Federalism and the Democratic Transition: Lessons from South Africa

By Robert P. Inman and Daniel L. Rubinfeld*

What path should a country follow in its transition from autocracy to democracy? The most recent and perhaps one of the more successful democratic transitions occurred in South Africa with the election of Nelson Mandela as President in April 1994. This paper argues that the federal form of governance, which allowed for locally elected provincial governments with significant fiscal policy responsibilities, was essential for this successful transition. An appropriately structured federal constitution can allow for democratic rule by the (often poor) majority, while protecting to a significant degree the economic interests of the previous ruling (usually rich) minority.

A federal constitution must resolve three questions:

1) The *Provincial* Question: Will there be more than one state or province, and if so, how many?

2) The *Assignment* Question: What powers, if any, should be assigned to the provinces?

3) The *Representation* Question: Will provinces be allowed to elect their own leadership independent of the central government, and will this leadership be represented at the center?

In the country's final constitution approved by the Constitutional Assembly on 11 October 1996, South Africa answered yes to all three questions and did so in a way that allowed both the original ruling National Party (NP) elite and the newly enfranchised African National Congress (ANC) majority to prefer democratic federalism to both a unitary, ANC-dominated democracy and to the original autocratic, NP-dominated apartheid regime.

I. Federalism and the Democratic Transition

While this analysis is motivated by the circumstances behind the South African transition, we view the argument as potentially quite general; we will focus therefore on the generic properties of a "typical" autocratic society contemplating a democratic transition. For the initially oppressed majority (the ANC in South Africa), either a unitary democracy or a federal democracy will be preferred to current autocratic rule. For the original ruling elite (South Africa's NP), we assume that maintaining the current autocratic regime is less desirable than a federally based democratic regime, but preferable to a unitary democratic society. It is this ranking by the elite (*federal democracy* > *autocracy* > *unitary democracy*) that creates the democratic impasse. We argue below that, if certain economic conditions hold and the federal constitution is appropriately designed, then the new ruling majority can be induced to most prefer a federal democracy, ranking regimes as *federal democracy* > *unitary democracy* > *autocracy*. If this ranking cannot be achieved and the ranking *unitary democracy* > *federal democracy* > *autocracy* is the only feasible ranking for the majority, then the majority can never credibly commit to a federal regime, the current elite will never agree to a democratic constitution, and the

† *Discussants:* Daniel Treisman, University of California; Gérard Roland, University of California–Berkeley.

* Inman: Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104-6367; Rubinfeld: School of Law and Department of Economics, University of California, Berkeley, CA 94720. From 1994 to 2000, Inman and Rubinfeld (with Richard Bird and Charles McLure) served as economic advisors to the Financial and Fiscal Commission, which was charged with designing the fiscal responsibilities of, and financing for, the nine newly created South African provinces. We owe a debt of gratitude to Junaid Ahmad of the World Bank, who facilitated our participation.

democratic transition will not occur.[1] But if an appropriately designed federal democracy is preferred by both parties, a transition can occur. When is this the case?

Once in the democratic state, the private economy will pay a wage w to each adult citizen in the once oppressed majority, and net a return on human and physical capital of π to each member of original ruling elite: $\pi > w$. In the democratic state, the elite's net return will also depend on the state of political unrest in their immediate local economy. The elite citizen earns $\pi(0)$ when the local economy is at peace and $\pi(1) < \pi(0)$ when the local economy suffers from political unrest, where the difference equals the cost to the elite of local political unrest.

In the democratic state, the public economy will impose a tax on the elite of τ per person, with revenues used to fund social services for the poor majority. Those in the elite may choose to stay within the country and pay the tax or to leave the country with their human and physical capital. The number of elite residents N who choose to remain within the country is given by $N = N(\tau)$, where $dN/d\tau < 0$. Total revenues available for social services is $R = \tau N(\tau)$, where maximal revenues, the peak of the revenue hill, is defined by the tax rate τ_c and $R_c = \tau_c N(\tau_c)$. There are M ($>N$) citizens who are members of the poor majority, all of whom are eligible to receive the social-service grant. The grant (which does not affect in- or out-migration by the poor) is distributed equally to all members of the majority, independent of location (or province if a federal constitution applies): $g = \tau N(\tau)/M$. The maximal social-service grant is defined by the maximum level of elite tax revenues: $g_c = \tau_c N(\tau_c)/M$. The objective of the new majority will be to maximize social-service payments to their constituents, conditional on creating a sustainable democracy. If a unitary democracy is sustainable, the maximal grant will be g_c. A sustainable federal democracy will only support a smaller grant, however, at some level g_f funded by a tax on the elite of τ_f: $g_f = \tau_f N(\tau_f)/M$, where $g_f < g_c$ as $\tau_f < \tau_c$.

With respect to the administration of the social-service grant, the pre-democracy elite has a potentially significant cost advantage over the new ruling majority. Grants to the majority are primarily to provide education and health-care services. If elite service providers manage the provision of the social-service grant, then the cost of providing the services to each majority recipient will be b_f per recipient. The net benefits received by each recipient will be $v_f = g - b_f$. If less well-trained or perhaps corrupt professionals provide the services, the cost to each majority recipient will be b_c per recipient, and net benefits will be $v_c = g - b_c$. We assume $b_c > b_f$. We also assume that, if the new government is organized as a unitary democracy, the elite providers shirk, and service costs will be b_c for all grant recipients. If government is organized as a federal democracy, however, then in any province controlled by the elite, the elite providers act efficiently and provide services to any majority citizens in their province at a cost of b_f. One can imagine in such provinces that public services are provided through "integrated" classrooms, hospitals, and clinics. In provinces where the elite do not have control, the province cannot attract elite providers, and the cost of services will be b_c.

A constitution establishing a unitary democracy will create a single central government dominated by the newly enfranchised majority. If sustainable, a unitary democracy will tax each member of the old ruling elite at the revenue-maximizing rate of τ_c, provide a social grant of g_c, incur administrative costs of b_c, and offer each member of the majority a net benefit of $v_c = g_c - b_c$.

A constitution establishing a federal democracy will have a central government and at least two provinces. One of these must be controlled by a majority from the elite, with the other province(s) assumed to be controlled by local, nonelite majorities. Taxation (τ) and funding for social-services grants (g) are assigned to the (newly enfranchised) majority-controlled central government. Without central financing for social services, federalism is not credible as a

[1] The oppressed majority (through armed insurgency) or outside nations (through tighter economic sanctions) can "up the ante" in the autocratic regime and shift the elite preferences to *federal democracy > unitary democracy > autocracy*. If so, the elite will prefer unitary democracy and a unitary constitution will be written. ANC leaders rejected continued military resistance as too costly, preferring the federal democratic alternative as a constitutional strategy (Patti Waldmeir, 1997 p. 165).

transition path. With provincial-only taxation, either the poor in the elite province risk receiving no services at all (de jure apartheid) or the elite province is devoid of poor citizens and becomes a fiscal refuge for the elite (de facto apartheid). Either way, the once oppressed majority will reject this federal assignment as a continuation of economic autocracy.[2] Finally, the responsibility for grants expenditure, and thus service provision, must be assigned to provincial governments.

To ensure that such a federal constitution is sustainable in the long run, and thus credible at the constitutional stage when the elite yields power to the newly enfranchised majority, four constraints must be met. First, the federal constitution must ensure a long-run, equilibrium tax rate τ_f on the ruling elite that is less than the maximal tax rate τ_c that would be chosen under a centralized unitary democracy. Without credible assurances that $\tau_f < \tau_c$, federalism becomes a hollow promise, unitary democracy becomes the de facto constitution, and the elite blocks the transition to democracy. Since $\tau_f < \tau_c$, the first constraint requires that $\tau_f N(\tau_f)/M = g_f < g_c = \tau_c N(\tau_c)/M$. Second, though $g_f < g_c$, g_f must be large enough that the long-run fiscal equilibrium under federalism is sufficiently attractive to the new majority that the majority prefers this alternative to a long-run regime of unitary (centralized) democracy. Since g_f cannot be 0, the second constraint sets a minimum for social-services grants of $g_f^{\min}$. Together, the first two constitutional constraints set the range for feasible assignment of spending responsibilities as $g_f^{\min} < g_f < g_c$.

There will always be a temptation in any one fiscal year for the majority to defect from the federal constitutional agreement and choose the maximal tax rate τ_c. To discourage defection, the elite needs a credible punishment strategy, which is available through its control of provincial social-service spending. If the central government chooses τ_c, then the elite province allocates all of the assigned grants revenues (g_c) to elite residents ("full fungibility"), leaving the newly enfranchised residents of the elite province with no social services. For this punishment strategy to be effective, there must be enough newly enfranchised residents ("hostages") in the elite province so that central-government defection is discouraged. The elite must have an incentive to punish, and this means there must be enough grant money at stake to make full fungibility attractive against the likely costs of political unrest, $\pi(0) - \pi(1)$. From the perspective of the majority, injuring a small number of poor residents in the elite province may be a small price to pay for increasing the grant to the rest of the new majority from g_f to g_c, but if the harm is widespread, the cost may be too high. The elite province must therefore have enough newly enfranchised residents ($m_e^{\min}$) for punishment to be effective, but not so many as to actually be a political majority; thus, $N \geq m_e \geq m_e^{\min}$. These two constitutional constraints define the required population size and geography of the elite province.

When these four federalism constraints are met, annual fiscal politics become a repeat-play prisoner's dilemma game in which the cooperative federal allocation is preferred to the noncooperative unitary allocation in the long run. The noncooperative unitary allocation in which both the national ruling majority and the elite province leadership defect from the federal compact (the central majority by choosing τ_c and g_c, and the provincial elite by adopting full fungibility) is likely to be inferior to the cooperative federal allocation for either of two reasons. First, if the central government allows the elite province to reallocate social-service grants from the poor to the elite, political unrest will follow at a social cost of $\pi(0) - \pi(1)$. Second, if the central government does not allow the elite province to reallocate grants, perhaps by imposing a centrally administered "receivership," then there will be additional administrative or corruption costs of $b_c - b_f$ per poor recipient. Either of these added costs make that "joint-defect" allocation socially inferior to the cooperative federal allocation.[3] Fortunately, when played as a repeated prisoner's dilemma game from one fiscal year to the next, the cooperative allocation can be enforced by a

[2] As the ANC did during South Africa's constitutional negotiations (see Waldmeir, 1997 p. 241).

[3] What is not included as a cost in this basic model is the inefficiency of high tax rates associated with elite worker productivity. In our specification, neither $\pi(0)$ nor $\pi(1)$ depends upon tax rates.

"grim" trigger strategy, wherein each party threatens its defect strategy forever if the other party defects (see Eric Rasmusen, 1989 pp. 91–94).[4]

The resulting federal constitution defines a credible long-run federal equilibrium. Further, the four constraints define the constitutional attributes required of any federal constitution: provincial size, assignment, and local political powers. Finally, as a credible long-run constitutional agreement, such an appropriately specified federal constitution allows the transition from an elite autocracy to a majority-rule democracy. South Africa's constitutional negotiations stand as a case in point.

II. Federalism and South Africa's Move to Democracy

The official beginning of South Africa's journey from autocratic white party rule to a multirace, majority-rule democracy was 2 February 1990, when President F. W. de Klerk released Nelson Mandela from jail and officially legalized the political activities of the ANC and the South African Communist Party. For the next three years, representatives of the NP, the ANC, and (off and on) Chief Buthelezi's Inkatha Freedom Party (IFP) met to negotiate the new South African Constitution. The central issue was minority rights (particularly economic rights) versus majority rule (see Waldmeir, 1997 pp. 79, 195). Initial negotiations were difficult, but on 18 November 1993, an interim constitution was approved. The heart of the agreement was a constitutional commitment to majority rule at all levels of government to satisfy the ANC and, to satisfy the minority parties, nine constitutionally created federal provinces with projected NP majorities in two (Western Cape and Northern Cape), an IFP majority in another (Kwazulu-Natal), and finally, an unspecified "handshake agreement" that those provinces would have real fiscal powers (Waldmeir, 1997 p. 228).

In drawing the provincial boundaries, a clear effort was made to ensure political control for minority economic and ethnic interests: white landowners in the Northern Cape, the white and Asian business and professional classes in the Western Cape, and the Zulu nation in Kwazulu-Natal (see Yvonne Muthien and Meshack Khosa, 1998). In the case of the Northern Cape, the lines were too closely cut. The NP underestimated their support from their once favored "coloured" population and the voter turn-out among illiterate white farmers. As a consequence, the province narrowly went ANC in the first elections. The NP won the Western Cape, and the IFP won Kwazulu-Natal, however, and the ANC opposition continues as majorities in both provinces today. The provincial boundary/population constraints for sustainable federalism were met.

To formalize the handshake on fiscal assignment, the interim constitution established an 18-member Financial and Fiscal Commission (FFC), with one member approved by the NP (or IFP for Kwa-Zulu Natal) and one approved by the ANC from each of the nine provinces. The appointees were required by the interim constitution to be experts in finance, economics, accounting, or public administration. From the beginning, the commission was committed to a central tax rate for provincial grants that was well below any plausible estimate of South Africa's revenue maximizing rate ($\tau_f < \tau_c$). The ANC-controlled central Finance Ministry had set the target tax burden on personal income (τ_f) at 40 percent, and the commission decided that any funding for provincial grants must remain within that target and be offset by an equal reduction in revenue applied to other national expenditures (e.g., defense, infrastructure, police). Each year a division of tax revenues between the national and provincial governments, known as the vertical division, was to be recommended by the FFC to the National Assembly after consultation with the Finance Ministry. Next, the commission established the provinces as the primary providers of K–12 education, health services, public housing, public transportation, and welfare services. The final recommended provincial grant averaged 1,600 Rand per resident in 1995 Rands (about $500) and accounted for approximately 50 percent of all government taxation and 20 percent of per-resident personal income (see Financial and Fiscal Commission, 1995). In an expanded version

[4] This is an example of what is needed for an "equilibrium social contract," the analysis of which first appeared (to our knowledge) in Rui de Figueiredo and Barry Weingast (1997) and in Robert Gibbons and Andrew Rutten (1998).

of this paper (in process), we argue that the commission's recommended grant is likely to satisfy, but just barely, the two assignment constraints for sustainable federalism; we estimate that $g_c \approx 2{,}100$ Rand $> 1{,}600$ Rand $\geq 1{,}565$ Rand $\approx g_f^{min}$.

Finally, it was well recognized by the primary negotiators for the NP and the ANC that the federalism agreement, embodied in the FFC proposals for provincial assignment and grants, could not be enforced by a constitutional court (see Waldmeir, 1997 pp. 228–30). The FFC proposal was a statutory, not constitutional, arrangement. Further, a two-thirds majority for the ANC in the National Assembly (and they are close) can redraw provincial boundaries. In the end, South Africa's federalism compact must be self-enforcing. At the moment it appears to be; both sides have credible punishment strategies to check defection, the ANC with its ability to impose τ_c or to encourage costly local protests and the original NP elite with its ability to adopt local full fungibility or to withhold more efficient social-service providers. To date provincial governments have remained significant, though perhaps fragilely positioned, players within South African public finance.

III. Concluding Remarks

Having played a role in designing the South African federal agreement, what seems most striking to us is how different the final fiscal arrangements are from those that originally came to mind, having been well educated in the Tiebout-Musgrave-Oates tradition. In contrast to the usual *efficiency* principles, as outlined in Inman and Rubinfeld (1997), the *transition* principles for sustainable federalism described here fully centralize the financing of what are typically seen as local services and decentralize spending decisions for important redistributive services. For the task of moving from autocracy to democracy, this seems just right. To escape autocracy it is essential to provide equitable public services for the new majority and to ensure the property and political rights for the once-ruling minority. In this light, our transition principles of federalism might be best seen as addressing a new democracy's "first-order" problems. The more familiar efficiency principles, developed as they were for established and secure democracies, are more appropriate for democracy's "second-order" agenda of growth and efficiency. Only as other democratic institutions mature (in particular, independent courts and opposition parties capable of winning national elections) can the institutions of federalism be directed toward the more familiar agenda of ensuring an efficient public sector.

REFERENCES

de Figueiredo, Rui and Weingast, Barry. "Self-Enforcing Federalism: Solving the Two Fundamental Dilemmas." Mimeo, Hoover Institution, Stanford University, 1997.

Financial and Fiscal Commission. "The Allocation of Financial Resources between the National and Provincial Governments: Recommendations for the Fiscal Year 1996/97." Mimeo, Financial and Fiscal Commission, South Africa, 8 September 1995.

Gibbons, Robert and Rutten, Andrew. "Equilibrium Social Contracts: Social Order with Self-Interested Rulers." Mimeo, Massachusetts Institute of Technology, 1998.

Inman, Robert P. and Rubinfeld, Daniel L. "Rethinking Federalism." *Journal of Economic Perspectives*, 1997, *11*(4), pp. 43–64.

Muthien, Yvonne and Khosa, Meshack. "Demarcating the New Provinces: A Critical Reflection on the Process," in Yvonne Muthien and Meshack Khosa, eds., *Regionalism in the new South Africa*. Brookfield, VT: Ashgate, 1998, pp. 23–56.

Rasmusen, Eric. *Games and information: An introduction to game theory*. New York: Blackwell, 1989.

Waldmeir, Patti. "Anatomy of a Miracle: The End of Apartheid and the Birth of the New South Africa." New York: Norton, 1997.