota \rightarrow \infty} \bar{\alpha}_{F \sim C}. \quad (\text{A176})$$

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  $\sigma$  because an increase in  $\sigma$  shifts down  $W_C$  while leaving  $W_D$  and  $W_F$  unaffected. Hence,  $\partial \bar{\alpha}_{F \sim C} / \partial \sigma > 0$  and  $\partial \bar{\alpha}_{D \sim C} / \partial \sigma > 0$ , while  $\partial \bar{\alpha}_{D \sim F} / \partial \sigma = 0$ .

## A.15. Proof of Corollary 4

Region  $l$ 's welfare depends on its residents' information according to

$$\frac{\partial \mathbb{E}u^l}{\partial \theta_l} = \frac{\alpha_0(1 - \chi_0) + (1 - \alpha_0)}{\theta_l} - \frac{\Delta [(1 - \chi_0)\alpha_0 + (1 - \chi_1)(1 - \alpha_0)]^2}{1 + \Delta [(1 - \chi_0)\alpha_0 + (1 - \chi_1)(1 - \alpha_0)]\theta_l} + \phi\sigma^2 \left[ \alpha_0^2(1 - \chi_0) + (1 - \alpha_0)^2 \left( 1 - \chi_1 + \frac{\chi_1}{L} \right) \right], \quad (\text{A177})$$

such that

$$\frac{\partial^2 \mathbb{E}u^l}{\partial \theta_l \partial \chi_0} = -\frac{\alpha_0}{\theta_l} (1 + \Xi\theta_l)^{-2} - \phi\sigma^2 \alpha_0^2 < 0 \quad (\text{A178})$$

and

$$\frac{\partial^2 \mathbb{E}u^l}{\partial \theta_l \partial \chi_1} = \frac{1 - \alpha_0}{\theta_l} [1 - (1 + \Xi\theta_l)^{-2}] - \phi\sigma^2 (1 - \alpha_0)^2 \left( 1 - \frac{1}{L} \right), \quad (\text{A179})$$

where for ease of notation

$$\Xi = \Delta [(1 - \chi_0)\alpha_0 + (1 - \chi_1)(1 - \alpha_0)]. \quad (\text{A180})$$

Moreover,

$$\frac{\partial^3 \mathbb{E}u^l}{\partial \theta_l^2 \partial \chi_1} = -(1 - \alpha_0) \Xi^2 \frac{3 + \Xi\theta_l}{(1 + \Xi\theta_l)^3} < 0 \quad (\text{A181})$$

and therefore

$$\sigma^2 \leq \frac{1 - (1 + \Xi)^{-2}}{\phi(1 - \alpha_0)} \frac{L}{L - 1} \Rightarrow \frac{\partial^2 \mathbb{E}u^l}{\partial \theta_l \partial \chi_1} > 0 \text{ for all } \theta_l < 1. \quad (\text{A182})$$

## A.16. Proof of Proposition 9

Recall that the expected ability of a local politician under decentralization is

$$\mathbb{E}\hat{\eta}_{l,p}^D = \phi\sigma^2 \left( 1 - \frac{L-1}{L}\xi_p \right) \alpha_p^l \theta_l, \quad (\text{A183})$$

while the expected ability of a central politician is

$$\mathbb{E}\hat{\eta}_p^C = \frac{\phi\sigma^2}{L} \sum_{l=1}^L \theta_l \alpha_p^l. \quad (\text{A184})$$

Thus

$$\mathbb{E}\hat{\eta}_p^C > \frac{1}{L} \sum_{l=1}^L \mathbb{E}\hat{\eta}_{l,p}^D \iff \xi_p > 0, \quad (\text{A185})$$

with

$$\frac{\partial}{\partial \xi_p} \left( \mathbb{E}\hat{\eta}_p^C - \frac{1}{L} \sum_{l=1}^L \mathbb{E}\hat{\eta}_{l,p}^D \right) = \phi\sigma^2 \frac{L-1}{L^2} \sum_{l=1}^L \alpha_p^l \theta_l > 0 \quad (\text{A186})$$

and

$$\frac{\partial^2}{\partial \xi_p \partial \theta_l} \left( \mathbb{E} \hat{\eta}_p^C - \frac{1}{L} \sum_{l=1}^L \mathbb{E} \hat{\eta}_{l,p}^D \right) = \phi \sigma^2 \frac{L-1}{L^2} \alpha_p^l > 0. \quad (\text{A187})$$

Proposition 2 established that rent extraction is lower under centralization than decentralization. Moreover

$$\frac{\partial}{\partial \xi_p} \left( \frac{1}{L} \sum_{l=1}^L \rho_l^D - \rho^C \right) = \frac{1}{L} \sum_{l=1}^L \frac{\partial \rho_l^D}{\partial \xi_p} = \frac{2\delta}{2-\delta} \phi \frac{L-1}{L^2} \sum_{l=1}^L \alpha_p^l \theta_l (\rho_l^D)^2 > 0 \quad (\text{A188})$$

and

$$\frac{\partial^2}{\partial \xi_p \partial \theta_l} \left( \frac{1}{L} \sum_{l=1}^L \rho_l^D - \rho^C \right) = \frac{1}{L} \sum_{l=1}^L \frac{\partial \rho_l^D}{\partial \xi_p} = \frac{2\delta}{2-\delta} \phi \frac{L-1}{L^2} \alpha_p^l (\rho_l^D)^2 > 0. \quad (\text{A189})$$

### A.17. Proof of Proposition 10

Suppose there is a set  $\Xi$  of public goods that generate externalities  $\xi$ , while the other generate none. Let  $\alpha_\Xi = \sum_{p \in \Xi} \alpha_p$  be the aggregate welfare share of externality-inducing public goods.

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

$$\beta_{l,p}^C = \frac{\alpha_p}{L} \quad (\text{A190})$$

with a uniformity constraint or

$$\beta_l^C = \frac{\alpha_p}{L} \left[ \xi + (1-\xi) \frac{\theta_l}{\theta} \right] \quad (\text{A191})$$

without. Aggregating across regions, the share of the good in the central budget is

$$\beta_p^C = \sum_{l=1}^L \beta_l^C = \alpha_p \quad (\text{A192})$$

regardless of the presence or absence of a uniformity constraint.

Under decentralization, the share of an externality-inducing good in each region  $l$  is

$$\beta_{l,p}^D = \alpha_p \frac{1 - \frac{L-1}{L} \xi}{1 - \frac{L-1}{L} \xi \alpha_\Xi} < \alpha_p \text{ for all } \alpha_\Xi < 1, \quad (\text{A193})$$

such that

$$\frac{\partial \beta_l^D}{\partial \xi} = - \frac{\alpha_p (1 - \alpha_\Xi) \frac{L-1}{L}}{\left(1 - \frac{L-1}{L} \xi \alpha_\Xi\right)^2} < 0. \quad (\text{A194})$$

## References

- [1] 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
- [2] Alesina, Alberto, Ignazio Angeloni, and Federico Etro. 2005. “International Unions.” *American Economic Review* 95(3): 602–15.
- [3] Alesina, Alberto, Ignazio Angeloni, and Ludger Schuknecht. 2005. “What Does the European Union Do?” *Public Choice* 123(3): 275–319.
- [4] Alesina, Alberto, and Enrico Spolaore. 1997. “On the Number and Size of Nations.” *Quarterly Journal of Economics* 112(4): 1027–56.
- [5] Alesina, Alberto, and Guido Tabellini. 2008. “Bureaucrats or Politicians? Part II: Multiple Policy Tasks.” *Journal of Public Economics* 92: 426–447.
- [6] Bardhan, Pranab, and Dilip Mookherjee. 2000. “Capture and Governance at Local and National Levels.” *American Economic Review* 90(2): 135–9.
- [7] ——. 2006a. “Decentralisation and Accountability in Infrastructure Delivery in Developing Countries.” *Economic Journal* 116(1): 101–27.
- [8] ——. 2006b. “Decentralization, Corruption, and Government Accountability.” In *International Handbook on the Economics of Corruption*, edited by Susan Rose-Ackerman, 161–88. Cheltenham: Edward Elgar.
- [9] 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.
- [10] Besley, Timothy, and Anne Case. 1995. “Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition.” *American Economic Review* 85(1): 25–45.
- [11] 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.
- [12] Besley, Timothy, and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton, NJ: Princeton University Press.
- [13] Besley, Timothy, and Michael Smart. 2007. “Fiscal Restraints and Voter Welfare.” *Journal of Public Economics* 91: 755–73.
- [14] Blanchard, Olivier, and Andrei Shleifer. 2001. “Federalism With and Without Political Centralization: China Versus Russia.” *IMF Staff Papers* 48(4): 171–9.
- [15] Bulte, Erwin, John List, and Mark Strazicich. 2007. “Regulatory Federalism and the Distribution of Air Pollutant Emissions”, *Journal of Regional Science*, 47(1): 155–78.

- [16] Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. “Robust Inference With Multiway Clustering.” *Journal of Business & Economic Statistics* 29(2): 238–49.
- [17] Charron, Nicholas, Lewis Dijkstra, and Victor Lapuente. 2013. “Regional Governance Matters: A Study on Regional Variation in Quality of Government within the EU.” *Regional Studies*, forthcoming.
- [18] Chay, Kenneth, and Michael Greenstone. 2003. “The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession.” *Quarterly Journal of Economics* 118: 1121–67.
- [19] ——. 2005. “Does Air Quality Matter: Evidence from the Housing Market.” *Journal of Political Economy*. 116: 376–424.
- [20] Del Monte, Alfredo, and Erasmo Papagni. 2001. “Public Expenditure, Corruption and Economic Growth: The Case of Italy.” *European Journal of Political Economy* 17: 1–16.
- [21] ——. 2007. “The Determinants of Corruption in Italy: Regional Panel Data Analysis.” *European Journal of Political Economy* 23: 379–96.
- [22] Dincecco, Mark. 2009. “Fiscal Centralization, Limited Government, and Public Revenues in Europe, 1650–1913.” *Journal of Economic History* 69: 48–103.
- [23] ——. 2011. *Political Transformations and Public Finances*. Cambridge: Cambridge University Press.
- [24] Dincecco, Mark, and Gabriel Katz. 2013. “State Capacity and Long-Run Performance.” Mimeo, University of Michigan.
- [25] Durante, Ruben, Giovanna Labartino, and Roberto Perotti. 2011. “Academic Dynasties: Decentralization and Familism in the Italian Academia.” NBER Working Paper No 17572.
- [26] 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.
- [27] 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.
- [28] Fredriksson, Per and Noel Gaston. 2000. “Environmental Governance in Federal Systems: the Effects of Capital Competition and Lobby Groups.” *Economic Inquiry*, 38(3): 501–14.
- [29] Gennaioli, Nicola, and Ilia Rainer. 2007a. “The Modern Impact of Precolonial Centralization in Africa.” *Journal of Economic Growth* 12(3): 185–234.

- [30] ——. 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.
- [31] Gennaioli, Nicola, and Hans-Joachim Voth. 2011. “State Capacity and Military Conflict.” CEPR Discussion Paper No. 8699.
- [32] Gentzkow, Matthew. 2006. “Television and voter turnout.” *Quarterly Journal of Economics* 121(3): 931–72.
- [33] Gentzkow, Matthew, Jesse Shapiro, and Michael Sinkinson. 2011. “The Effect of Newspaper Entry and Exit on Electoral Politics.” *American Economic Review* 101: 2980–3018.
- [34] 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*, forthcoming.
- [35] 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.
- [36] Golden, Miriam A., and Lucio Picci. 2005. “Proposal for a New Measure of Corruption, Illustrated with Italian Data.” *Economics and Politics* 17(1): 37–75.
- [37] Graber, Doris A. 1984. *Processing the News: How People Tame the Information Tide*. New York, NY: Longman.
- [38] Greenstone, Michael. 2002. “The impacts of environmental regulations on industrial activity: evidence from the 1970 and 1977 clean air act amendments and the census of manufacturers.” *Journal of Political Economy*. 110: 1175–1219.
- [39] ——. 2004. “Did the Clean Air Act Cause the Remarkable Decline in Sulfur Dioxide Concentrations?” *Journal of Environmental Economics and Management*. 47: 585–611.
- [40] Henderson, J. Vernon. 1996. “Effects of air quality regulation.” *American Economic Review* 86: 789–813.
- [41] Holmström, Bengt. [1982] 1999. “Managerial Incentive Problems: A Dynamic Perspective.” *Review of Economic Studies* 66(1): 169–82.
- [42] Kahn, Matthew E. 1996. “New evidence on trends in vehicle emissions.” *Rand Journal of Economics* 27(1): 183–96.
- [43] Lindbeck, Assar, and Jörgen W. Weibull. 1987. “Balanced-Budget Redistribution as the Outcome of Political Competition.” *Public Choice* 52: 273–97.
- [44] List, John, and Daniel Sturm. 2006. “How Elections Matter: Theory and Evidence from Environmental Policy.” *Quarterly Journal of Economics*. 121: 1249–1281.



- [45] Lockwood, Ben. 2002. “Distributive Politics and the Costs of Centralization.” *Review of Economic Studies* 69: 313–37.
- [46] ——. 2006. “The Political Economy of Decentralization.” In *Handbook of Fiscal Federalism*, edited by Ehtisham Ahmad and Giorgio Brosio, 33–60. Cheltenham: Edward Elgar.
- [47] Nelson, Randy A, Tom Tietenberg, and Michael R Donihue. 1993. “Differential Environmental Regulation: Effects on Electric Utility Capital Turnover and Emissions.” *Review of Economics and Statistics* 75 (2): 368–73.
- [48] Oates, Wallace E. 1972. *Fiscal Federalism*. New York, NY: Harcourt Brace Jovanovich.
- [49] ——. 1999. “An Essay on Fiscal Federalism.” *Journal of Economic Literature* 37: 1120–49.
- [50] Persson, Torsten, and Guido Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge, MA: MIT Press.
- [51] Ponzetto, Giacomo A. M. 2011. “Heterogeneous Information and Trade Policy.” CEPR Discussion Paper No. 8726.
- [52] Ponzetto, Giacomo A. M., and Ugo Troiano. 2014. “Social Capital, Government Expenditures, and Growth.” CEPR Discussion Paper No. 9891.
- [53] Poole, Keith T., and Howard Rosenthal. 1985. “A Spatial Model For Legislative Roll Call Analysis.” *American Journal of Political Science* 29(2): 357–84.
- [54] Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
- [55] ——. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York, NY: Simon & Schuster.
- [56] Seabright, Paul. 1996. “Accountability and Decentralization in Government: An Incomplete Contracts Model.” *European Economic Review* 40: 61–89.
- [57] Snyder, James M., and David Strömberg. 2010. “Press Coverage and Political Accountability.” *Journal of Political Economy* 118(2): 335–408.
- [58] Strömberg, David. 2004. “Radio’s Impact on Public Spending.” *Quarterly Journal of Economics* 99: 189–221.
- [59] Treisman, Daniel. 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.
- [60] U.S. Department of Commerce. 1971. “Environmental Quality Control.” *State and Local Governments Special Studies*. U.S. Bureau of the Census.

- [61] U.S. Department of Commerce. 1980. “Environmental Quality Control.” *State and Local Governments Special Studies*. U.S. Bureau of the Census.
- [62] U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards. 2000. “National Air Pollutant Emission Trends, 1900–1998.” *EPA-454/R-00-002*. Research Triangle Park, NC.
- [63] U.S. Senate. 1970. “Hearings of the subcommittee on air and water pollution of the Committee on Public Works.” U.S. Government Printing Office.
- [64] Wallis, John J. 2000. “American Government Finance in the Long Run: 1790 to 1990.” *Journal of Economic Perspectives*, 14(1): 61-82.
- [65] ——. 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.
- [66] 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.

**Table 1 – The Evolution of SO<sub>2</sub> Emissions**

Year	Mean	Std. dev.	Min	Max
1960	0.1514627	0.1885529	0.01564	1.09353
1961	0.1477979	0.1831783	0.01545	1.08493
1962	0.1500735	0.1806726	0.01502	1.09312
1963	0.1543552	0.1808964	0.01428	1.11131
1964	0.1596994	0.1848661	0.01355	1.13862
1965	0.1668798	0.1909361	0.01354	1.16977
1966	0.1764556	0.1991029	0.01398	1.19072
1967	0.1746923	0.2021702	0.01445	1.21226
1968	0.1823104	0.2087637	0.01595	1.22756
1969	0.1843952	0.2120315	0.01683	1.22926
1970	0.1875458	0.2163975	0.01719	1.23151
1971	0.1743233	0.1939584	0.02091	1.05714
1972	0.1756837	0.1918978	0.01711	0.96356
1973	0.1855456	0.2141701	0.00308	0.9634
1974	0.1663952	0.1793253	0.01274	0.81897
1975	0.1502231	0.1591823	0.01186	0.76124
1976	0.1546588	0.1574864	0.00209	0.73957
1977	0.1553815	0.1535385	0.00209	0.71944
1978	0.140569	0.1271132	0.01784	0.64687
1979	0.1398869	0.1247468	0.01734	0.61479
1980	0.1327542	0.1207243	0.013	0.58772
1981	0.1203604	0.1206859	0.00915	0.59906

*Notes:* SO<sub>2</sub> Emissions in short tons per capita across the 48 continental United States.

**Table 2 – Summary Statistics**

Variable	Obs.	Mean	Std. dev.	Min	Max
SO <sub>2</sub> Emissions per Capita	48	.1875458	.2163975	.01719	1.23151
Newspaper Circulation	48	.2729864	.0539867	.1411094	.3912679
Income	48	3812.125	590.3823	2628	5071
Population (thou.)	48	4207.841	4372.018	333.795	20023.18
Land Area	48	61645.9	46807.72	1044.93	261797.1
White Population %	48	.8958356	.0885186	.6284792	.9960007
DW-Nominate	48	-.0626302	.2455653	-.5345	.5895
% Republicans in Office	48	.4913194	.3417551	0	1
% Republican Votes	48	.4592273	.121013	0	0.6279
Polluting Manufacturing %	48	.0635479	.0387529	.0077965	.2034424
Manufacturing %	48	.2222823	.0934438	.0432321	.3747373
Utilities %	48	.024722	.0055251	.0113665	.040028
Fossil Fuels (log)	48	2.969011	.4110257	2.357892	4.101017
Coal %	48	.1724713	.1638502	0	.6422735
Motor Gasoline %	48	.2045679	.0623228	.0654172	.356464

*Notes:* All variables are 1970 values. The DW-Nominate score is the average for the state's U.S. senators. The share of Republicans in office is computed considering the governor, the leaders of the two state legislatures, the two U.S. senators, and the majority of the state delegation to the U.S. House of Representatives. The share of Republican votes is the average in gubernatorial elections from 1968 to 1972. Polluting Manufacturing % is the share of value added represented by: Paper and allied products (SIC 26), Chemicals and allied products (SIC 28), Petroleum and coal products (SIC 29), Stone, clay, glass, and concrete (SIC 32), Primary metal industries (SIC 33). Fossil Fuels is the log of consumption relative to income. Coal % and Motor Gasoline % are shares of fossil-fuel consumption.

**Table 3 – Information and the Effects of the Clean Air Act**

Dependent Variable: SO <sub>2</sub> Intensity of Income (log), 1960 to 1981							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Newspaper Circ.	-1.641	-0.155	-0.158	-0.092	0.430	-0.102	0.392
× after 1970	(1.095)	(0.799)	(0.774)	(0.836)	(1.127)	(0.773)	(1.100)
Newspaper Circ.	0.319*	0.389*	0.390**	0.424**	0.517**	0.357**	0.431***
× t since 1970	(0.179)	(0.192)	(0.185)	(0.191)	(0.211)	(0.141)	(0.132)
Income (log)		-0.767	-0.535	-0.776	-0.337	-0.740	-0.198
× after 1970		(0.587)	(0.549)	(0.595)	(0.640)	(0.576)	(0.609)
Income (log)		-0.099	-0.130	-0.104	-0.018	-0.115	-0.048
× t since 1970		(0.087)	(0.091)	(0.085)	(0.105)	(0.069)	(0.069)
Pop. Density (log)		-0.048	-0.099	-0.046	-0.019	-0.039	-0.053
× after 1970		(0.057)	(0.073)	(0.056)	(0.062)	(0.052)	(0.059)
Pop. Density (log)		0.010	0.017*	0.012	-0.001	0.005	-0.002
× t since 1970		(0.008)	(0.009)	(0.007)	(0.011)	(0.006)	(0.007)
Polluting Mfg. %			3.278*				2.108
× after 1970			(1.611)				(1.246)
Polluting Mfg. %			-0.441				0.056
× t since 1970			(0.261)				(0.164)
DW-Nominate				0.066			0.001
× after 1970				(0.190)			(0.164)
DW-Nominate				0.036			-0.004
× t since 1970				(0.037)			(0.018)
Midwest					0.428**		0.394*
× after 1970					(0.201)		(0.195)
Midwest					-0.022		0.007
× t since 1970					(0.024)		(0.021)
South					0.600**		0.553**
× after 1970					(0.240)		(0.215)
South					0.029		0.038
× t since 1970					(0.035)		(0.026)
West					0.454		0.410
× after 1970					(0.286)		(0.285)
West					-0.026		0.002
× t since 1970					(0.044)		(0.034)
SO <sub>2</sub> Intensity (log)						0.080	0.032
× after 1970						(0.061)	(0.052)
SO <sub>2</sub> Intensity (log)						-0.049***	-0.050***
× t since 1970						(0.007)	(0.006)
Observations	1056	1056	1056	1056	1056	1056	1056
R <sup>2</sup>	0.939	0.941	0.942	0.941	0.943	0.946	0.949

*Notes:* Robust standard errors in parentheses, with two-way clustering by state and by year; p = 10%\*, 5%\*\*, 1%\*\*\*. All regressions include time f.e., state f.e., and state-specific linear time trends. Independent variables are 1970 values. Polluting Manufacturing % is the share of value added represented by: Paper and allied products (SIC 26), Chemicals and allied products (SIC 28), Petroleum and coal products (SIC 29), Stone, clay, glass, and concrete (SIC 32), Primary metal industries (SIC 33). The DW-Nominate score is the average for the state's U.S. senators.

**Table 4 – Alternative Controls for Economic Conditions**

Dependent Variable: SO <sub>2</sub> Intensity of Income (log), 1960 to 1981							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Newspaper Circ.	-0.158	-0.251	-0.164	-0.073	-0.183	-0.257	-0.264
× after 1970	(0.774)	(0.805)	(0.835)	(0.845)	(0.837)	(0.827)	(0.818)
Newspaper Circ.	0.390**	0.419**	0.389*	0.391*	0.458**	0.447**	0.486**
× t since 1970	(0.185)	(0.179)	(0.188)	(0.190)	(0.187)	(0.199)	(0.186)
Income (log)	-0.535	-0.647	-0.756	-0.254	-0.205	-0.785	-0.710
× after 1970	(0.549)	(0.620)	(0.625)	(0.607)	(0.605)	(0.762)	(0.819)
Income (log)	-0.130	-0.135	-0.098	-0.089	-0.119	0.073	-0.006
× t since 1970	(0.091)	(0.092)	(0.079)	(0.095)	(0.091)	(0.121)	(0.096)
Density (log)	-0.099	-0.087	-0.047	-0.031	-0.045	-0.099	-0.162
× after 1970	(0.073)	(0.078)	(0.058)	(0.051)	(0.058)	(0.078)	(0.105)
Density (log)	0.017*	0.022**	0.011	0.011	0.019**	0.031***	0.034**
× t since 1970	(0.009)	(0.010)	(0.008)	(0.007)	(0.007)	(0.011)	(0.016)
Polluting Mfg. %	3.278*						1.598
× after 1970	(1.611)						(1.559)
Polluting Mfg. %	-0.441						-0.230
× t since 1970	(0.261)						(0.277)
Manufacturing %		0.719					0.571
× after 1970		(0.728)					(0.785)
Manufacturing %		-0.217					0.024
× t since 1970		(0.139)					(0.162)
Utilities %			0.952				-8.200
× after 1970			(9.497)				(10.469)
Utilities %			0.057				-0.216
× t since 1970			(1.957)				(1.985)
Fossil Fuels (log)				0.375**	0.321**	-0.200	-0.229
× after 1970				(0.153)	(0.135)	(0.329)	(0.414)
Fossil Fuels (log)				0.007	0.040	0.183*	0.162*
× t since 1970				(0.036)	(0.029)	(0.089)	(0.083)
Coal %					0.354		-0.004
× after 1970					(0.320)		(0.275)
Coal %					-0.215***		-0.181**
× t since 1970					(0.043)		(0.074)
Motor Gas. %						-3.280	-3.332
× after 1970						(1.982)	(2.320)
Motor Gas. %						1.000*	0.660
× t since 1970						(0.547)	(0.547)
Observations	1056	1056	1056	1056	1056	1056	1056
R <sup>2</sup>	0.942	0.941	0.940	0.942	0.944	0.943	0.945

*Notes:* Robust standard errors in parentheses, with two-way clustering by state and by year; p = 10%\*, 5%\*\* , 1%\*\*\* All regressions include time f.e., state f.e., and state-specific linear time trends. All independent variables are 1970 values. Manufacturing %, Polluting Manufacturing %, and Utilities % are shares of state GDP. Polluting manufacturing industries are: Paper and allied products (SIC 26), Chemicals and allied products (SIC 28), Petroleum and coal products (SIC 29), Stone, clay, glass, and concrete (SIC 32), Primary metal industries (SIC 33). Fossil Fuels is the log of consumption relative to income. Coal % and Motor Gasoline % are shares of fossil-fuel consumption.

**Table 5 – Alternative Controls for Political Conditions**

Dependent Variable: SO <sub>2</sub> Intensity of Income (log), 1960 to 1981							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Newspaper Circ.	-0.092	1.707*	-0.132	0.320	0.272	-0.206	1.678
× after 1970	(0.836)	(0.966)	(0.783)	(0.750)	(0.780)	(0.848)	(1.221)
Newspaper Circ.	0.424**	0.484**	0.397*	0.375*	0.421*	0.381*	0.513**
× t since 1970	(0.191)	(0.228)	(0.197)	(0.203)	(0.205)	(0.185)	(0.235)
Income (log)	-0.776	-0.188	-0.727	-0.502	-0.641	-0.830	-0.452
× after 1970	(0.595)	(0.509)	(0.751)	(0.629)	(0.607)	(0.591)	(0.611)
Income (log)	-0.104	-0.069	-0.085	-0.106	-0.089	-0.068	-0.064
× t since 1970	(0.085)	(0.098)	(0.085)	(0.094)	(0.092)	(0.093)	(0.103)
Density (log)	-0.046	-0.136**	-0.052	-0.062	-0.057	-0.047	-0.140**
× after 1970	(0.056)	(0.065)	(0.054)	(0.057)	(0.057)	(0.055)	(0.065)
Density (log)	0.012	0.006	0.009	0.011	0.010	0.011	0.008
× t since 1970	(0.007)	(0.010)	(0.008)	(0.007)	(0.008)	(0.008)	(0.009)
DW-Nominate	0.066						-0.311
× after 1970	(0.190)						(0.206)
DW-Nominate	0.036						0.048
× t since 1970	(0.037)						(0.042)
White Pop. %		-2.741***					-3.789**
× after 1970		(0.948)					(1.418)
White Pop. %		-0.139					-0.065
× t since 1970		(0.156)					(0.180)
% R. in Office			-0.038				0.416
× after 1970			(0.253)				(0.262)
% R. in Office			-0.013				-0.041*
× t since 1970			(0.022)				(0.024)
% R. Votes				-0.749*			-0.434
× after 1970				(0.379)			(0.544)
% R. Votes				0.022			0.294**
× t since 1970				(0.102)			(0.135)
Elect. Evenness					-0.586		0.721
× after 1970					(0.357)		(0.475)
Elect. Evenness					-0.043		-0.290
× t since 1970					(0.130)		(0.173)
Sen. Class 1+3						0.047	-0.041
× after 1970						(0.130)	(0.100)
Sen. Class 1+3						0.006	0.012
× t since 1970						(0.019)	(0.021)
Sen. Class 2+3						-0.009	-0.126
× after 1970						(0.101)	(0.118)
Sen. Class 2+3						0.034	0.036
× t since 1970						(0.023)	(0.023)
Observations	1056	1056	1056	1056	1056	1056	1056
R <sup>2</sup>	0.941	0.943	0.941	0.941	0.942	0.941	0.945

*Notes:* Robust standard errors in parentheses, with two-way clustering by state and by year; p = 10%\*, 5%\*\* , 1%\*\*\* All regressions include time f.e., state f.e., and state-specific linear time trends. All independent variables are 1970 values. The DW-Nominate score is the average for the state’s U.S. senators. The share of Republicans in office is computed considering the governor, the leaders of the two state legislatures, the

two U.S. senators, and the majority of the state delegation to the U.S. House of Representatives. The share of Republican votes is the average in gubernatorial elections from 1968 to 1972. Election evenness is the average of  $1 - |\% \text{ R. Votes} - 0.5|$  in gubernatorial elections from 1968 to 1972.



**Table 6 – Channels of Pollution Abatement**

	Income	Pop. Den.	Poll. %	Mfg. %	Util. %	F. Fuel	Coal %	Gas. %
Newspaper Circ.	0.059	0.048	0.020*	0.022	0.011	-0.147	0.223	-0.017
× after 1970	(0.172)	(0.037)	(0.010)	(0.034)	(0.007)	(0.274)	(0.150)	(0.036)
Newspaper Circ.	0.002	-0.018	0.000	-0.009	0.001	0.097*	0.041	-0.013
× t since 1970	(0.025)	(0.024)	(0.005)	(0.009)	(0.002)	(0.052)	(0.025)	(0.008)
Income (log)	-0.067	-0.050*	-0.009	-0.018	0.003	0.216	-0.031	-0.035
× after 1970	(0.058)	(0.028)	(0.006)	(0.010)	(0.003)	(0.175)	(0.093)	(0.029)
Income (log)	0.022**	-0.023**	0.001	0.012***	-0.000	-0.008	0.001	0.003
× t since 1970	(0.010)	(0.010)	(0.002)	(0.004)	(0.001)	(0.019)	(0.013)	(0.004)
Pop. Den. (log)	-0.018**	0.004	-0.000	-0.004**	0.001**	-0.004	-0.015	0.004
× after 1970	(0.007)	(0.003)	(0.001)	(0.002)	(0.000)	(0.018)	(0.010)	(0.003)
Pop. Den. (log)	-0.003**	-0.005***	0.000	0.000	0.000	-0.001	-0.003**	0.001**
× t since 1970	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.002)	(0.001)	(0.000)
Observations	1056	1056	912	912	912	1056	1056	1056
R <sup>2</sup>	0.989	1.000	0.987	0.991	0.929	0.983	0.979	0.975

*Notes:* Robust standard errors in parentheses, with two-way clustering by state and by year; p = 10%\*, 5%\*\* , 1%\*\*\* All regressions include time f.e., state f.e., and state-specific linear time trends. Independent variables interacted with the 1970 break are 1970 values. Income and Population Density are logarithms. Manufacturing %, Polluting Manufacturing %, and Utilities % are shares of state GDP. Polluting manufacturing industries are: Paper and allied products (SIC 26), Chemicals and allied products (SIC 28), Petroleum and coal products (SIC 29), Stone, clay, glass, and concrete (SIC 32), Primary metal industries (SIC 33). Fossil Fuels is the log of consumption relative to income. Coal % and Motor Gasoline % are shares of fossil-fuel consumption.

**Table A1 – Information and the Effects of the CAA on NO<sub>x</sub> Emissions**

Dependent Variable: NO <sub>x</sub> Intensity of Income (log), 1960 to 1981							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Newspaper Circ.	0.041	0.121	0.120	0.150	0.557	0.165	0.552
× after 1970	(0.256)	(0.316)	(0.296)	(0.335)	(0.395)	(0.325)	(0.392)
Newspaper Circ.	0.160**	0.190**	0.190**	0.187**	0.189*	0.164**	0.157**
× t since 1970	(0.067)	(0.082)	(0.082)	(0.085)	(0.097)	(0.059)	(0.065)
Income (log)		0.020	0.087	0.016	0.055	0.142	0.138
× after 1970		(0.173)	(0.177)	(0.180)	(0.210)	(0.183)	(0.233)
Income (log)		-0.005	-0.006	-0.004	-0.032	-0.078***	-0.081***
× t since 1970		(0.026)	(0.028)	(0.026)	(0.036)	(0.022)	(0.016)
Pop. Density (log)		-0.015	-0.030	-0.014	-0.009	-0.008	-0.013
× after 1970		(0.020)	(0.026)	(0.020)	(0.025)	(0.020)	(0.032)
Pop. Density (log)		-0.003	-0.003	-0.003	0.001	-0.008***	-0.008**
× t since 1970		(0.004)	(0.005)	(0.004)	(0.005)	(0.002)	(0.003)
Polluting Mfg. %			0.939				0.640
× after 1970			(0.586)				(0.454)
Polluting Mfg. %			-0.013				0.132
× t since 1970			(0.116)				(0.091)
DW-Nominate				0.031			0.004
× after 1970				(0.058)			(0.061)
DW-Nominate				-0.004			-0.004
× t since 1970				(0.015)			(0.011)
Midwest					0.010		-0.007
× after 1970					(0.038)		(0.034)
Midwest					0.001		0.016*
× t since 1970					(0.009)		(0.008)
South					0.115*		0.084
× after 1970					(0.063)		(0.055)
South					-0.006		0.014
× t since 1970					(0.014)		(0.010)
West					0.105		0.091
× after 1970					(0.067)		(0.064)
West					0.013		0.018
× t since 1970					(0.017)		(0.014)
NO <sub>x</sub> Intensity (log)						0.071*	0.042
× after 1970						(0.039)	(0.028)
NO <sub>x</sub> Intensity (log)						-0.043***	-0.050***
× t since 1970						(0.011)	(0.009)
Observations	1056	1056	1056	1056	1056	1056	1056
R <sup>2</sup>	0.973	0.973	0.973	0.973	0.974	0.976	0.976

*Notes:* Robust standard errors in parentheses, with two-way clustering by state and by year; p = 10%\*, 5%\*\* , 1%\*\*\* All regressions include time f.e., state f.e., and state-specific linear time trends. Independent variables are 1970 values. Polluting Manufacturing % is the share of value added represented by: Paper and allied products (SIC 26), Chemicals and allied products (SIC 28), Petroleum and coal products (SIC 29), Stone, clay, glass, and concrete (SIC 32), Primary metal industries (SIC 33). The DW-Nominate score is the average for the state’s U.S. senators.

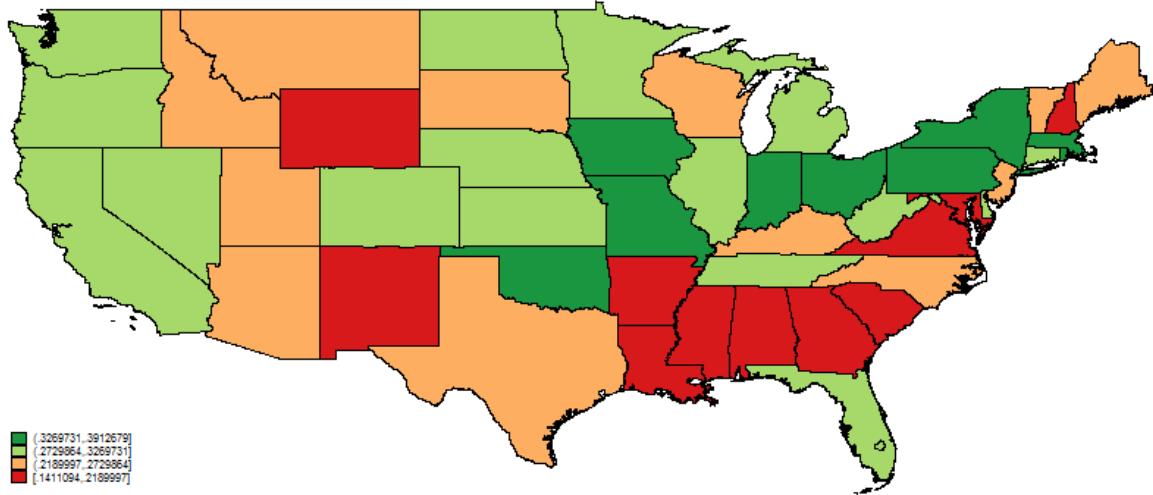
**Table A2 – Alternative Dependent Variables**

	SO <sub>2</sub> /cap.	SO <sub>2</sub> /cap.	SO <sub>2</sub> /sq.m.	SO <sub>2</sub> /sq.m.	SO <sub>2</sub> /\$	SO <sub>2</sub> /\$
Income (log) at $t$					-0.720*	-0.690**
					(0.416)	(0.328)
Pop. Den. (log) at $t$					-0.820	-0.525
					(1.132)	(1.270)
Newspaper Circ.	-0.096	0.352	-0.048	0.402	-0.073	0.391
× after 1970	(0.741)	(1.055)	(0.729)	(1.039)	(0.750)	(1.066)
Newspaper Circ.	0.392*	0.443***	0.374*	0.449***	0.377*	0.442***
× $t$ since 1970	(0.194)	(0.144)	(0.200)	(0.156)	(0.196)	(0.150)
Income (log)	-0.834	-0.221	-0.884	-0.294	-0.856	-0.252
× after 1970	(0.580)	(0.579)	(0.568)	(0.558)	(0.578)	(0.565)
Income (log)	-0.076	-0.035	-0.099	-0.066	-0.101	-0.055
× $t$ since 1970	(0.090)	(0.072)	(0.094)	(0.078)	(0.103)	(0.096)
Pop. Density (log)	-0.066	-0.074	-0.062	-0.066	-0.058	-0.063
× after 1970	(0.056)	(0.057)	(0.054)	(0.053)	(0.054)	(0.053)
Pop. Density (log)	0.007	-0.004	0.002	-0.009	0.004	-0.006
× $t$ since 1970	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.007)
Polluting Mfg. %		1.948		1.901		1.973
× after 1970		(1.245)		(1.219)		(1.234)
Polluting Mfg. %		0.108		0.122		0.099
× $t$ since 1970		(0.179)		(0.177)		(0.173)
DW-Nominate		0.015		0.049		0.028
× after 1970		(0.158)		(0.156)		(0.160)
DW-Nominate		-0.006		-0.004		-0.004
× $t$ since 1970		(0.017)		(0.018)		(0.018)
Midwest		0.415*		0.414**		0.408*
× after 1970		(0.200)		(0.198)		(0.198)
Midwest		0.006		0.001		0.004
× $t$ since 1970		(0.023)		(0.024)		(0.022)
South		0.559**		0.547**		0.551**
× after 1970		(0.210)		(0.208)		(0.209)
South		0.038		0.035		0.037
× $t$ since 1970		(0.027)		(0.027)		(0.026)
West		0.383		0.380		0.390
× after 1970		(0.274)		(0.268)		(0.274)
West		0.009		0.012		0.008
× $t$ since 1970		(0.034)		(0.035)		(0.035)
SO <sub>2</sub> Intensity (log)		0.036		0.037		0.035
× after 1970		(0.051)		(0.049)		(0.051)
SO <sub>2</sub> Intensity (log)		-0.050***		-0.050***		-0.050***
× $t$ since 1970		(0.006)		(0.006)		(0.006)
Observations	1056	1056	1056	1056	1056	1056
R <sup>2</sup>	0.936	0.945	0.974	0.977	0.941	0.949

*Notes:* Robust standard errors in parentheses, with two-way clustering by state and by year;  $p = 10\%$ \*,  $5\%$ \*\* ,  $1\%$ \*\*\* All regressions include time f.e., state f.e., and state-specific linear time trends. Dependent variables are logarithms. Independent variables interacted with the 1970 break are 1970 values. Polluting Manufacturing % is the share of value added represented by: Paper and allied products (SIC 26), Chemicals and allied products (SIC 28), Petroleum and coal products (SIC 29), Stone, clay, glass, and concrete (SIC 32), Primary metal industries (SIC 33). The DW-Nominate score is the average for the state's U.S. senators.

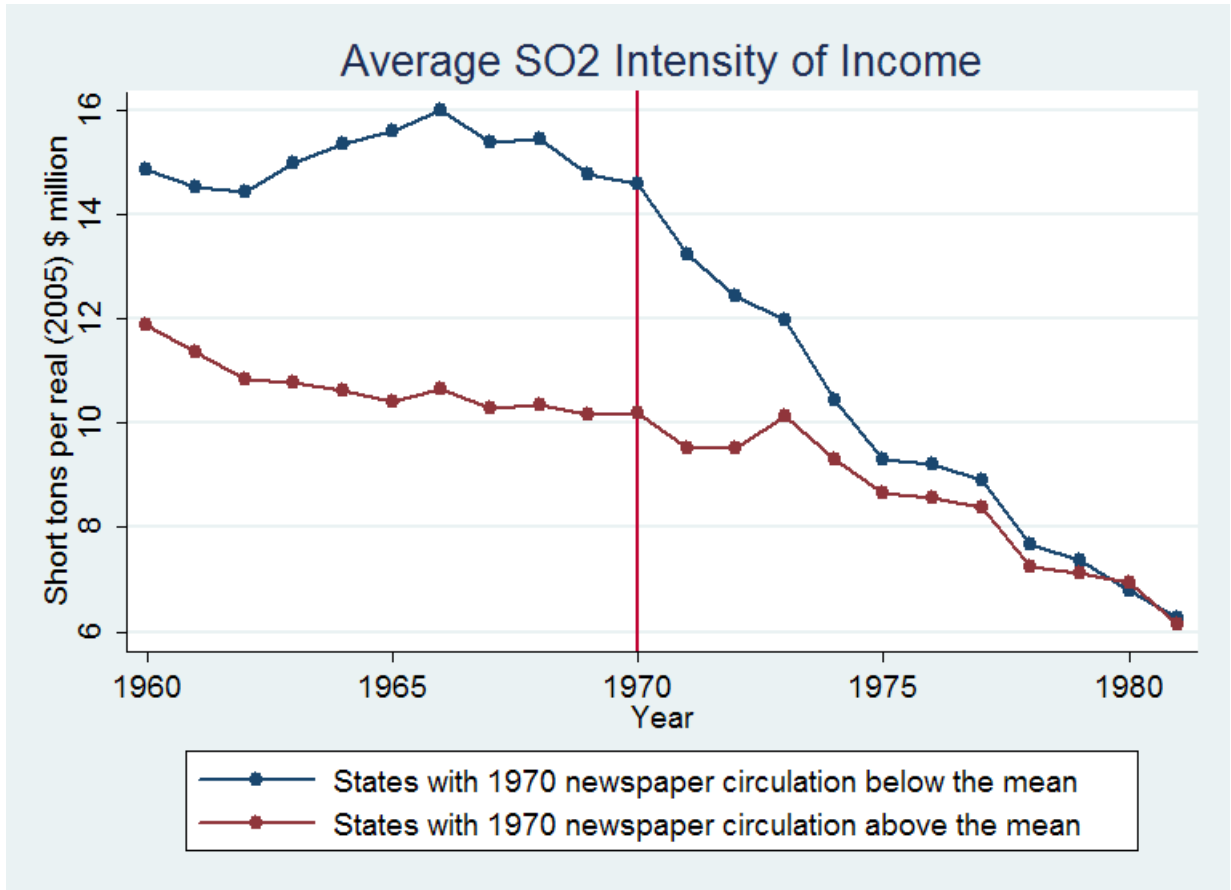
Figure 1 – Information across the United States

Average Daily Newspaper Circulation per Capita, 1968-72



Source: Gentzkow, Shapiro, and Sinkinson (2011)

Figure 2 – Information and the Impact of the Clean Air Act



Sources: Emissions are from the EPA, newspaper circulation from Gentzkow, Shapiro, and Sinkinson (2011), and personal income from the BEA Regional Economic Accounts.