

## CHAPTER 4

---

# The political economy of federalism

ROBERT P. INMAN AND DANIEL L. RUBINFELD

Recent changes in the institutions of government economic policymaking in western and eastern Europe, efforts to write new constitutions for Russia and South Africa, and the less dramatic, but no less important, pressures to redefine policy responsibilities between levels of government in such contrasting economies as the United States and China suggest that now is a good time to revisit a long-standing topic of traditional political economy: the design of federalist political institutions.

The framing in 1787 of the United States Constitution marked the beginning of the contemporary debate, pitting Montesquieu's ideal of a decentralized "confederate" republic composed of sovereign member city-states against the vision of Madison and the other Federalists of a "compound" republic with an overarching central government responsible to the union's common citizenry. The tension between the confederate model of independent city-states each with an effective veto over central government actions and the compound model of a central government capable of acting against local interests remains at the center of today's debates over the design of federalist constitutions. Can contemporary political economy help us to understand better the important trade-offs implicit within the choice between a confederate and a compound republican constitution? This survey seeks to provide one answer.<sup>1</sup>

The authors would like to thank the Smith Richardson Foundation and the National Science Foundation (SES-9022192) for financial support and the Center for Advanced Study in the Behavioral Sciences for a congenial environment in which to pursue our joint work on federalism. The comments on an earlier version of this paper by William Fletcher, Michael Fitts, David Lieberman, Dennis Mueller, and seminar participants at Washington University of St. Louis and the Center for Advanced Study (Stanford) are gratefully acknowledged.

1. The focus of this survey will be on the "political" theory of federalism – a theory of federalist institutions that maps the multiplicity of citizen interests onto policy decisions. The political theory of federalism is grounded in social choice theory and the new political economy.

There is a parallel literature in economics and public administration seeking to develop an "administrative" theory of federalism – that is, a theory of institutions that

Section 1 reviews the political and economic arguments for the confederate republic, while Section 2 reviews the economic theory behind the compound republic. Two federalist dimensions of any republic's constitution are identified: representation (*R*) of the local city-states to the central government and the assignment (*A*) of governmental tasks to the city-states or the central government. We suggest one approach for choosing the efficient compound republic as a preferred combination of *R* and *A*.

The analysis in Sections 1 and 2 will reveal a central tension between economic efficiency and democratic rights and virtues. Choosing values of *R* and *A* ultimately requires a balancing of these potentially competing economic and political objectives. The decentralized federalist structure of the confederate republic favors the goals of democracy; the more centralized compound republic places a greater weight on economic efficiency.

Section 3 explores the stability of federalist constitutions: When might an individual city-state secede from a confederate or a compound republic? Economic conditions and political institutions required to hold a federalist republic together are identified.

Section 4 offers a concluding comment on directions for new research.

## **1. City-states and political union: The confederate republic**

### *1.1 The Political case for the city-state: Protecting democratic rights and promoting civic virtue*

From Plato and Aristotle through Rousseau and Montesquieu to contemporary federalist legal scholars,<sup>2</sup> one hears continuing praise for the political virtues of small government. Small governments are seen by many to protect the rights of the individual against the tyranny of the majority. Small governments are also seen to encourage political participation, the mutual accommodation of all views, political compromise, and the value of community.

Controlling the ability of the majority to overrule the rights of the

turns policy decisions into allocative outcomes. The latter draws its theoretical structure from the principal-agent and transactions cost literatures; see Holmström and Tirole (1989) generally, Tirole (1994) for an application of these literatures to government, and Bender and Mookherjee (1987) and Klibanoff and Morduch (1993) for an application to the design of federalist institutions. The large empirical literature on the effects of incentive grants on local government allocations will provide a useful basis for the application of the new administrative theory of federalism. See Inman (1979) and more recently Wildasin (1986) for surveys of that literature.

2. See, for example, Frug (1980, 1987, 1993), Rapaczynski (1986), and Amar (1987).

individual is a fundamental challenge for democratic societies. For several reasons, small democratic city-states are often proposed as a solution. First, small city-states are argued to be more fluid politically; that is, open to the vagaries of the voting cycles. With no stable majority, the distinction between the ruler and the ruled is blurred. Tyranny is less attractive if there is a significant risk of being tyrannized in return (Frug 1980, 1069). Second, exit from a repressive state to a less repressive state is possible. Mobile individuals can join that city-state having their most preferred set of rights and responsibilities (McConnell 1987, 1503). Third, no small city-state is likely to be able to defeat and oppress another city-state without high military cost.

In addition to protecting individual rights, small governments also are seen as the wellspring of the virtuous public life. To the Greeks, to Rousseau and Montesquieu, to de Tocqueville and Dewey, political participation is the source of public values.<sup>3</sup> Through political participation citizens become educated and come to know fully, and to respect, the points of view of others. For classical and contemporary communitarians, only small governments are likely to encourage the required level of participation and thus foster the valued public virtues.

### *1.2 The economic case for a political union*

Though small governments may protect individual rights and foster public virtue, it has long been appreciated that small governments are vulnerable in modern commercial economies. Larger nations can overrun a small city-state militarily, unless it is protected geographically by impassable mountains or narrow harbors. Even without a military threat, small states might be disadvantaged in trade and commerce through their imposition of self-defeating competitive tariffs or their failure to control the overutilization of their shared natural resources. Missing was a theory of government that balanced the political advantages of the small city-state with the necessity of providing for their common economic interests. Montesquieu's vision of a confederate republic sought to meet this need.

To Montesquieu "if a republic be small it is destroyed by a foreign force; if it be large, it is ruined by an internal imperfection [of a] thousand private views" (quoted in Beer 1993, 219, 220). Montesquieu sought a solution to this dilemma through a "confederate republic" com-

3. Frug (1980, 1067–73) provides a useful overview of the importance of participation for political life in the writings of Aristotle, Rousseau, de Tocqueville, and Arendt. For more detail, see Dahl and Tufte (1973, chap. 1) and Pateman (1976). Frug (1980, 1987) provides a contemporary application of these theories to U.S. urban policy.

posed of small city-states acting in unison when their common interests are at risk. The United Netherlands and the Swiss cantons were his models. Montesquieu's ideal confederate republic is a mutual agreement between city-states to pursue joint economic objectives, each state having the right to exit the union when it feels its private interests are harmed by the action of the republic. While secession is a right, decision making is by bargaining.<sup>4</sup> How well the confederacy performs its tasks depends upon how well the union meets the conditions for efficient bargains. From a contemporary perspective, Montesquieu's ideal confederate republic can be seen as a bargain among independent city-states. The theory can be understood and evaluated on these terms.

In a powerful challenge to the need for strong central governments in modern economies, Ronald Coase (1960) suggests that market failures created by economic spillovers or the inefficient provision of public goods can be resolved by successful bargaining between the affected parties to the extent that the property rights of the parties are well specified and the costs of bargaining are low (see Coater 1987a). Coase's theory has direct application to Montesquieu's confederate republic as a bargain between city-states. The assumptions underlying the Coasian framework are

- (C1) There are no resource costs associated with reaching agreement.
- (C2) Preferences over bargaining outcomes and the resources of participants are common knowledge.
- (C3) Bargaining agents perfectly represent the economic interests of their constituents.
- (C4) All bargaining agreements are costlessly enforceable.
- (C5) The parties will agree to a division of the economic surplus from bargaining.

Assumptions (C1) through (C5) ensure that economic efficiency will be achieved in a fully decentralized confederation of city-states. Assumption (C1) ensures that bargainers can get together and reach agreements costlessly – there are no decision costs. Assumptions (C2) and (C3) guarantee that all parties are fully informed – there are no costs of revelation. Assumption (C4) establishes that all agreements are valid and enforceable – there are no monitoring or enforcement costs. Finally,

4. Montesquieu discusses alternative models for organizing confederate republics, some of which allow a majority of member states to discipline a corrupt member. His ideal, however, remains decision making by unanimity. See Beer (1993, chap. 7) for a summary of Montesquieu's theory of federalism.

assumption (C5) ensures that bargains do not unravel over the question of how to share the economic surplus from agreement.

In Montesquieu's confederate republic the task of the central government is to provide property rights so as to protect the integrity of participating city-states and facilitate and enforce all city-state agreements for the control of interstate spillovers (e.g., barriers to free trade) and the provision of public goods (e.g., a common defense). Coasian bargains in the halls of government ensure that economic efficiency obtains. The tasks of the small city-states are to protect individual freedom and to encourage political participation. Political debate and open discussion fosters a spirit of community and respect for the views of others. In the end, the confederate republic is hypothesized to be both free and efficient.

### *1.3 Are confederate republics free and efficient?*

How valid is the theory of the confederate republic, resting as it does on a belief in the virtues of community and the efficiency of Coasian bargains? There are a number of reasons to be skeptical.

The evidence relating small governments to the protection of individual rights and to good government is mixed at best. History is replete with counterexamples wherein small governments repress the rights of minorities and larger, more encompassing, governments protect those rights. But there are valid examples that make the opposite point.<sup>5</sup> And there is no compelling evidence that countries that explicitly protect small governments through confederate constitutions have better human rights records (see Riker 1964, 14–16).

Studies that have sought to draw the relationship between the size of the state and political participation also have given mixed results. For example, there appears to be no systematic evidence that voter turnout varies with the size of the electorate (Dahl and Tufte 1973, 45; Verba, Nie, and Kim 1978, chap. 13). Surveys do show that citizens are more likely to seek to influence their local governments than national governments through nonelectoral channels and that these informal activities are more likely in stable rural communities than in large, urban centers (Verba, Nie, and Kim 1978, chap. 13). Further, these local efforts are likely to have a greater effect of policy (Dahl and Tufte 1973, 58–9). Still, no more than 28 percent of those surveyed in the United States have ever tried to influence their local governments, and of those only a

5. U.S. southern states' support of slavery and recent Serbian efforts at ethnic cleansing are examples where small states suppressed rights, while the larger state – United States and Yugoslavia – protected rights. Conversely, Germany and the Holocaust, and the United States's historical treatment of Native Americans, suggest the opposite.

bit more than half felt it was likely they would have a significant effect on policy; the percentages were substantially lower in the other countries surveyed.<sup>6</sup>

Against this evidence, it is difficult to see the small city-state as the necessary protector of individual rights and promoter of democratic values. Certainly, small government can contribute to these political goals, but today these institutions are far removed from the Greek ideal. If we are to make a decisive case for the small city-state in a theory of federalism, additional arguments must be advanced.

The dependence on efficient and costless bargaining of the confederate republic theory is also problematic. One can be critical of each of the five Coasian assumptions for efficient bargains. Contrary to (C1), representatives must communicate positions and consider alternatives, which is often a time-consuming, costly task. Rarely is there full information about the parties' preferences and endowments as required by (C2). Parties may fail to find the best allocations because they do not know the full range of possible trades. If the parties make poor estimates of each other's threat points when bargaining, they may underestimate the potential surplus to be gained and consequently take a hard line in the bargaining process. To the extent that parties succumb to such strategic behavior, Coasian bargains may not occur (see Mailath and Postlewaite 1990). Preferences and economic endowments of the bargainers can be revealed, but only at a cost (see Laffont 1987). Again, not all inter-governmental bargains will be efficient.

Assumption (C3) assumes perfect agency on the part of city-state representatives to the confederate republic, but this too seems unlikely. Representatives might have an agenda independent of the citizens of the city-state they represent. These imperfect agents must be identified and replaced – at a cost. How will the agents be chosen and whom will they represent? If the citizens within each city-state have distinct preferences for the outcomes of the republic's bargains, then, as Bowen (1943) first showed, there is no guarantee that the median's representatives under majority rule voting will approve fully efficient resource allocation for all the citizens of each city-state.

Without enforceability – assumption (C4) – city-states will have the incentive to break agreements as conditions or interests change over

6. See Dahl and Tufte (1973, 58–9). A more recent survey by the Advisory Commission on Intergovernmental Relations of citizen participation found that most local participation occurs through selected advisory committees (60 percent) rather than through general citizen discussion (16 percent) as envisaged by the advocates of political participation; ACIR (1979, 223). Only 3 percent of the surveyed officials said citizen participation resulted in a “substantial” change in the budget; 66 percent said “very little change”; see ACIR (224).

time; consequently, members of the confederation may have little incentive to reach agreements in the first place.<sup>7</sup> Control of cheating can be restored, but only at a cost of monitoring the agreement and such costs can be substantial (see Williamson 1985).

Without assumptions (C1–C4) there will be potentially important revelation, decision, monitoring, and enforcement costs when reaching agreements in Montesquieu's confederate republic. Some bargains that were originally efficient under assumptions of costless communication and full information may no longer be efficient when transactions costs are positive. More importantly, when transactions costs are significant, the efficient federalist structure may no longer be a confederate republic of many city-states (see Section 2.3).

Finally, if assumption (C5) no longer holds (even if transaction costs are zero), the parties may not be able to divide the economic surplus. Every Coasian bargain does two things: it establishes an efficient exchange creating economic surplus *and* it distributes that surplus among the bargaining parties. Proponents of Coase emphasize the first but have largely ignored the implications of the second (see Cooter 1982). Yet it is well known that the division of any economic pie is a bargaining problem that may have no solution. Only in the special case of two players with a pre-established order to their offers is an efficient allocation likely (see Rubinstein 1982). When the number of participants is greater than two, a shared sense of fairness – called a “focal point” – may be required (see Schelling 1960 and Haller 1986). Assumption (C5) is meant to guarantee such a focal point. Without it, however, confederate republics may fail to achieve full economic efficiency.

In summary, while the theoretical case for the confederate republic can be made, available evidence is far from decisive that such confederations will achieve their goals. Modern city-states do not appear to be unique protectors of democratic rights and civic virtues. Efficient bargains that underlie the economic case for the confederation's central government are unlikely to be achieved in practice.<sup>8</sup> Against these facts,

7. A point appreciated by Madison when criticizing the Articles of Confederation; see Beer (1993, 248–9).

8. One telling example is the failure of the Continental Congress during the period 1777 to 1785 adequately to provide for the financing of the U.S. War of Independence and to effectively manage foreign affairs following the war. The Congress had been established by the Articles of Confederation as a vehicle for “Coasian agreements” between the colonies. These failures led to the calling of the Constitutional Convention in 1787 and, ultimately, to the formation of Madison's compound republic; see Rakove (1989, 5–6).

Sandler and Murdoch (1990) offer a more contemporary example of the failure of Coasian bargains to efficiently provide public goods. In their study of NATO allocations, they conclude that for the period 1956 to 1987, NATO members behaved inefficiently, free riding on other nations's contributions to the collective defense of western Europe.

confederate republics may just as easily be unjust and inefficient. Alternatives must be considered.<sup>9</sup>

## 2. The political economy of the compound republic

While the classical and contemporary communitarians championed the city-state for its potential to promote democratic rights and civic virtues, James Madison valued the small city-state for its ability to satisfy the economic needs arising from local circumstances. A relatively strong central government, in sharp contrast to the fragile alliance of a confederate republic, is the core of Madison's compound republic. The central government receives its legitimacy not from the unanimous consent of all city-states but from the majority approval of all its individual citizens. The preferences of citizens are expressed through the local election of representatives to a national legislature and through the national election of a single executive. The executive implements the laws approved by the legislature. While Montesquieu might have asked "Why have a Union?" Madison would have asked "Why have states?" In contrast to Montesquieu, Madison's case for city-states was economic not political; they better serve the "local and particular."<sup>10</sup> Public finance economists know this argument today as the Tiebout (1956) theory of local government.

### 2.1 *The economic case for the city-state: Efficiency*

In the Tiebout model of intergovernmental competition, small cities competing for mobile residents ensure the efficient supply of local public goods.<sup>11</sup> Five conditions define the Tiebout economy:

9. Were the theory of a confederate republic to rest only upon the efficiency advantages of Coasian bargains between small city-states it would not be a theory of federalism. Logically, if the Coase assumptions hold, both a single central government or a loose confederation of states could achieve the fully efficient outcome (the former through bargains among representatives to the national legislature, the latter through bargaining among city-states). Montesquieu rightly finds his reasons for local governments in their hypothesized contribution to the democratic good of civic virtue and citizen devotion to the public welfare.
10. Madison feared that local economic interests would be lost within the national legislature unless separate states were established to provide those interests with "superintending care"; again, see Beer (1993, 293–4) for a full review of Madison's economic argument.

In contrast to Montesquieu and the Anti-federalists Madison was skeptical of the political arguments for small governments. He saw the central government as the individual's best protection against tyranny; see McConnell (1987, 1500–7).

11. We only sketch the argument here, simply to place Tiebout's argument within the broader literature on federalist political institutions. Rubinfeld (1987) and Wildasin (1986) provide detailed surveys of this literature.



- (T1) Publicly provided goods and services are produced with a congestible technology.
- (T2) There is a perfectly elastic supply of jurisdictions, each capable of replicating all attractive economic features of its competitors.
- (T3) Mobility of households among jurisdictions is costless.
- (T4) Households are fully informed about the fiscal attributes of each jurisdiction.
- (T5) There are no interjurisdictional externalities.

A Tiebout equilibrium arises when each household resides in the jurisdiction of its choice, and where no household can improve its economic welfare by moving elsewhere.<sup>12</sup>

With a congestible technology (T1), publicly provided goods are rival public goods, and the per-household cost of providing each level of a public good first decreases and then increases as more households move into the jurisdiction. For each level of output, there is a technically efficient population size that minimizes the average cost per household of providing that service. Ensuring that these efficient alternatives exist in the Tiebout economy is the task of public sector entrepreneurs. Implicit in assumption (T2) is the fact that these entrepreneurs are fully capable of providing all public goods and private amenities (including jobs or job access) in each jurisdiction.

Assumptions (T3) and (T4) ensure that there are informed and mobile citizens to discipline the public sector entrepreneurs. When inefficiencies arise, households move to otherwise similar jurisdictions with efficient production. The no spillover assumption (T5) ensures that all public goods can be provided within these efficient jurisdictions and that no citizens can consume the public service without paying their full marginal cost of consumption.

Current empirical evidence suggests that most public services satisfying (T1) and (T5) can be efficiently provided at minimum cost in communities with populations of five to ten thousand residents (see Rubinfeld 1987, for a summary of that literature). Further, such communities are allocatively efficient, providing all citizens with their preferred levels of the public goods at each good's minimum production costs.

What happens to the argument if one or more of the Tiebout assumptions is lost? Relaxing assumption (T1) does not fundamentally alter the

12. The Tiebout argument focuses on static efficiency. Not considered are issues of dynamic efficiency. Weingast (1993), however, has made the argument that Tiebout fiscal competition will limit the taxation of mobile capital enhancing investment and economic development. Tiebout competition is also seen by many as source of innovation in public goods technologies and management, though Rose-Ackerman (1980) is skeptical.

efficiency advantages of a fully decentralized Tiebout economy, *if* assumptions (T3) and (T5) still hold. Thus, even if the production of government services has a public component, efficiency remains, provided citizens can costlessly relocate to consume the service (T3) and exclusion (T5) is still possible. Entrepreneurs can sell their excess capacity to other jurisdictions at competitive prices. In effect, one local jurisdiction can become the public supplier for the other jurisdictions. Governments will vary in size depending on the technology of the services supplied (Olson 1969).

Assumptions (T2–T5) are more fundamental. With the loss of Tiebout assumption (T2) – the elastic supply of jurisdictions – three potentially adverse consequences for economic efficiency are likely to arise. First, because their number is limited, each jurisdiction is composed of citizens with different demands for local public goods. Local politics now becomes important. If local decisions are made by majority rule, for example, the median voter's choice need no longer be economically efficient – a point first emphasized by Bowen (1943). Second, when the public goods economy can no longer replicate attractive local jurisdictions, economic rents can be earned. The presence of these publicly created rents in turn introduces an additional layer of economic bargaining – between the fiscal entrepreneur and the owners of the scarce factor. The presence of rents requires lump-sum taxes and transfers for their efficient distribution. Without such policy, bargaining may lead to inefficiencies in the public goods economy (see Epple and Zelenitz 1981 and Wildasin 1986). Third, once the supply of jurisdictions is fixed, the analogy of the decentralized model of federalism to free market competition is no longer valid. Each local government may now see itself as a Cournot competitor, aware of the potential effect of its action on other jurisdictions, while assuming no direct competitive response. In this framework, the actions of each local jurisdiction, even if small, can have a substantial effect on the welfare of the rest of the decentralized world. Essentially, a self-interested decision by those in power in one jurisdiction can result in a series of distortions arising throughout the affected network of other jurisdictions.<sup>13</sup>

When mobility becomes costly and assumption (T3) no longer applies, the perfect Tiebout economy cannot guarantee overall (market plus public sector) efficiency. When the production of the public good is site-specific *and* when the relocation of the household is costly on the *margin* (i.e., the relocation affects private resource allocations), household relo-

13. For example, states might be encouraged to relax their environmental controls to encourage business migration, or to forestall the loss of business to other states with relaxed environmental regulations; see Oates and Schwab (1988).

cations to achieve a more efficient public sector will create real externalities for the private economy. The usual example is labor leaving one jurisdiction for less productive employment in another fiscally more attractive jurisdiction; the consequence is a less efficient private economy (see Buchanan and Goetz [1972] for the initial presentation of this argument and the summaries in Boadway and Flatters [1982] and Wildasin [1986]).

When assumption (T4) no longer holds, that is, when there is incomplete information about the costs and levels of public goods production, there are additional difficulties for the Tiebout economy. Households have two options: go it alone and risk being exploited by better informed fiscal entrepreneurs and local factors of production *or* hire an agent (called a “city manager” or “mayor” for households in place, and a “real estate agent” for relocating households) to protect household interests. To the extent that the market for “agents” works perfectly – all inefficient or exploitive providers are exposed, and only truthful agents survive – the final public goods outcome will still be efficient, but some households’ fiscal surpluses will be allocated to the agents as payment for services performed.

Complications arise, however, when information about agent performance is costly and some deceitful agents survive, or when the expertise provided by agents to households cannot itself be kept private. In both instances externalities in the market for agents will cause an inefficient amount of information to be provided about the fiscal performance of the Tiebout communities.<sup>14</sup> Inefficient information leads to inefficient public sector allocations.

Assumption (T5) – no spillovers – ensures that all communities can exclude consumers who free ride by not contributing to the costs of providing their public goods. With free riding there will typically be an underprovision of public services by competitive city-states (see Pauly 1970).

What is the evidence that competitive city-states do enhance the economic performance of the public goods economy? First, with competitive governments, production costs for any level of output should be minimized, as should rent transfers from one group to another. Thus,

14. With costless information, agent efforts can be monitored (avoiding moral hazard) and agent abilities will be known (avoiding adverse selection). When information about agents is costly, however, there is no guarantee that the agent market will be efficient; see Holmström (1985). Similarly, when the information provided by the agents to households cannot be kept private, the information market will be inefficient; see Allen (1990).

were decentralized and centralized public goods economies providing the same levels of public good outputs to be compared, the decentralized public economy composed of many city-states would have lower spending and taxes than the centralized single-state economy, all else equal.

Testing this efficiency prediction has recently become known as the search for Leviathan. (The presence of a powerful central rent-seeking authority, Leviathan, would be evidence of inefficiency.) Oates's (1985) pioneering efficiency study regresses total state and local government spending per capita in a state on the aggregate number of local governments per capita in the state and finds no significant negative effect on spending. Using the Oates's approach and data, Nelson (1987) and Zax (1989) find that more governments per square mile and more general purpose governments per capita reduce government spending per capita, while states with many special district governments and more governments per capita – typically rural states – have higher spending per capita, all else equal. Finally, Forbes and Zampelli (1989) seek to explain county government spending by the number of counties in the SMSA and find no effect.

There is a pattern in these conflicting results. Very small governments (special districts in rural states) are too small to take advantage of economies of scale. A few large governments (counties) are not sufficiently competitive to ensure efficiency. Between these extremes – namely, general purpose governments in urbanized states – competition does appear to affect spending.

Although the Leviathan efficiency results are suggestive, they are not decisive. One cannot infer cost and demand performance from aggregate expenditure studies without very strong identifying restrictions.<sup>15</sup> A preferred approach is to test directly for the effects of competitive governments on public sector costs. Inman (1982) and Zax (1988) provide two tests of the effects of a competitive suburban fringe on center city labor costs, confirming the expected negative effect of competition on costs.<sup>16</sup>

15. If a decentralized governmental structure determines the cost of public goods and the demand for public goods depends on a vector of demand variables and costs, then the expenditure equation to be estimated will involve unspecified interactions of cost and demand variables.

16. Zax (1988) finds that more suburban governments reduce aggregate labor expenditures in their adjoining central city. Inman (1982) refines the test by allowing for income differences between city and suburban residents, noting that when city incomes are significantly greater (a rich city with a rural fringe) or significantly less (a poor city with a rich fringe) than suburban incomes, then suburbs do not offer an effective competitive alternative for current city residents. A U-shaped relationship between the ratio of city to suburban income and wage and benefits is predicted and confirmed.

Even if costs are minimized, however, there is no guarantee that small city-states provide residents with their preferred levels of publicly provided goods. Specifically, there is no certainty that the Samuelson condition for allocative efficiency within city-states (the sum of the individual residents' marginal benefits being equal to the marginal cost of production) will be satisfied. Again, there are indirect and direct tests of this proposition.

In a series of papers Brueckner (1979, 1982, 1983) shows that, under certain restrictive conditions, communities that provide public goods so as to maximize the value of residential property will satisfy the Samuelson condition.<sup>17</sup> Brueckner (1979, 1982) finds that education is overprovided in New Jersey communities, while other services are efficiently provided. The latter conclusion also applies to Massachusetts.

Barlow (1970) was the first to offer a direct test of allocative efficiency within small city-states. He used estimated citizen demand curves for education obtained from a median voter model and aggregate community data to infer marginal benefit schedules for community residents. Knowing the distribution of demand characteristics within each community, Barlow then calculated the sum of residents' marginal benefits for education at the community's current level of school spending. Finally, comparing the sum of marginal benefits to the marginal cost of producing education, Barlow concluded that education was overprovided (from the perspective of community residents).

Using sophisticated microeconomic techniques to estimate family demands from survey data, Bergstrom et al. (1988) repeated the Barlow test for intrajurisdictional allocative efficiency using a sample of Michigan school districts. Importantly, they were able to identify a potential source of overprovision in some communities – the U.S. federal government's provision of federal tax subsidies for school spending. They concluded, however, that government provision of education was locally efficient within the majority of communities, given the rate of federal subsidy.

Though not fully decisive on the point, the current empirical evidence suggests competitive local governments can provide an efficient level of congestible (local) public goods. In such economies, production costs are minimized and citizens often do vote for the efficient level of the local public good. These are encouraging results.

17. Brueckner (1983) qualifies his efficiency result by noting that the property value maximization test assumes each household is a "community taker," where the allocation of households across communities is given. The value maximization condition is therefore a local efficiency condition. However, reallocation of households across communities may improve the welfare of all households. Eberts and Gronberg (1981) provide one effort to test this global efficiency condition, concluding that families collect in communities by their demands for public goods.

What is not assured is the efficient allocation of public goods with significant spillovers. In this case, a subsidy is needed to internalize the externalities. But any such policy to control interjurisdictional spillovers would require the agreement of the competitive city-states. For such agreements we must look to more encompassing political institutions. In Madison's compound republic this is the representative central government.<sup>18</sup>

## 2.2 *The political economy of representative government*

The focal institution of Madison's compound republic is the central government legislature composed of representatives elected from subsets of the city-states. Henceforth, we call these groups of city-states "states." Within the central legislature representatives from the states are free to fashion coalitions and set policies, subject only to the constraint that policies receive the approval of a majority of the legislators. Though numerous specifications of this legislative game are possible, two are prominent in the political economy literature – the minimum-winning-coalition legislature (MWC) and the universalistic (U) legislature. We explore implications of each for the design of federalist constitutions.

*The minimum-winning-coalition legislature.* Policies chosen in the MWC legislature reflect the preferences of the winning 51 percent majority. When policies are one dimensional (e.g., spending on a single, national public good) the outcome will be that preferred by the median representative in the legislature. When policies are multidimensional (e.g., spending on projects with local benefits such as income redistribution or local public goods) then additional legislative structure will be necessary to ensure stable, majority rule outcomes. MWC legislatures find this additional structure in agenda rules.

In legislatures run by open-rule agendas – the agenda setter's proposal can be amended by a 51 percent majority of the legislature – power is diffuse, residing in all members of the legislature who are allowed to offer proposals.<sup>19</sup> Baron and Ferejohn (1989) offer a theory

18. A point made explicitly by James Wilson to the Pennsylvania ratifying convention for the new U.S. Constitution: "[w]hatever the object of government extends, in its operation, *beyond the boundaries* of a particular state, should be considered belonging to the government of the United States" (italics in the original; quoted in McConnell 1987, 1495).

19. In contrast, in legislatures run by closed-rule agendas, the agenda setter's proposal cannot be amended on the floor of the legislature. Here power rests with the select few who are given control over the agenda, for example, legislative committees (Shepsle

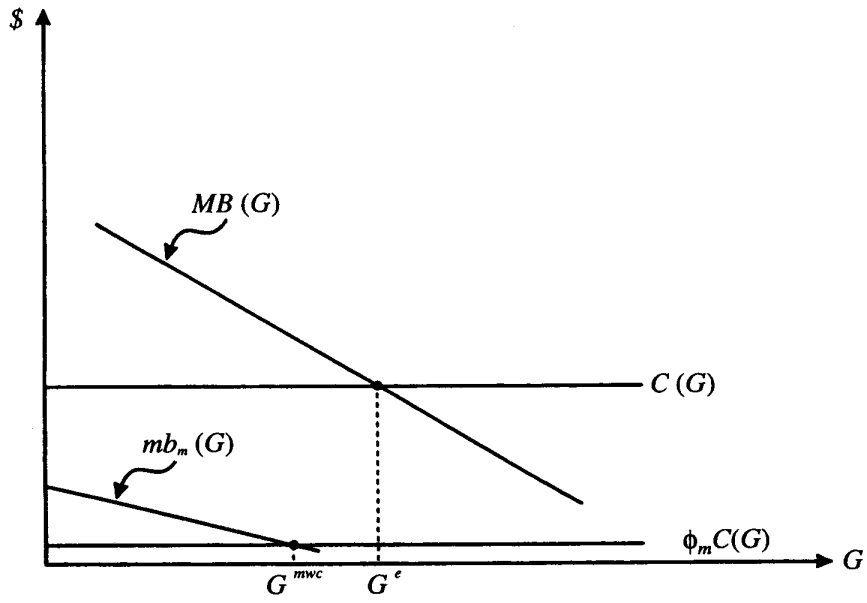
of legislative choice over government economic policies in open-rule legislatures. Only two rules apply: (1) amendments must be germane to the economic policy issue under consideration and (2) all members of the legislature have an equal probability of being recognized to offer a new, amended proposal. Baron (1993a) has characterized the equilibrium allocations for the provision of pure (national) public goods and for particularistic (local) goods by this central legislature run by open-rule agendas. Figures 1a and 1b illustrate the allocations and their likely consequences for economic efficiency.

In the case of a single pure public good ( $G$ ), the chosen policy is the one that satisfies the legislature's decisive median representative and his constituents. Figure 1a shows the resulting allocation at the point  $G^{mwc}$ , where the marginal benefit curve for the median representative's constituents – say, the median voter's marginal benefit of  $mb_m(G)$  – equals that representative's constituents' (e.g., median voter's) share ( $\phi_m$ ) of the social cost of producing the national public good:  $mb_m(G^{mwc}) = \phi_m C(G^{mwc})$ . The socially efficient allocation is at point  $G^e$ , where the social marginal benefit of the public good equals the good's social marginal cost:  $MB(G^e) = C(G^e)$ . Whether the national public good is overprovided, underprovided (as shown in Figure 1a), or set at the efficient level ( $G^{mwc} = G^e$ ) will depend on the distribution of citizen demands for the national public good and the distribution of tax burdens.<sup>20</sup>

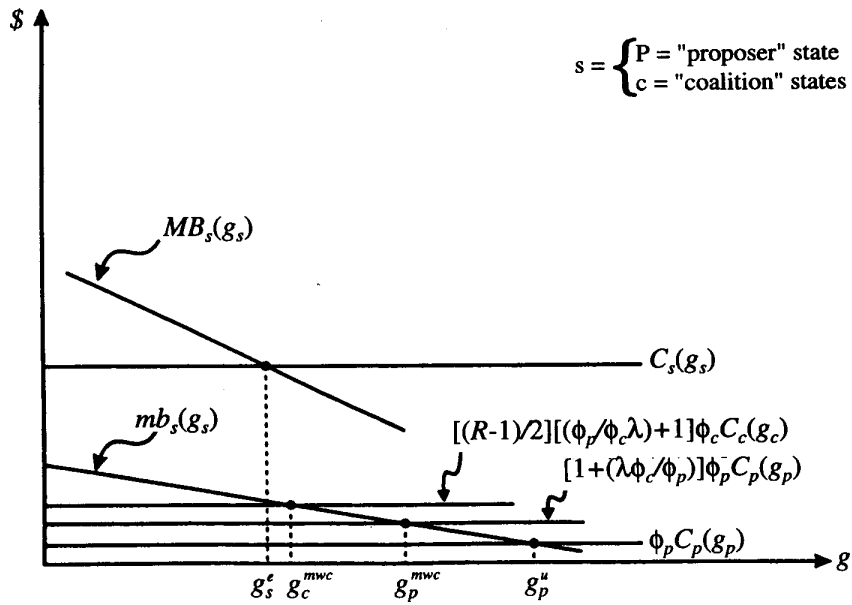
Central government legislatures are not limited to providing national public goods, however. Nothing prevents these legislators from using central government taxation to provide local, or "particularistic," public goods ( $g$ ) that benefit the residents of only one state or small group of city-states. Local public goods financed nationally are necessarily redistributive, as a national tax finances a good that benefits only a few residents.

1979) or the leadership of organized political parties (Cox and McCubbins 1993). What remains unexplained in these closed-rule models is the source of the agenda setter's monopoly power over policymaking, though recent efforts by Krehbiel (1992) and Cox and McCubbins (1993) are promising. We do not explore the federalist implications of their work here, though this is a promising avenue of future research.

20. This point was made originally by Bowen (1943). Define the relationship between the social marginal benefit and the decisive constituent's marginal benefit by the proportional relationship:  $\Psi_m MB(G) \equiv mb_m(G)$ . Note that  $\Psi_m MB(G) \equiv mb_m(G)$  by construction, and  $mb_m(G) = \phi_m C(G)$  by the political process. Thus,  $\Psi_m MB(G) \equiv mb_m(G) = \phi_m C(G)$  or  $(\Psi_m/\phi_m)MB(G) = C(G)$ . The demand share parameter for the decisive constituent ( $\Psi_m$ ) is defined by the distribution of citizens' demands for the national public good, while the cost share parameter for the decisive constituent ( $\phi_m$ ) is defined by national tax rates and the distribution of tax base. Public goods are overprovided ( $MB(G) < C(G)$ ), efficiently provided ( $MB(G) = C(G)$ ), or underprovided ( $MB(G) > C(G)$ ) as  $(\Psi_m/\phi_m) > 1$ ,  $= 1$ , or  $< 1$  – that is, as the decisive constituent's share ( $\Psi_m$ ) of national benefits exceeds, equals, or is less than his share ( $\phi_m$ ) of national costs.



1a. "National" public goods.



1b. "Local" public goods.

Figure 1. Economic policy in open-agenda MWC legislatures.



Redistributive public policies are susceptible to voting cycles. How will an open-rule MWC legislature allocate these goods? If chosen to propose a public policy, each legislator's first choice is to spend nothing on the particularistic goods of other states and only to spend on his own state's preferred project ( $g_p$ ). The first choice is  $g_p^u$  in Figure 1b, where the proposer's ( $p$ 's) constituents' marginal benefit from the local public good equals his constituents' marginal cost:  $mb_p(g_p^u) = \phi_p C_p(g_p^u)$ .

If offered, however, such a proposal is sure to lose. To carry the necessary 51 percent majority, a proposer must offer some benefit to legislators from one half of the other states. Baron (1993) shows that these benefits come as a smaller project for the proposer's own state at level  $g_p^{mwc}$  and as modestly sized projects for each of one-half of the other states in the proposer's coalition ( $c$ ) at levels  $g_c^{mwc}$ .<sup>21</sup> Those states not included in the proposer's coalition receive no local project from the central government.

Either inefficient oversupply or inefficient undersupply may result from the central government's provision of state particularistic public goods. As an example, assume that all states have identical marginal benefit and cost curves from the provision of  $g$ . Then, in Figure 1b,  $g_s^e$  is the efficient level of output of the local public good to state  $s$ , defined

21. The need to build a minimal winning coalition imposes an additional "shadow cost" on a proposer above the direct tax costs of  $\phi_p C_p(g_p)$  in Figure 1b. Since each dollar spent on the proposer's own project imposes a tax burden on others in the minimal winning coalition, they must be compensated through increased spending on their projects. This compensation to coalition members makes the proposer's own spending more costly, as his project's marginal costs now equal his project's direct costs of  $\phi_p C_p(g_p)$  plus the required coalition compensation at a rate  $\lambda$ . Thus, the full marginal costs of another unit of  $g_p$  are  $[1 + (\lambda\phi_c/\phi_p)]\phi_p C_p(g_p)$  shown in Figure 1b.

In a legislature of size  $R$  a support coalition of  $(R - 1)/2$  legislators is needed to pass the proposer's budget. To win this support the proposer must supply particularistic goods ( $g_c$ ) in the  $(R - 1)/2$  coalition states at a cost to the proposer of  $[(R - 1)/2]\phi_p C_c(g_c)$  in direct tax costs. (For simplicity Figure 1b assumes project costs  $C_s(g_s)$ , where  $s = c$  or  $p$ , are identical in all states.) This spending also imposes an added tax burden on a typical decisive citizen in each coalition state, but that burden is offset by the benefits received by that decisive citizen from the increase in  $g_c$ :  $[R - 1)/2]\phi_c C_c(g_c) - mb_c(g_c)$ . (For simplicity Figure 1b assumes project benefits  $MB_s(g_s)$  and  $mb_s(g_s)$ , where  $s = c$  or  $p$ , are identical in all states.) Again the proposer must offer compensation at rate  $\lambda$  for this added net burden on citizens in the coalition. A small increase in  $g_c$  costs the proposer the sum of these direct costs plus compensation:

$$[(R - 1)/2]\phi_p C_c(g_c) + \lambda\{[(R - 1)/2]\phi_c C_c(g_c) - mb_c(g_c)\}$$

or

$$\{[(R - 1)/2][(\phi_p/\phi_c\lambda) + 1]\phi_c C_c(g_c) - mb_c(g_c)\}.$$

Aggregate costs of coalition building to the proposer are minimized when this marginal cost of coalition building is equal to zero, or when:

$$mb_c(g_c) = [(R - 1)/2][(\phi_p/\phi_c\lambda) + 1]\phi_c C_c(g_c).$$

This is shown as project size  $g_c^{mwc}$  in Figure 1b.

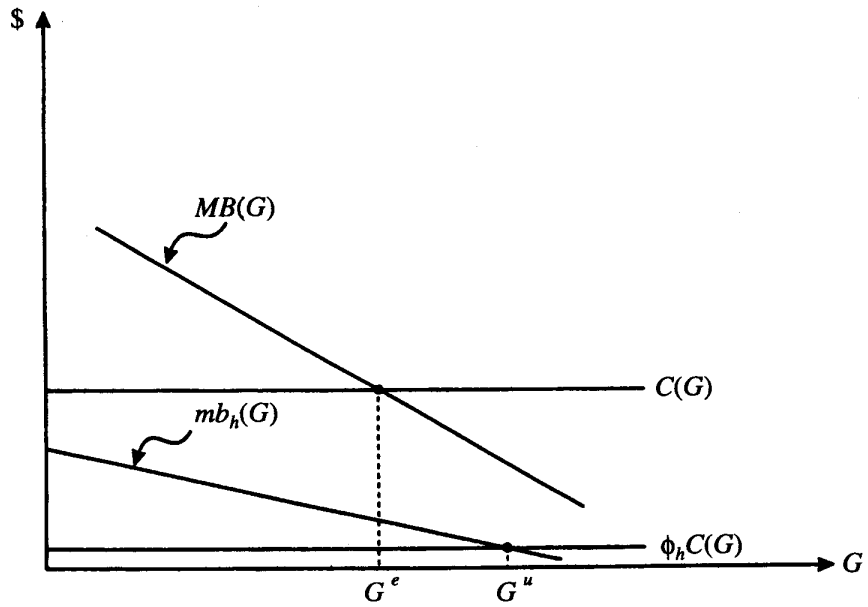
where the marginal benefit of the good to residents of the state equals the good's marginal production cost:  $MB(g_s^e) = C(g_s^e)$ . As shown, the proposer and all states in the winning coalition receive more than the efficient amount:  $g_s^e < g_c^{mwc} < g_p^{mwc}$ .<sup>22</sup> The states excluded from the 51 percent majority receive no spending on their local public goods from the central government. This latter inefficiency is reduced, however, if local supplementation to central government provision is allowed. With supplementation, the decisive voter in each excluded state now buys the local public good using local tax revenues. Thus, the states in the 49 percent minority receive their median voter allocations (which may or may not be efficient).

*The universalistic legislature.* While MWC legislatures overcome the difficulty of the voting cycle through agenda rules, universalistic legislatures control the cycle through the adherence to an informal norm of deference. Under the norm each legislator defers to the choices of all the other legislators. If any legislator or group of legislators fails to defer, the norm requires that all legislators penalize the defectors by denying their first choices. This norm of deference – “You scratch my back, I’ll scratch yours” – results in legislative proposals that are approved unanimously.<sup>23</sup> For this reason such legislatures are often called “universalistic” (U) legislatures. Empirical evidence suggests that policy allocations in U legislatures typically favor the high demander on each well-defined policy dimension (see Weingast and Marshall 1988, and Hall and Grofman 1990).

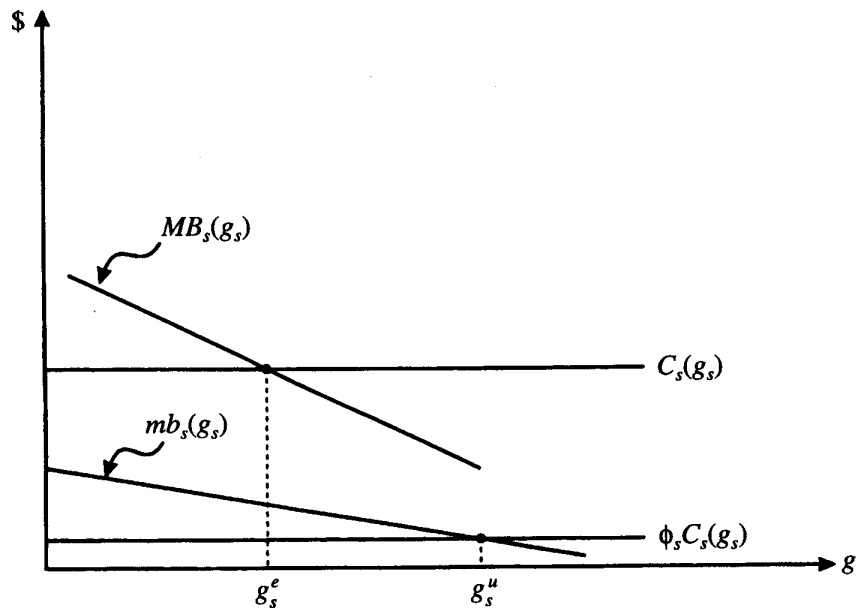
Figure 2a illustrates a universalistic legislature’s allocation of a national public good ( $G$ ). The downward sloping  $mb_h(G)$  schedule measures the marginal benefit from a national public good of size  $G$  to the decisive (e.g., median) voter in the high demand state. Similarly, the  $MB(G)$  schedule measures the aggregate social marginal benefit from the provision of the public good. The horizontal curve  $C(G)$  measures the marginal social cost of providing the project of size  $G$ , while the lower curve  $\phi_h C(G)$  measures the share ( $\phi_h$ ) of the cost born by the median voter in the decisive high demand state. Social efficiency is again defined by point  $G^e$  where  $MB(G^e) = C(G^e)$ . The predicted allocation for the universalistic legislature is at the high demander’s personally

22. Baron (1993a) shows this pattern of inefficiencies result for the case of constant marginal benefit curves and rising social marginal costs.

23. Weingast (1979) and Niou and Ordeshook (1985) have shown that legislative allocations under this norm of deference will be preferred by legislators to those expected from a MWC legislature with open agendas. Chari and Cole (1993) derive the legislative equilibrium under a norm of deference. Weingast and Marshall (1988) provide institutional details for how the norm of deference might be enforced.



2a. "National" public goods.



2b. "Local" public goods.

Figure 2. Economic policy in U legislatures.

preferred allocation, where  $mb_h(G) = \phi_h C(G)$  – point  $G^u$  in Figure 2a. Whether national public goods are overprovided (as shown in Figure 2a), underprovided, or efficiently provided depends upon the distribution of citizen demands for the national public good and the distribution of tax burdens, and the identity of the decisive representative.<sup>24</sup>

The allocation of particularistic public goods ( $g$ ) to individual states in a U legislature is shown in Figure 2b. The aggregate marginal benefit to all residents in state  $s$  is shown as  $MB_s(g_s)$ , while the benefit to the state's decisive (e.g., median) voter is given by  $mb_s(g_s)$ . The social marginal cost of the particularistic public good is  $C_s(g_s)$ , and the decisive voter's share of that cost when financed by national taxation is  $\phi_s C_s(g_s)$ . Social efficiency requires  $MB_s(g_s^e) = C_s(g_s^e)$ , given by the allocation  $g_s^e$ . The U legislature operating under a norm of deference will provide  $g_s^u$  in each state, where  $mb_s(g_s^u) = \phi_s C_s(g_s^u)$ .

Again, the efficiency of these allocations depends upon the distribution of citizen demands relative to the distribution of tax burdens. Since the median voter's share of state benefits from  $g_s$  is likely to be much larger than his or her share of the national cost of  $g_s$ , an overprovision of statewide public goods seems the most likely outcome – that is,  $g_s^u > g_s^e$  as shown in Figure 1b.<sup>25</sup> Inman (1988) and Inman and Fitts (1990) present evidence in support of this prediction.

*Federalism and representative legislatures.* The constitutional relationship between city-states and the central government in federalist republics can have significant consequences for the economic and political performance of the representative central legislature. Two dimensions of the constitution have been emphasized in the federalism literature as

24. National public goods are overprovided ( $MB(G) < C(G)$ ), efficiently provided ( $MB(G) = C(G)$ ), or underprovided ( $MB(G) > C(G)$ ) as  $(\Psi_h/\phi_h) > 1, = 1,$  or  $< 1$  – that is, as the high demander's share ( $\Psi_h$ ) of national benefits exceeds, equals, or is less than the high demander's share ( $\phi_h$ ) or national costs.
25. The efficiency analysis parallels that for national public goods. Local public goods are overprovided ( $MB_s(g) < C_s(g)$ ), efficiently provided ( $MB_s(g) = C_s(g)$ ), or underprovided ( $MB_s(g) > C_s(g)$ ) by the national legislature as  $(\Psi_s/\phi_s) > 1, = 1,$  or  $< 1$  – that is, as the state's median voter's share ( $\Psi_s$ ) of state benefits exceeds, equals, or is less than that median voter's share ( $\phi_s$ ) of national tax costs.

When drafting the U.S. Constitution the Anti-federalists noted this potential inefficiency when the "Impartial Examiner" argued that each state would prefer national to local taxation so as "to raise revenue by such means, as may appear least injurious to its own interest." Alexander Hamilton, in response, stressed that national taxation must be allocated for services that are "General not local; its operation extending in fact, or by possibility, throughout the Union, and not being confined to a particular spot" (Quoted by McConnell 1987, 1496–7).

particularly influential: the extent of *representation* of the city-states to the central legislature and the *assignment* of policy responsibility between the state-local and central levels of government (Rose-Ackerman 1981).

The representative structure ( $R$ ) of the central government may vary from a “town-meeting” format in which each of  $N$  city-states sends one representative to the central legislature ( $R = N$ ) to a “presidential” format in which all city-states vote in a single central election for one central representative ( $R = 1$ ). More typically, representative legislatures group city-states into subsets of legislative districts called “states” – say,  $n$  cities to one state – and then each of  $R (= N/n)$  legislative states sends one representative to the central legislature.

As the federalist constitution creates more legislative states, *representation* ( $R$ ) increases. The effects of changes in  $R$  on resource allocations in the central legislature depend on technology (a national or particularistic public good) and on how the legislature chooses to do business (U or MWC).

In U legislatures, an increase in  $R$  is likely to increase the level of aggregate spending by the central legislature on national and particularistic public goods. Spending on the national public good rises if an increase in  $R$  means a “finer” partition of representation of citizen demands for  $G$ , giving direct representation to a new, even higher, high demand coalition –  $mb_n(G)$  shifts upward in Figure 2a. Particularistic spending rises with  $R$  since more city-states now have direct representation in the legislature and each new state gets to choose a level of  $g_s = g_s^u$ . The hypothesized bias in U legislatures towards inefficiently high government spending (Figures 2a and 2b) is likely to increase as  $R$  increases.

In open-agenda MWC legislatures, increased representation has an uncertain effect on national and local public goods. Spending on national public goods changes only to the extent that increases in  $R$  alter the location of the decisive representative in the legislature. If the increase in  $R$  adds representatives equally to the low and high ends of the distribution of demand for  $G$ , then  $G^{mwc}$  will remain unchanged. Adding more (less) representation to the low end of the distribution will decrease (increase) the provision of the national public good by a MWC legislature.

For particularistic public goods, increasing  $R$  means each proposer must build a larger supporting coalition ( $(R - 1)/2$ ) to reach his majority; this increases aggregate spending on  $g_s$ . However, building such a coalition is more costly to the proposer, since his own taxes must rise to pay for increased spending. This reduces  $g_s$  for each state in the proposer’s coalition. On balance, therefore, aggregate spending and eco-

economic efficiency may rise or fall as  $R$  increases in open-agenda MWC legislatures.<sup>26</sup> We conjecture, however, that if the demand for local public goods is price inelastic – as the evidence seems to indicate (Inman 1979) – then the positive effect of larger coalitions on aggregate spending offsets the negative effect of higher costs of coalition building, and spending on local goods in MWC legislatures rises with  $R$ .

If representative central legislatures overspend on particularistic public goods, and if this overspending rises with increased representation of local interests in the legislature, then a compound federalist republic faces a trade-off between economic efficiency and the democratic advantages – enhanced participation, protection of rights, and developed civic virtues – that have been hypothesized by political theorists (Section 1) to follow from more extensive political representation. How might this trade-off be resolved? One answer is to look for additional federalist institutions that retain a level of representativeness while controlling economic inefficiencies. The constitutional *assignment* of economic tasks to the central or local levels of government is one such institution.

Specifying the assignment of central and local government responsibility for economic policy has been the primary concern of previous economic theories of federalism (see Oates's classic text, *Fiscal Federalism* 1972). When economic spillovers are significant and goods are national, small city-states are likely to provide those services inefficiently. If economic spillovers are absent and goods are local, then provision by large central legislatures is likely to be inefficient.<sup>27</sup> For economic efficiency, the assignment principle becomes: constitutionally assign national public goods to the central legislature and local public goods to city-states. Application of the principle requires us to define the extent of spillovers from government activities.

New evidence on spillovers, coupled with a growing appreciation for the inefficiencies of central legislatures, is raising important questions about how best to assign government activities within the federalist hierarchy. Old conclusions may no longer hold. Monetary, deficit, and

26. Adding to the uncertainty is the fact that changes in representation also may alter the location of the decisive voter in each state.

27. We have offered two theories – one each for the U and MWC legislatures – explaining why the central government will inefficiently provide local public goods. Oates (1972, 54–63) using the work of Tullock (1969) offers a different approach that assumes that the central government must provide the same level of a local public good to everyone in the nation. One needs a compelling reason to impose this equal provision constraint, however. Since the constraint is not technological – local goods can be provided at different levels in different areas – the constraint must be found in a behavioral model of legislatures.

redistribution policies, once considered the sole domain of the central government, are now seen as potentially valid local government activities (see Gramlich 1987 and Pauly 1973). Conversely, education policy, once seen as a task for only local governments, arguably now can be added to the national agenda as human capital externalities are identified (see Bénabou 1993). If the reach of a given activity's spillover becomes open to dispute, then application of the assignment principle becomes political.<sup>28</sup> If the central legislature assumes responsibility for deciding assignment, then assignment no longer stands as a feasible control to limit legislative inefficiencies. If assignment ceases to work as an independent federalist institution, then we must look for other institutions to "soften" the trade-off between the economic efficiency and the democratic advantages of representation.

In his study of federalism, Riker (1964, 101–3) argues for a strong executive to protect the nation's interest in economic efficiency. Through constitutional powers giving the nationally elected executive the right to veto and to execute legislative policies, the national interest in efficient policies can be protected against excessive local spending on particularistic public goods. Fitts and Inman (1992) model executive influence over universalistic legislatures and provide econometric evidence (Inman and Fitts 1990) that U.S. presidents who have used their executive powers effectively have controlled particularistic spending by congress.

The risk, of course, is that strong executive leadership dominates the central legislature in national policymaking. Rather than easing the trade-off between efficiency and democratic outcomes allowing us to achieve more of both, the strong, nationally elected executive may achieve its efficiency gains by sacrificing democratic rights and civic virtue. Only one (e.g., the national median) voice ( $R = 1$ ) is heard on any issue.

Riker's (1964, 91–101) solution is to add a second political institution to balance executive influence and to express local interests. Locally run political parties can be organized into effective majorities in the central legislature. Through their control of the legislature these parties can "block" policies of the powerful executive, unless at least a majority of those local interests are respected.

Sunstein (1988) offers yet another approach. His analysis seeks to strengthen the hand of local interests in executive actions directly by

28. The framers of the U.S. Constitution intended the Supreme Court to be the enforcer of the Constitution's assignment principle. The Court has found this an impossible task, however, and has essentially conceded the job to the political arena. See *Garcia v. San Antonio Metropolitan Transit Authority*, 469 US 528 (1985) and the discussion in Rapaczynski (1986).

opening the executive's administration of existing laws and the executive's deliberations on new laws to public hearings. More efficient public policies with increased representation – and all attending democratic virtues – might then result.

### 2.3 *Choosing a federalist constitution for a compound republic*

Two dimensions of constitutional design are uniquely federalist: (1) The creation of hierarchical levels of government – for example, central and state, state and local, or central, state, and local – each with separately *assigned* spheres of policy responsibility, and (2) the specification of *representation* for local interests within the central government, typically through a representative central legislature.

The assignment dimension allocates policy to the higher, central ( $A = 1$ ) or the lower, state or local levels ( $A = 0$ ), where the responsible level of government has the right to finance and execute a policy without fear of a unilateral “veto” by the other level of government. Representation ( $R$ ) is defined (for example) most simply by equally sized subsets of city-states forming states or provinces, where each state sends a single representative to the central legislature; thus,  $R = N/n$ , where the  $N$  city-states in the union are divided into  $R$  states of  $n$  communities each with  $1 \leq R \leq N$ . In the simplest case, the  $R$  states both provide assigned government services and send one representative to the central legislature to set central government policies. The federalist constitution defines  $A$  for each government activity and specifies a (usually, single) value of  $R$ .<sup>29</sup>

To specify a federalist constitution that is economically efficient, a careful detailing of the costs and benefits of alternative specifications of  $A$  and  $R$  is required. Breton and Scott (1978) were the first to systematically pursue the design of federalist constitutions from this strictly economic perspective. Their unique contribution was to add the transactions cost of government to the discussion of federalist constitutions. They identified four transactions costs that must be borne in any federalist structure: (1) the costs of revealing positions and selecting a representative for the central legislature (*revelation costs* =  $\alpha(R,A)$ ); (2) the costs of reaching agreements in the state and central legislatures (*decision costs* =  $\delta(R,A)$ ); (3) the costs of monitoring executive or bureaucratic implementation of

29. Typically, the assignment dimension lists those activities that are uniquely the responsibility of the central government (e.g., defense policy, commercial policy, trade, and foreign policy) and then specifies a residual rule such as the U.S. Constitution's Tenth Amendment. The representation dimension can be defined for a unicameral or bicameral legislatures. We assume a unicameral legislature for simplicity.



the legislature's decisions (*monitoring costs* =  $\sigma(R,A)$ ); and (4) the costs to citizens of moving from one state (or city) to another to achieve preferred policy outcomes (*moving costs* =  $\mu(R,A)$ ).

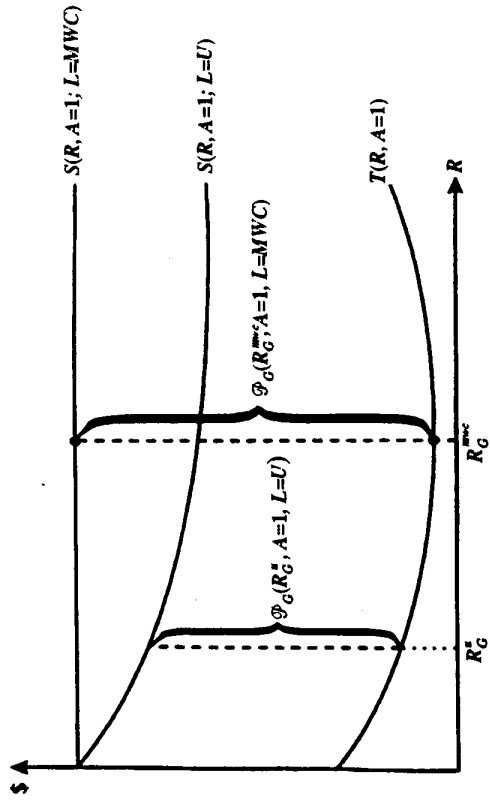
The total transactions cost of government is simply  $T(R,A) = \alpha(R,A) + \delta(R,A) + \sigma(R,A) + \mu(R,A)$ . While Breton and Scott focus on the effects on costs of variations in  $A$ , both  $A$  and  $R$  are likely to be important in a fully specified model of governmental transaction costs. Inman and Rubinfeld (1994) provide empirical evidence that each of Breton and Scott's separate transaction costs and therefore total costs,  $T(A,R)$ , are all likely to be U-shaped in  $R$  and, for each value of  $R$ , rising in  $A$ . Figures 3a–3d shows plausible specifications for  $T(R,A = 1)$  and  $T(R,A = 0)$ , respectively.

The economic benefit from a particular federalist structure is also shown in Figures 3a–3d, assuming an exogenous policy set of public goods and legislative institutions ( $L$ ). Benefits are measured by the aggregate consumer surplus ( $S$ ) generated from the provision of the public activity. Both representation ( $R$ ) to the central legislature and the assignment of responsibility for the elements of the policy set ( $A$ ) can affect the allocative performance of federalist governments. Thus,  $S = S(R,A; L)$ .<sup>30</sup> Consumer surplus will be maximized when the federalist structure provides the efficient level of national ( $G^e$ ) and particularistic ( $g^e$ ) public goods.

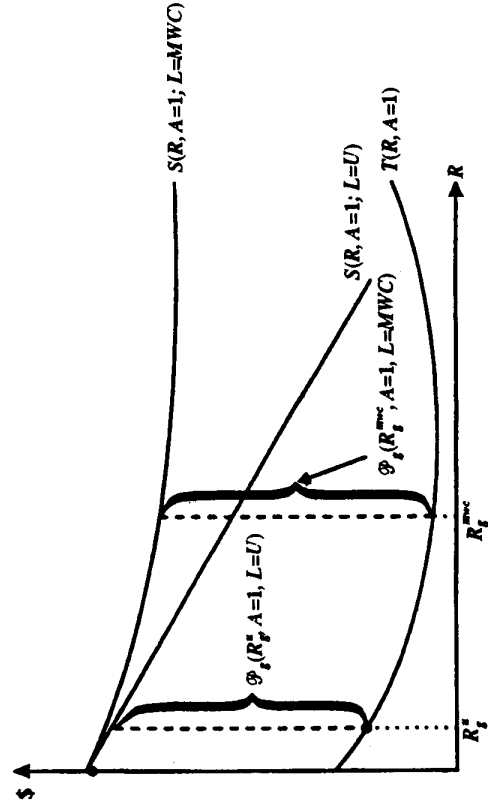
Figure 3a illustrates two plausible patterns of allocative efficiency for national public goods for the MWC and  $L$  legislatures when the central government has assignment responsibility for these goods ( $A = 1$ ). In the case of the  $L = MWC$  legislature, if increased representation leaves the national median in the legislature unaffected – new members equally represent positions from the left and the right of the median – then as  $R$  increases, allocations are fixed and allocative efficiency,  $S(\cdot)$ , is unaffected. In the case of the  $L = U$  legislature, increased representation brings in additional high demanders and spending on the national public good is likely to increase beyond  $G^e$ ; thus  $S(\cdot)$  declines as  $R$  rises.

Figure 3b illustrates  $S(R,A = 1; L = MWC$  or  $U)$  for particularistic public goods. As noted above (2.2.3),  $S(\cdot)$  is likely to decline for both the minimum winning coalition ( $L = MWC$ ) and universalism ( $L = U$ ) legislatures; we conjecture that the decline in efficiency as  $R$  increases is larger for the universalistic legislature.

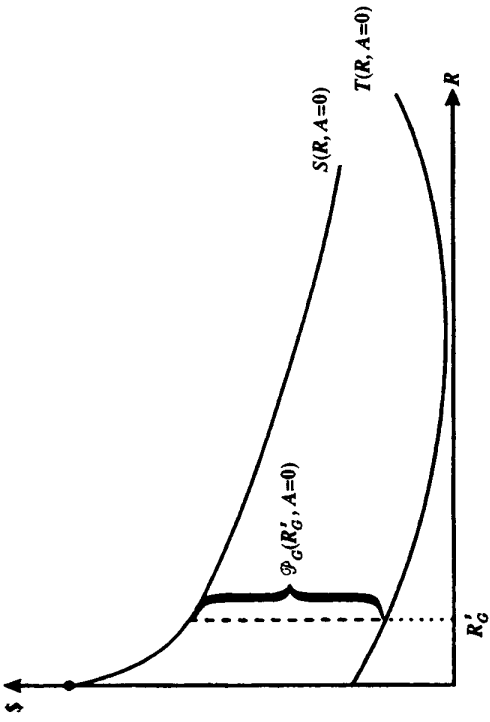
30. The policy set and legislative institutions can be set by the constitution or they can be left undefined. If left undefined, they become endogenous and functions of  $R$  and  $A$ . For one model of how the policy set might be determined, see Becker (1983). There is a vast literature on endogenous legislative institutions, beginning with Riker (1982a); the best recent survey is Krehbiel (1992).



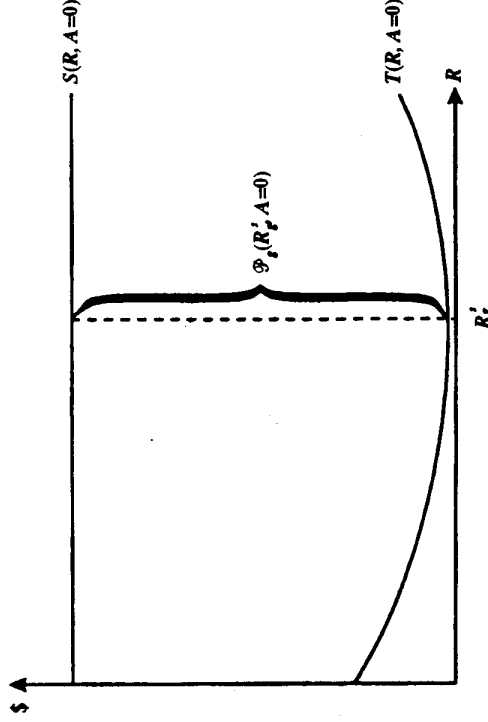
3a.  $A=1$  for "national" public goods.



3b.  $A=1$  for "local" public goods.



3c.  $A=0$  for "national" public goods.



3d.  $A=0$  for "local" public goods.

Figure 3. Representation and assignment.

Figures 3c and 3d show how economic efficiency might change when both national and particularistic public goods are assigned to the states only ( $A = 0$ ).<sup>31</sup> For national public goods, allocative efficiency is likely to decline when only states provide those goods, since free-rider problems discourage efficient interstate bargaining. This is shown in Figure 3c by the steep downward decline in  $S(R, A = 0)$  as more states ( $R$ ) encourage free-riding behavior (see Olson 1965).<sup>32</sup> State-only provision of particularistic public goods is likely to enhance economic efficiency, however (see Figure 3d). When states pay for their own state public goods, they balance the full marginal cost against state benefits; allocations closer to  $g_g^e$  are likely to result. Here increasing the number of states ( $R$ ) will have no effect on  $S$  if the benefits of the state good are fully internal to state residents; thus,  $S(R, A = 0)$  is horizontal in Figure 3d.

Overall economic efficiency in the compound republic is maximized when the gains from allocative efficiency  $S(R, A)$  less the transactions costs of running the government  $T(R, A)$ ,  $\mathcal{P}(R, A) = S(R, A) - T(R, A)$ , is maximized. In Figures 3a and 3b,  $\mathcal{P}(R, A = 1)$  is maximized at the levels of representation  $R_G^{mwc}$  and  $R_G^u$  and  $R_g^{mwc}$  and  $R_g^u$  respectively. In Figures 3c and 3d,  $\mathcal{P}(R, A = 0)$  is maximized at the levels  $R_G^s$  and  $R_g^s$  respectively.

When the central government provides public goods ( $A = 1$ ), the MWC legislature is more efficient. Importantly, the ability of MWC legislatures to control excessive government spending allows the public sector to be more representative ( $R_G^{mwc} > R_G^u$ ; see Figure 3a). When only the states provide national and particularistic public goods ( $A = 0$ ),  $R_G^s$  is likely to be small for national public goods to internalize spillovers. A comparison of Figures 3a and 3c for national public goods, shows that central provision is likely to be preferred to state provision, despite the potential allocative imperfections of central legislatures:

$$\begin{aligned} \mathcal{P}_G(R_G^{mwc}, A = 1; L = MWC) &> \mathcal{P}_G(R_G^u, A = 1; L = U) \\ &> \mathcal{P}_G(R_G^s, A = 0). \end{aligned}$$

31. Since the central legislature does not allocate the public goods, economic efficiency in this case will be independent of central government legislative institutions – that is,  $S = S(R, A)$ .
32. Bell (1989) presents a model of state provision of a national public good in which  $S(R, A = 0)$  has an inverted U shape. In Bell's model state governments are Leviathans capable of capturing all consumer surplus through lump-sum taxation. When  $R$  is small, individual states can exploit citizens. As  $R$  rises, state fiscal competition increases and this makes citizens better off. However, the increase in  $R$  means more states provide the national public good, which is inefficient. Bell shows that for plausible consumer preferences and spillover technologies a finite value of  $R$  greater than 1 is second-best efficient.

Thus, for national public goods,  $A = 1$  is the efficient assignment in this case.<sup>33</sup> A similar comparison of Figures 3b and 3d for particularistic public goods favors state provision:

$$\begin{aligned} \mathcal{P}_g(R_g^s, A = 0) &> \mathcal{P}_g(R_g^{mwc}, A = 1; L = MWC) \\ &> \mathcal{P}_g(R_g^u, A = 1; L = U). \end{aligned}$$

Thus, for particularistic public goods,  $A = 0$  is the efficient assignment.

With the possibility of efficient assignment, an efficient level of representation can be specified for each public goods technology. Ideally, we would select  $R_g^s$  from Figure 3d and, say,  $R_G^{mwc}$  from Figure 3a, allowing for a separate partitioning of the national population based on the goods to be allocated.<sup>34</sup> If only one partition is allowed so that each state must both provide state goods and select central legislators, then a single value of  $R$  must be chosen. This will be that value of  $R$  that maximizes  $\mathcal{P}_g(R, A = 0) + \mathcal{P}_G(R, A = 1; L = MWC \text{ or } U)$ . Finally, if restricting the domain of the central government through assignment is constitutionally difficult – as the U.S. Supreme Court has discovered (see Rapaczynski 1986) – then  $A = 1$  for both particularistic and national public goods and  $R$  has meaning only as it determines the number of central government legislators. In this case, the second-best efficient  $R$  is that which maximizes  $\mathcal{P}_g(R, A = 1; L = MWC \text{ or } U) + \mathcal{P}_G(R, A = 1; L = MWC \text{ or } U)$ .

While the analysis here clarifies how assignment and representation affect resource allocations in a federalist public economy, the particular conclusions from Figures 3c–3d are, of course, only illustrative. More detailed information about the determinants of the transactions costs of government and additional studies of how central and state legislatures allocate public sector resources are required before confident specifications of the efficient  $A$  and  $R$  are possible.

Even this information offers only a beginning to setting the federalist constitution. Legal scholars and new communitarians have stressed that important democratic values of protecting rights, encouraging participa-

33. This is not the only possible efficiency ordering of federalist institutions, however. For example:

$$\mathcal{P}_G(R_G^{mwc}, A = 1; L = MWC) > \mathcal{P}_G(R_G^s, A = 0) > \mathcal{P}_G(R_G^u, A = 1; L = U),$$

could well result. In this case, we should only centralize the provision of the national public good if we are confident the relatively more efficient MWC regime holds for the central legislature. Otherwise, we should decentralize the provision of the national public good and set  $A = 0$ .

34. Here the recommendation of separate governments for separate technologies is only partly due to the extent of spillovers. The ability of legislatures to efficiently allocate goods and services given the technologies also matters. This analysis extends the usual spillover-only model (Olson 1969) to accommodate the efficiency effects of political decision making.

tion, and developing civic virtues may be affected by the choice of representation and assignment. Economic equity may weigh in the balance too. Important trade-offs between these values and economic efficiency may emerge as representation and assignment impact jointly on all valued outcomes. If so, hard choices must be made (see Cooter 1987b or Hausman and McPherson 1993). Finally, as economic and political circumstances change over time the preferred federalist structure may change as well. A federalist constitution must be a flexible document capable of responding to new economic and social orders (Mueller 1973).

### **3. Stability of the federalist constitution**

Having chosen a federalist constitution by specifying values of  $R$  and  $A$ , the question remains: Will that constitution survive the pressures of domestic politics, or will one or more city-states opt to secede from the federalist union? Stability is an issue for both the confederate republic of Montesquieu and for the compound republic of Madison.

#### *3.1 Stability of the confederate republic*

Why should a confederate republic that requires unanimity of all participating city-states for any central government action be concerned with secession? One answer is the frustration individual city-states might feel with the process of Coasian bargaining, the cornerstone of public decision making in the compound republic. For reasons noted in Section 2.3, unanimous agreements between city-states, particularly when  $N$  is large, may be impossible. Yet even when there is collective agreement as to how to improve the collective welfare of the member city-states, disagreements are sure to arise over the distribution of the resulting economic surplus. City-states that consistently get the smallest share of the benefits from collective action because of high taxes or low services may consider the option of leaving the union. Exit will occur and the confederate republic will be unstable, when the benefits of exit to one or more members are greater than the costs.

Formal models of exit from federal confederate republics have been developed by Buchanan and Faith (1987), Wei (1991), Austin (1994a), and Berkowitz (1994). The models clarify the likely sources of instability. Exit occurs when (1) the long-term fiscal exploitation of a city-state by the union is substantial; (2) the city-state offers few nonfiscal benefits to others in the union; (3) the city-state itself receives few nonfiscal benefits from the union; and, (4) the city-state has a comparative advan-

tage in defending its territory against military or economic (trade war) attack.<sup>35</sup>

Wei (1991) uses his model of exit to explain the decline of the Soviet Republic in which the richer states (Russia, Ukraine, Belorussia, Lithuania, Latvia, and Estonia) have the strongest incentive to exit from a union. Austin (1994b) studied the voting behavior by republics in the All-Union Referendum of March 1991 and found that republics enjoying the greatest net economic benefits from the Union were the most likely to support the Union. Litvack (1994) shows that the same difficulties remain within the new Russian Republic, since the regions rich in natural resources (Bashkortostan, Tatarstan, and Yakutia) have withheld taxes, ignored export regulations, and threatened independence.

Wei (1991) notes that the fiscal pressures for secession present in the old Soviet Union also exist within China today. In China, however, the central government has acted by reducing fiscal transfers from the wealthy south to the poorer regions of the north and west. Africa, too, experimented with the confederate republic model of federalism when Tanganyika, Kenya, and Uganda joined together in the 1960s to form the East Africa Federation. Coasian bargains proved difficult, however, when the parties could not agree on the division of the surplus from

35. Buchanan and Faith (1987), Wei (1991), and Berkowitz (1994) consider the case of just two players, the potential seceder and the remainder of the polity. The candidate for exit balances the benefits to residents if the city-state exits the union (the saving in fiscal exploitation [ $F$ ] less any nonfiscal benefits such as access to inexpensive natural resources [ $T$ ]) against the cost of exit.

Costs are incurred to withstand an "attack" by the rest of the polity seeking to hold the city-state within the union. The remainder of the polity will spend up to  $A$  dollars to attack, where  $A \leq$  the value to the polity of holding the potential seceder within the union. That value equals the noneconomic benefits of having the seceder in the union ( $V$ ) plus the net fiscal surplus earned from the seceder ( $F$ ). Thus,  $A \leq V + F$ . These  $A$  dollars can buy  $\alpha$  units of attack, where  $\alpha = A/p_\alpha$  and  $p_\alpha$  is the cost of the attack. To be successful, the potential exiter must buy at least  $\delta (\geq \alpha)$  units of defense. Defense  $\delta$  can be bought by spending  $D = p_\delta \delta$  dollars. Thus, the successful exiter must spend at least  $D = (p_\delta/p_\alpha)(V + F)$  dollars to leave the union.

Finally, exit will occur if benefits are equal to or greater than costs, or if  $B = F - T \geq (p_\delta/p_\alpha)(V + F) = D$  or when

$$F[1 - (p_\delta/p_\alpha)] \geq (p_\delta/p_\alpha)V + T.$$

Exit occurs when fiscal exploitation ( $F$ ) is large, the seceder offers few nonfiscal benefits to others in the union ( $V$  low), gets little in return ( $T$  low), and has a comparative advantage in defending its territory ( $p_\delta/p_\alpha$  is small).

These two-player models reveal the central forces which lead to instability, but they are inadequate for predicting whether a confederate republic will actually unravel. More complicated models of coalition formation and stability with  $N \geq 3$  city-states will be needed; see Austin (1994a) for a start.

shared facilities – sugar production plants (Leys and Robson 1965). The union dissolved over this and other distribution issues.<sup>36</sup>

### 3.2 *Stability in the compound republic*

The compound republic overcomes the difficulties of reaching unanimous agreements in confederacies by requiring simple majorities of representatives from the city-states – not unanimity – for its decisions. Although decision making by majority rule helps ease central government inaction, it increases the likelihood of fiscal exploitation of minorities by majorities. The economic pressures to exit noted above should be even stronger here than in the confederate union run by unanimity because each city-state loses the protection of its legal veto over redistributions. Perhaps the most telling example of attempted exit from a compound republic is the U.S. Civil War.

There is another, more subtle threat, to the federalist constitution in the compound republic, however. Riker (1964) has called it the “overawing” of the states by the central government. Effective federalism must give some responsibilities to the  $R (\leq N)$  states and protect that agenda through the assignment dimension of the constitution (recall that  $A = 0$  for those activities assigned to the states). Riker (1964, chap. 4) examines forces within the republic which act to undo this assignment. Inman (1988) has documented the strong drift toward central government financing and regulation of state and local government activities in the United States over the last fifty years. One major cause has been the shift at the central government level from a MWC legislature to a U legislature. The temptation of locally elected legislators to use central government financing to share the costs of state and local services is present in a compound republic, and under universalistic legislative structures those legislators are not sufficiently constrained. Although the constitution says  $A = 0$ , political realities give  $A = 1$  for many goods and services.

What can be done to protect the integrity of assignment? Riker (1964, 102–3) is skeptical that the judicial branch can effectively enforce assignments. The recent experiences of the U.S. courts provides ample evidence for his concern (see Rapaczynski 1986 and Merritt 1988). For Riker (1964, chap. 4) the national political process itself is the only effective check.

36. There is an extensive historical and political science literature on secession and the formation of nations of which the federalism literature is only a part; see Bookman (1992) and Hobsbawm (1990) for an introduction to these literatures.

To protect assignments, three conditions must hold. First, the desired assignments to the states must be clearly understood and supported in the populace as a whole. Here the federalist constitution, and perhaps the judiciary, plays a crucial role. Second, information about violations of assignment and their consequences must be available to citizens. Here economic analyses of the efficient use of local, state, and federal governments are important (Rivlin 1992). Finally, political institutions that are capable of restoring the assignment balance must be established outside the centralizing forces of government.

Studying the explosion of U.S. central government activity following the 1930s depression, Riker (1964, 91–101) focused on executive branch politics as the cause of central government influence over the states. He looked to strong political parties rooted in local politics to protect state assignment against executive interventions.<sup>37</sup> In contrast, using data from a second period of central government growth in the 1960s and 1970s, Inman (1988) and Fitts and Inman (1992) found the centralizing forces to lie within a locally dominated national legislature, and they advocate nationally elected executives to internalize the inefficiencies of universalism.

Weingast (1993) has brought a broader historical perspective to the question. Weingast finds, contrary to recent U.S. experience, that an independent judiciary establishing local political authority was the key to protecting seventeenth-century England's federalist structure. His analysis of nineteenth-century U.S. federalism stresses, like Riker's analysis of contemporary politics, the importance of strong national political parties grounded in a commitment to limited federal government.

This brief historical record makes clear there is no unique political institution certain to protect assignment in the compound republic. Yet in each of these examples when a political institution did rise to the defense of the federalist constitution – whether it was the executive, the judiciary, or a legislative political party – a single truth prevailed: the vast majority of the citizens in that society believed the federalist constitution to be worth protecting (Elazar 1993).

#### **4. Conclusion**

Originally the central issue in the writing of the United States Constitution in 1787, the design of federalist political institutions has once again emerged as crucial for many of the world's economies. The emerging

37. A view currently shared by the U.S. Supreme Court in their decision to place the protection of state activities within the legislatures of the central government; see Rapaczynski (1986) and Merritt (1988).



economies in the Russian Republic, Eastern Europe, and South Africa are each struggling to define an appropriate federalist constitution to guide their public choices. The established and stable economies of China, the United States, and Western Europe are searching for new federalist structures for their existing public sectors.

This survey reveals two important dimensions for any federalist constitution: local representation to the central government ( $R$ ) and assignment ( $A$ ) of responsibility. Two objectives of any federalist constitution are also identified: protecting rights/promoting civic virtue and achieving economic efficiency. Unfortunately, choices of  $R$  and  $A$  may not permit the ideal on each objective. Our review of the Coasian bargaining and political economy literatures suggests high values of  $R$  and  $A = 0$  thought to enhance democratic rights and civic virtues may discourage public sector efficiency. Conversely, low values of  $R$  and values of  $A = 1$  may enhance efficiency but might diminish citizen representation in government choices. Federalist constitutions must make hard choices and be responsive to the potentially changing economic and political conditions that lay behind those choices.

We reviewed the arguments for two federalist constitutions that come down on opposite sides of the trade-off between democratic rights and virtues, and economic efficiency. Montesquieu strongly favors the goal of citizen representation and thus his confederate constitution is strongly decentralized, setting  $R = N$  and  $A = 0$ . Madison, on the other hand, sought a balance between efficiency and representation. His more centralized compound republic sets  $R < N$  and  $A = 1$  for many important public activities.

Which of these federalist constitutions is best? We cannot resolve this issue here. We hope, however, that our survey has identified the important questions to be answered before this choice is made. First, how do representation and assignment affect the important political values of participation, protection of individual rights, and the development of civic virtues? Second, how do representation and assignment affect the allocation of goods and services, and thus economic efficiency? Third, knowing how representation and assignment affect rights and efficiency, what combination of these important values do we desire? Fourth, having chosen a constitution, what safeguards can be created to ensure it survives the pressures of daily politics?

Finding answers to these questions will require careful analyses from a full range of scholarly talents – political science, economics, history, philosophy, and law. Each discipline has already made important contributions. We hope this review encourages new research and facilitates a dialogue among scholars working on the federalist question.