April 2, 2011

Federal Institutions and the Democratic Transition: Learning from South Africa

Robert P. Inman, *University of Pennsylvania*
Daniel L Rubinfeld, *Berkeley Law*

Journal of Law, Economics, and Organization Advance Access published April 2, 2011

Federal Institutions and the Democratic Transition: Learning from South Africa

Robert P. Inman[*]
University of Pennsylvania

Daniel L. Rubinfeld[**]
University of California, Berkely

We present a model of a peaceful transition from autocracy to democracy using federal governance as a constitutional means to protect the economic interests of the once ruling elite. Under "democratic federalism," the constitution creates an annual policy game where the new majority and the elite each control one policy instrument of importance to the other. The game has a stable stationary equilibrium that the elite may prefer to autocratic rule. We apply our analysis to South Africa's transition from white, elite rule under apartheid to a multi-racial democracy. We calibrate our model to the South African economy at the time of the transition. Stable democratic equilibria exist for plausible estimates of redistributive preferences and rate of time preference ("impatience") of the new majority during the early years of the new democracy. The future of the democratic federal bargain is less certain under the new populist presidency of Jacob Zuma (JEL H11, H77, P48).

[*] Finance, Economics, and Public Policy, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA. Email: inman@wharton.upenn.edu.

[**] Department of Law and Economics, University of California, Berkeley, CA, USA and Department of Law, New York University. Email: drubinfeld@law.berkeley.edu.

The work here has benefited significantly from presentations at Berkeley Law School, Cornell, Michigan, NYU Law School, Norwegian Institute of Technology (Trondheim), Penn Law School, Stanford, Wesleyan, Wharton Applied Economics Summer Workshop, and the Wissenschaftszentrum Seminar on Federalism. The comments of John Ferejohn, Jean-Pierre Benoit, our referee, and editor (Pablo Spiller) significantly improved the presentation and focus of our argument. Special thanks are due Carolyn Ballay, Grant Long, and Jon Stott for extraordinarily valuable research assistance. Finally, the authors had the very unique opportunity to serve as advisors to the constitutional commission—known as the Financial and Fiscal Commission (FFC)—that established the economic ground rules for new federal constitution (1994–1996) and to then advise and assist the Departments of Finance, Education, and Welfare (1996–2000) in implementing the financial arrangements of the new constitution. In this regard, we thank Dr. Junaid Ahmad of the World Bank for facilitating our work with the Commission and Mr. Murphy Morobe and Dr. Antony Melck, cochairs of the Commission, for their openness to our ideas and their patience as we sought to learn the details of the South Africa's new political economy. As is often the case, academics learn the most from those who do it for real. Our experiences are no exception. We hope this effort at generalization does justice to their wisdom and leadership in shaping South Africa's new democracy.

The Journal of Law, Economics, & Organization
doi:10.1093/jleo/ewr001

> (I)t may not be enough to work purely on one-person one-vote, because every national group would like to see that the people of their flesh and blood are in the government. The ordinary man . . . must look to our structures and see that as a colored man I am represented . . . and an Indian must also be able to say, 'I am represented' . . . and the whites must say, 'I have representation.' . . . *Especially in the first few years of the democratic government we may have to do something to show that the system has got an inbuilt mechanism which makes it impossible for one group to suppress the other.* Nelson Mandela, from a speech in Stellenbosch, May, 1991 (As quoted in Waldmeir, pp. 213–214; Italics added.)

1. Introduction

A central challenge for those seeking a peaceful transition from autocracy to democracy is to provide credible protections for the civil and economic rights of the once-ruling minority. Setting promises aside, the once oppressed majority will have the political power, and perhaps the inclination, to expropriate elite assets and incomes once the new democracy is in place. The elite understands this possibility and may therefore continue suppression and autocratic rule, despite sizeable costs. The current political economy literature on democratic transitions has suggested three mechanisms for protecting elite interests in a new democracy: (1) Continued elite control of the military (Acemoglu and Robinson 2001), (2) an upper legislative chamber controlled by the elite with veto powers over redistributive legislation (Lijphart 1984), and (3) the gradual extension of the franchise timed to match the growth of a propertied middle class (Conley and Temimi 2001; Boix 2003; Lizzeri and Persico 2004).

None of these three strategies are likely to be embraced by today's oppressed majorities, however. Allowing the elite continued control over the military or one legislative chamber favors the status quo's distribution of economic resources; both grant the current elite veto power over any reallocations it does not favor. Limiting the franchise to the middle class requires opening that class to the current poor majority either through affirmative action or through education, a path that again depends upon elite (now private sector) decisions and may take generations.

South Africa's successful transition from apartheid to a multi-racial democracy suggests a fourth approach for guaranteeing minority economic rights—a federal constitution with minority elite control over important redistributive services in at least one politically important province.[1] Sections 2–4 spell out the fundamental components of a formal game-theoretic model of how a federal constitution might provide credible protections for elite economic interests.[2] Section 2 describes the requirements necessary for democratic federalism to be the mutually preferred constitution by both the new majority and the old ruling minority. Section 3 specifies a political economy model of fiscal redistribution in which taxes on the high-income elite and redistributive services and transfers to the poor majority are set by democratic politics. If governed by a unitary constitution, tax rates will be set to fully exploit the elite minority and to maximize redistributive transfers to the poor majority. Under a federal constitution, provincial boundaries can be drawn to create at least one elite-run province responsible for redistributive services and transfers to the province's poor residents. If so structured, federal governance creates a "hostage" game in fiscal policy between a majority controlled central government and the elite-run province(s).

Section 4 presents our central result; Proposition 1 provides the necessary and sufficient conditions for such a federal constitution to implement an equilibrium level of redistributive taxation that is less than the fully exploitive tax rate observed under unitary governance. First, we assign responsibility for providing redistributive services to provincial governments such that the assigned services (i) are important to the poor majority's economic welfare and (ii) are most efficiently provided by the public employees of the elite-run provinces. Allocating such services to provinces, importantly to elite-run provinces, gives the elite the ability to harm majority "resident-hostages." We call this first condition the *Assignment Constraint*. Second, we define the boundaries of the provinces so that in equilibrium, the elite remains in political control of their provinces but with a sufficient number of majority resident-hostages so that the majority controlled central government views any threat to harm those hostages as credible. We call this second condition the *Border Constraint*. Together, the Border and Assignment Constraints define the conditions whereby democratic federalism can be an institutional strategy for facilitating the democratic transition.

We define the Assignment and Border Constraints for two alternative political regimes. The first (specified by Lemma 1) we call the exogenous

[1] Waldmeir (1997) and Steytler and Mettler (2001) provide detailed accounts of the bargaining positions of the two major parties to South Africa's transition—the African National Congress (ANC) headed by Nelson Mandela and the National Party (NP) led by F. W. de Klerk—as they negotiated the interim and final constitutions. Mandela made it very clear from the beginning of the negotiations that the ANC would control the police and the army in the new democracy, that there will be no upper chamber controlled exclusively by the elite, and that all citizens would be entitled to vote immediately; see Waldmeir (1997, Chapters 10–13). With those options off the table, a new approach was needed. A federal constitution with significant provincial powers and elite control of at least one province was the answer. For the details of the negotiations over provincial powers, see Waldmeir (1997: 193–197 and 241–244). For negotiations over provincial boundaries to ensure elite control of at least one province (see Muthien and Khosa 1998).

[2] For a more complete description of the detailed model specification, see our NBER working paper, Inman and Rubinfeld (2011).

q-Regime; it has the level of assigned redistributive services, $q = \mathbf{q}$, set by nonmajoritarian politics, say by a constitutional standard enforced by an nonmajoritarian court or by an elected president using agenda powers. The second (specified by Lemma 2) we call the endogenous q^*-Regime; it has the level of redistributive services, $q = q^*$, set by majoritarian politics to satisfy the demands of the typical (presumably, median) majority resident. We then show (Lemma 3) that the Border and Assignment Constraints for the endogenous q^*-Regime are "tighter" than those for exogenous **q**-Regime.

In Section 5, we use the model to evaluate the South African transition. We find that for our specification of the South African political economy at the time of its democratic transition, the agreed to federal constitution appears to have set the provincial borders and service assignments so that democratic federalism could check an inclination by the majority to maximally tax the once ruling elite. Our evidence suggests Presidents Mandela (1994–2000) and Mbeki (2001–2007) used their agenda powers to moderate redistributive demands of the poor majority, implying a **q**-Regime. The new president, Jacob Zuma (2008-), may need to respond to majoritarian demands. If so, then South Africa's fiscal politics will become a q^*-Regime. Lemma 3 would then apply. Fiscal politics that placed budgets within the **q**-Regime's Border and Assignment Constraints may become politics setting budgets outside the smaller q^*-Regime's constraints. If so, democratic federalism alone may no longer be a sufficient constraint on maximal redistribution and alternative but complementary institutions may be needed. We suggest two possibilities.

Section 6 concludes our analysis by considering, first, the potential applicability of democratic federalism to other emerging democracies and then, second, how our analysis of federal governance as a transition institution fits into the new political economy analysis of federalism generally.

2. Federalism and the Democratic Transition: An Overview

Six assumptions specify the underlying political economy at the time of the transition:

1. The oppressed majority does not have sufficient military strength to defeat the current autocratic regime and unilaterally impose a constitution,
2. Once a democratic constitution is in place, the current ruling elite turns over control of its military to the new majority,
3. The oppressed majority will be a demographic and political majority in the new democracy,
4. Elite residents are free to leave the country and/or adopt tax avoidance strategies,
5. Both the majority and the elite understand each others' motives for setting policies and have full information about each others' economic positions, and

6. Constitutional negotiators for the elite and the majority seek a democratic constitution that protects the long-run welfare of their average constituent.

Assumptions (1) and (2) restrict the analysis to peaceful democratic transitions. Assumption (3) ensures that the new (poor) majority will be decisive over central government economic policy. Assumption (4) makes redistributive taxation less than fully efficient. Assumptions (5) and (6) require the choice of the new democratic constitution to be a rational long-run choice by both the elite and the new majority within a full information, dynamic redistribution game. We will focus our attention on a description of the subgame perfect, Nash equilibrium to this game.

The specification of a successful federal constitution at the time of transition proceeds in two stages. In the *constitutional stage*, the residents or their representatives choose either of two democratic constitutions. The first is a *unitary* constitution with all policies decided and administered by a democratically elected central government. The second is a *federal* constitution, where policy responsibilities are shared between the national government and constitutionally created provinces. We assume that simple majority rule determines policies in both cases.[3] At the time of South Africa's provincial creation, it was expected that the rural Northern Cape and the urban Western Cape (centered in Cape Town) would be controlled by the white elite. Because of weak voter turnout by white farmers, the African National Congress (ANC) was (and continues to be) victorious in the Northern Cape. The Western Cape, however, has remained a non-ANC province with the educated white and Asian middle class as the dominant partners in the ruling coalition government. For this reason, we specify the Western Cape as the elite-controlled province when applying our analysis to the South African transition.

Under the federal system, the constitution specifies provincial boundaries that allocate a share of all majority residents to live within the elite province (e.g., the Western Cape). Once provincial boundaries are drawn, all residents are free to move in response to provincial policy choices. Redistributive taxation is always assigned to the central government; otherwise, all elite residents would move to the elite province and set redistributive taxes at zero. In our specification of the federal system, responsibility for the provision of redistributive services is given to the provinces. Elite residents do not receive services from the redistributive budget, although they may consume those services from a separately funded general service budget. The specification of provincial boundaries and the assignment of tax and spending responsibilities define a federal constitution.

Given the specification of the fiscal constitution, the second stage of the transition game is an *annual redistributive policy subgame*. Each year the central government chooses a redistributive tax rate borne fully by

[3] Both unitary and federal constitutions allow amendments. In this sense, the constitutional rules and institutions must be self-enforcing.

the elite to fund the redistributive services and lumpsum transfers to the poor majority. If unitary governance were the rule, the central government would provide services and transfers directly. Under federal governance, however, the central government uses redistributive tax revenues to fund intergovernmental grants for provincially provided redistributive services and lumpsum transfers. Consistent with the view that provincial governments have the ability to reallocate at least some of those central government transfers to their own uses, the elite-controlled province can spend a portion of its transfer revenues on services for elite residents. We call such a reallocation "elite capture" or "elite shirking."[4] Since the elite does not control the national government or majority-run provinces, there is no elite capture in majority provinces.

To evaluate the equilibrium in this two-stage game, we use backward induction. First, at the policy stage (in any year), we determine a set of policy outcomes that define the utilities of the typical poor majority resident, $\omega_t(\cdot)$, and typical elite resident, $y_t(\cdot)$. The policy outcomes must be sustainable as an equilibrium of this repeated fiscal policy game. Given these outcomes, the value of any constitution will then be the discounted present value of all future utilities that follow from the choice of policies of the majority-controlled central government and, under a federal constitution, the elite-controlled provinces. The discount factor, δ bounded as $0 < \delta \leq 1$, may vary for the poor majority and the wealthy elite. The set of democratic constitutions that will allow a peaceful transition from autocracy to democracy will be those for which the discounted stream of annual welfare for both the majority and the elite are greater than or equal to their welfare under the autocratic status quo, assumed to be exogenous.

3. The Annual Policy Game

In this section, we show that the annual policy game played by the elite minority and the poor majority can be characterized as a noncooperative game in which the poor majority chooses a redistributive tax rate and the elite minority chooses whether or not to capture substantial intergovernmental revenues in their elite controlled province. This will allow us to specify the constitutional conditions that will support the outcome that maximizes social welfare—a federal system with minimal shirking by the elite and a redistributive tax rate that is less than the maximal tax rate. We begin by describing the budgetary constraints and cost considerations that limit the available alternatives. This allows us to determine the annual utility that each group will achieve in pursuing each of its strategic alternatives.

[4] There is an extensive literature for United States local governments estimating capture of intergovernmental transfers intended for lower income recipients; see, for example, Duggan (2000) and Gordon (2004). For evidence on the presence of capture in developing economies, see Reinikka and Svensson (2004) for Uganda and Bardhan and Mookherjee (2006) for India.

3.1 Budget Choices

The majority-controlled central government chooses an aggregate redistributive tax per elite resident (τ) whose proceeds are allocated to the provincial governments as a redistributive grant (g) to provide services for majority residents. The central government also sets national standards for the constitutionally assigned, provincially provided redistributive service inputs (q), which are provided at a cost $s(q)$. Service input standards are assumed to be fully enforced.

Service input standards for constitutionally assigned redistributive services may be set in response to: (i) a constitutional requirement to provide a "fair" or "adequate" service level to all citizens successfully enforced by a constitutional court,[5] or (ii) presidential preferences enforced by agenda powers,[6] or (iii) majority citizen preferences enforced by majority rule median voter politics.[7] Standards for the provincially assigned redistributive services that come from a constitutional court or an agenda-setting president we call the *exogenous* $\mathbf{q}$-Regime. Standards set by majority rule politics define an *endogenous* q^*-Regime; in this case, q^* is set by the central government so as to maximize the welfare of the median majority resident. Our analysis will specify the feasibility and sustainability of democratic federalism under both the nonmajoritarian $\mathbf{q}$-Regime and the majoritarian q^*-Regime.

After satisfying the required service standard, however decided, provinces are free to allocate the remainder of their redistributive grant to services of their own choosing. All fiscal policies are decided subject to an aggregate redistributive budget constraint which requires that spending on redistributive services and transfers be financed by centrally raised and administered redistributive taxation.[8]

The taxpaying elite is free to leave the country or to adopt tax avoidance strategies as the redistributive tax rate increases.[9] Tax avoidance is the primary means by which the elite reduces its tax payments. There is a revenue hill for redistributive taxation. Revenues initially increase as the tax rate per elite resident (τ) rises, reach maximum at τ_U, and then decline. Majority dominated unitary governments always select the maximum rate. Given the revenue

[5] For an analysis of how judicial preferences can enforce outcomes different from the majority's preferred outcome (see Ferejohn and Weingast 1992).

[6] See Romer and Rosenthal (1979) for a model where agenda-setting powers, a veto coalition of 33% in the legislature, and a sufficiently low reversion for the provision of redistributive services will be sufficient for a President to check Parliament's preferences for redistributive services. We argue below that this was the case for the Mandela and Mbeki presidencies in South Africa.

[7] To ensure a voting equilibrium, we are implicitly assuming the service standard is one-dimensional—say interchangeable inputs—or national politics are decided by citizen-candidate elections as in Besley and Coate (1997).

[8] The model can also allow for payments to, or receipts from, third parties. In the early post-Apartheid years, payments were made to KwaZulu-Natal to ensure the participation of Chief Buthelezi and Inkatha in the new democracy. These payments are included in our specification of the model for the South African transition.

[9] Although emigration is a possibility, it has not proven to be significant in the case of South Africa. In our application, we focus on tax avoidance behaviors.

potential of national redistributive taxation, it will be important to see if democratic federalism will support an equilibrium redistributive tax rate, denoted τ_F, that is less than τ_U.

3.2 The Cost of Providing Redistributive Services

The primary inputs used by the provinces to provide redistributive services are public employees: teachers, doctors, nurses, social workers, and public administrators. All public employees are assumed to be paid a common civil service wage, which is only imperfectly related to their individual productivity.[10] More productive public employees will therefore be less expensive when providing any required service input bundle. We assume that elite public employees from the old autocratic regime are well trained and therefore have cost advantage over majority, less well-trained public employees. It is this "inherited" productive advantage of elite public employees working in the elite province that will prove crucial to the elite's ability to check redistributive taxation.[11] The majority needs the elite and therefore has an incentive to retain the elite's participation in the provision of redistributive public services.

We assume that if the unitary regime is chosen, the well-trained elite teachers, nurses, doctors, and civil servants will reduce their effort, or more likely, simply exit the public sector for comparable employment in the private economy. This assumption is important. It is the elite's cost advantage that protects the attractiveness to the majority of the federal form of governance, and it is only within federal governance that elite has any ability to hurt the majority if they adopt too high a redistributive tax rate.[12]

[10] Having wages fully independent of employee productivity is not essential, but an imperfect matching of wages to productivity is important. As a consequence of the decision to not discriminate by race, South Africa does have a common wage structure for positions in the civil service, without careful regard for background or training. See Dixit (1997: 94–98) on the use for low-power incentives not tied to marginal product in bureaucracies.

[11] We assume that public services are provided by a common linear technology proportional to the training-adjusted level of public employees: $q = a(X/M)$, where (X/M) is public employees (X) per majority resident (M) and a is employee productivity measured by years of training. As an example, if there is 1 employee for every 25 majority adult residents and that employee has 14 years of training, then $q = 14(1/25) = 0.56$. We assume that the elite public employees have a level of training of a_e, the majority public employees have a level of training of a_m, and that there may be the need to use untrained employees with training of a_u, where $a_e > a_m > a_u$. The cost of provision is $s(q) = S(X/M)$, so that $s_e(q) = S(q/a_e) < s_m(q) = S(q/a_m) < s_u(q) = S(q/a_u)$.

[12] In our analysis, the costs of providing redistributive public services under unitary governance is specified as $s_U(q) = m \cdot s_m(q) + (1 - m).\ s_u(q)$, where m is the share of public employees under unitary governance who are formally trained, whereas $(1 - m)$ is the residual number of public employees hired with only limited training. The elite does not provide public services under unitary governance, or if they do work in the public sector, they mimic the behavior of the majority employees. The motivation for this assumption follows from the theoretical work of Akerlof and Kranton (2005). It finds its empirical support in the extensive work on the adverse consequences for organizational efficiency of racial and educational diversity between managers and workers; see Williams and O'Reilly (1998).

3.3 Redistributive Fiscal Effort

We assume that the central government can successfully monitor the inputs allocated by the provinces to redistributive services, so once the standard for public service provision has been set by the central government, the provinces comply. What the central government cannot monitor, however, is the level of redistributive transfers meant for the poor after the required level of service inputs for education, health care, and the like have been satisfied. These extra or "free" redistributive revenues (r) can be "captured" by the elite in the elite-run province for services consumed by the elite residents. The share so captured ($0 \leq \phi \leq 1$) measures a *lack of* redistributive effort by the province. In the public finance literature, ϕ is often called the "provincial capture" or "flypaper effect" of targeted grants.[13] The majority-run central government and the majority provinces would like minimal provincial shirking with $\phi = 0$. We assume that majority-run provinces allocate all free redistributive revenues to their poor constituents. However, in elite-controlled provinces there is shirking as the elite seeks to push ϕ as high as possible.

We assume there is lower value of fiscal effort ϕ_L, perhaps very small, that the elite province can always allocate to elite residents without detection or penalty by the majority, but there is an upper limit ϕ_H as well. The upper limit defines maximum shirking and is set by the fact that majority residents in the elite-run province can leave the province and relocate to a majority-run province where there is no shirking. Given a cost of exit, the upper limit is set to equalize the welfare of a typical poor resident in an elite-run province with shirking and a majority-run province without shirking. If the majority does leave, then the elite-run province will receive no redistributive transfers from the central government, have no redistributive responsibilities, and thus no ability to influence the central government setting the national redistributive tax rate. Thus, the elite will not exceed this upper limit. As a result, $0 < \phi_L \leq \phi \leq \phi_H \leq 1$. Finally, choosing a level of capture above the lower bound, ϕ_L, is not costless for the elite. When the rate of capture exceeds its lower bound and services or transfers to lower income residents are noticeably reduced, poor residents within the elite province impose a "protest" penalty (ρ) on each elite resident. These costs come as the consequence of spontaneous marches or riots or from formally organized strikes. The costs of such protests may discourage redistributive "shirking" via high capture.

3.4 Economic Welfare

The economic welfare of elite residents will equal their pretax incomes, Y (assumed exogenous) minus their redistributive tax payments (τ) plus any resources "captured back" through reduced fiscal effort ($\phi \cdot r$) in the elite province:

[13] The current empirical literature, both for developed and less developed countries, estimate the rate of capture of central government grants by the provincial governments for their own use to range from $0.30 \leq \phi \leq 1.0$.

$$y(\tau,\ \phi_{\mathrm{L}}) = Y - \tau + \phi_{\mathrm{L}} \cdot r_{\mathrm{e}}(\tau; q),$$

under federalism with low capture;

$$y(\tau,\ \phi_{\mathrm{H}}) = Y - \tau + \phi_{\mathrm{H}} \cdot r_{\mathrm{e}}(\tau; q) - \rho,$$

under federalism with high capture less a protest penalty, and

$$y(U) = Y - \tau_{\mathrm{U}},$$

under unitary governance. All elite residents are assumed to live in the elite province. The elite leadership wishes to maximize $y(\cdot)$.

The economic welfare of a typical majority resident will be the sum of private sector income, W (assumed exogenous), the utility value of redistributive services, denoted $\upsilon(q)$, and any "free" redistributive revenues not captured by the provincial government, $(1 - \phi) \cdot r$. The majority does not pay the redistributive tax. For a majority resident living in an elite province with capture ϕ ($= \phi_{\mathrm{L}}$ or ϕ_{H}):

$$\omega_{\mathrm{e}}(\tau, \phi) = W + \upsilon(q) + (1 - \phi) \cdot r_{\mathrm{e}}(\tau; q),$$

whereas for the majority resident living in a majority province:

$$\omega_{\mathrm{m}}(\tau, \phi) = W + \upsilon(q) + r_{\mathrm{m}}(\tau; q).$$

Since the provision of public services in the elite provinces is more efficient, $r_{\mathrm{e}}(\tau;\ q) > r_{\mathrm{m}}(\tau;\ q)$. In equilibrium, this advantage must be sufficient to just compensate poor residents of the elite province for elite capture.[14] In a federal equilibrium, a fraction (μ) of the majority residents will live in the elite province and $(1 - \mu)$ of the residents will live in majority-run provinces. We assume that the majority leadership wishes to maximize the welfare of the average majority resident, defined as

$$\omega(\tau, \phi) = \mu \cdot \omega_{\mathrm{e}}(\tau, \phi) + (1 - \mu) \cdot \omega_{\mathrm{m}}(\tau, \phi),$$

under federalism, and

$$\omega(U) = W + \upsilon(q) + r(\tau_{\mathrm{U}}; q),$$

under unitary governance with no provinces ($\mu = 0$). The majority leadership wishes to maximize $\omega(\cdot)$.

We focus on finding sustainable constitutions that implements *democratic federalism*. A federal constitution with elite-run provinces is not by itself sufficient to ensure that provinces have influence. The majority-run central government can always set a maximal redistributive tax rate, τ_{U}, while still using provinces to provide redistributive services. Or, stronger still, the central

[14] The equality $\omega_{\mathrm{e}}(\tau,\ \phi) + E = \omega_{\mathrm{m}}(\tau,\ \phi)$ will define the equilibrium maximum value of $\phi = \phi_{\mathrm{H}}$, where E is an exogenous "exit cost" for each majority resident leaving the elite province. The more attractive is the elite province to the majority, the higher will be $\phi_{\mathrm{H}.}$

government can choose maximal redistribution and use a central bureaucracy to provide redistributive services. Here, provinces are irrelevant to the policy outcomes; this is de facto *unitary democracy*. Only under democratic federalism are elite policy preferences respected.

3.5 Credible Elite Punishments

Two conditions, which we define more precisely in Section 4, must hold for high capture to be a credible elite punishment in those instances when the majority leadership selects maximal redistributive taxation, τ_U. The first condition is the Assignment Constraint. It requires that constitutionally assigned redistributive public services, q, be attractive enough that the majority controlled central government still prefers to use provinces, and in particular, the low-cost elite-run provinces, even if the elite adopts high capture. If the Assignment Constraint did not hold, then when the elite province adopted high capture, the majority controlled central government could simply move to de facto unitary governance, supply redistributive services centrally, and deny the elite any access to high capture of free redistributive transfers. Given assignment, the annual level of q required by the central government will be decided by either nonmajoritarian ($\mathbf{q}$-Regime) or majoritarian (q*-Regime) politics.

The second condition is the Border Constraint. This constraint sets a lower and upper bound on the number of majority residents who live in the elite province. If too few majority residents are in the elite province, then the elite's threat to adopt high capture is ineffective as the "pain" of high capture impacts only a few majority residents and can be compensated for by adopting the maximal tax rate. But if there are too many majority residents in the elite province, then the majority can out vote the elite in setting provincial policies and again high capture ceases to be a credible elite punishment.

When the Assignment and Border Constraints are met, then high capture becomes a credible elite punishment. The federal constitution specifies these constraints. If they are met, then the resulting fiscal policy game becomes a hostage game.[15] The majority controls the central government's tax rate and thus holds the elite's income hostage. Through the Assignment and Border Constraints, the elite controls redistributive services to a significant share of the majority population and thus holds the welfare of the average majority resident hostage. Proposition 1 in Section 4 specifies when this hostage game will result in a less than fully redistributive fiscal equilibrium. If so, then democratic federalism has the potential to provide a peaceful transition from autocracy to democracy.

4. Protecting the Elite Through Democratic Federalism

Democratic federalism provides the institutional framework for implementing a hostage strategy within the annual fiscal policy subgame. Its success depends

[15] Schelling (1960: 135–136) first proposed the use of hostages as a means for enforcing incomplete contracts (see also Williamson 1983).

on a credible threat by the elite to harm majority residents through the adoption of the high capture strategy when the majority adopts maximal taxation. This is only possible if the elite controls provincial fiscal policies and prefers a high capture strategy if the majority opts for maximal tax rates. Formally:

Definition. (Credible elite punishment). The high capture strategy will be a credible elite punishment strategy when:

(i) The elite prefers the high capture strategy to low capture when the majority adopts the maximal redistributive tax rate: $y(\tau_U, \phi_H) > y(\tau_U, \phi_L)$;
(ii) The elite remains a political majority in at least one province when the majority adopts the maximal redistributive tax rate—$N(\tau_U) \geq M_{e,}$, where $N(\tau_U)$ is the elite population in the elite province with maximal taxation and M_e is the majority's population in the elite province; and
(iii) The majority prefers provinces and federalism to unitary governance even when the elite adopts a high capture strategy: $\omega(\tau_U, \phi_H) > \omega(U)$.

Requirements (i) and (ii) specify the Border Constraint on the share of the majority population (μ) that is assigned to live in the elite province. Requirement (iii) specifies the Assignment Constraint as a limit on the level of redistributive services that are required of provincial governments; see Appendix A.2. If both constraints are met, then the high capture strategy becomes a credible elite punishment under federalism. The exact specifications for the constraints depend upon political regime within which redistributive services are decided.

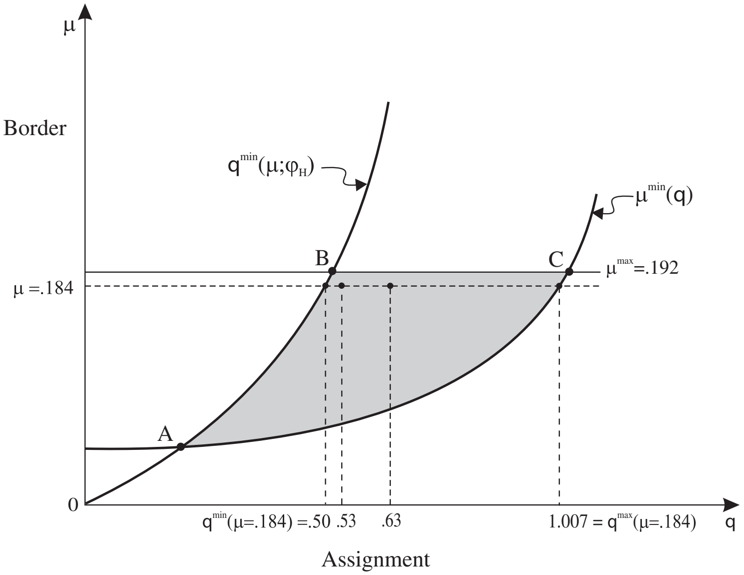
4.1 Border and Assignment Constraints in the **q**-Regime

This political regime applies when a nonmajoritarian political process sets the level of redistributive services. Condition (i) holds if the elite province has a sufficiently large population of majority residents receiving redistributive transfers. If so, then benefits of high capture to the elite compensates for the potential protest costs imposed on each elite resident by the denied majority. Condition (ii) requires the elite always be a political majority in its province(s). Thus, the majority population in the elite province cannot be too large. Together these two conditions define the Border Constraint for the **q**-Regime specified as limits on the share (μ) of the majority population residing in the elite province:[16]

$$\mu^{\max} \geq \mu > \mu^{\min}(\mathbf{q}). \tag{1}$$

Condition (iii) requires that if the central government defects or punishes the elite, it does so within a federal structure where provinces still have fiscal responsibilities. This condition holds when **q** is large enough that the cost advantage of using efficient elite provinces to provide redistributive services more than offsets the loss in the majority's welfare from high capture in those

[16] The Appendix A.2 provides the details.



Mandela Assignment q=.53

Mbeki Assignment (FY 05/06) q=.63

Coordinates for Points:

A: μ = .050; q = .20
B: μ = .192; q = .51
C: μ = .192; q = 1.02

Figure 1.

provinces. For condition (iii) to hold, therefore, we require $\mathbf{q} > \mathbf{q}^{min}(\mu;\ \phi_H)$. But the constitutionally mandated level of redistributive services cannot be set too high either. For each tax rate, as **q** increases the level of free redistributive transfers available for capture declines, possibly reducing the returns to elite capture below the burden of protest costs. If so, high capture will no longer be a credible strategy. The maximum value of **q** that protects ϕ_H as a credible punishment strategy will be that $\mathbf{q}^{max}(\mu)$. The Assignment Constraint in the **q**-Regime is defined by:[17]

$$\mathbf{q}^{max}(\mu) \geq \mathbf{q} > \mathbf{q}^{min}(\mu;\ \phi_H). \tag{2}$$

Lemma 1 follows.

Lemma 1. (Credible elite punishments in the **q**-regime). For political economies satisfying the **q**-Regime Border and $\mathbf{a}_0$ Assignment Constraints, the high capture strategy will be a credible punishment strategy for the elite whenever

[17] The Appendix A.2 provides the details.

the majority adopts a revenue-maximizing redistributive tax rate. (Proof: See Appendix A.3.)

Figure 1 illustrates the feasible constitutional values of μ and **q** sufficient to ensure credible elite punishments at the time of the South African transition decision. The Border Constraint requires that μ lie above the $\mu^{\min}(\mathbf{q})$ curve and below the $\mu^{\max}$ line. The Assignment Constraint requires that **q** lie to the right of the $\mathbf{q}^{\min}(\mu;\ \phi_{\mathrm{H}})$ curve and be at or to the left of $\mathbf{q}^{\max}(\mu)$. The shaded area shows the constitutional assigned values of μ and **q** where both constraints for a credible elite punishment are satisfied.

4.2 Border and Assignment Constraints in the *q**-Regime

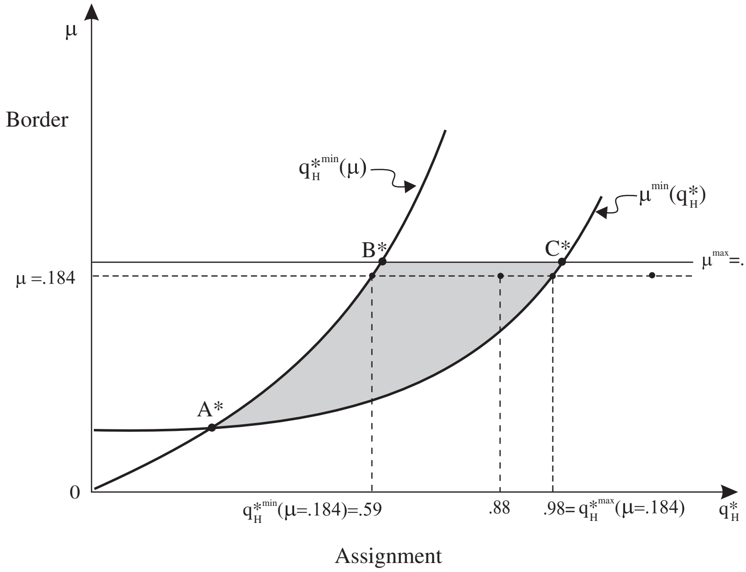
This political regime applies when the majority is free to choose the level of redistributive services that provinces must supply. Services are assigned by the constitution, but their level is endogenous to the preferences of the majority controlling the central government. Since q is one-dimensional in this model, we have in mind a majority voter with the median "taste" for redistributive services as setting q^*.

As before, the q^*-Border Constraint follows from condition (i) that the elite province rationally adopt the high capture strategy when the majority defects to a maximal tax rate. Condition (ii) requiring elite political control continues to define the upper bound for of μ. Conditions (i) and (ii) together again define the Border Constraint for the q^*-Regime:

$$\mu^{\max} \geq \mu > \mu^{\min}(q_{\mathrm{H}}^*). \tag{3}$$

The q^*-Assignment Constraint ensures that condition (iii) holds when the majority can freely choose q^*. In this case, the Assignment Constraint is specified as a comparison of consumer surpluses for the majority under federalism and unitary governance. Consumer surplus comparisons depend upon elite cost advantages and the relative attractiveness of the constitutionally assigned redistributive services. More highly valued services—that is, with high values of $\upsilon'(q) > 0$—increase the consumer surplus of redistributive services to the majority, thereby increasing the relative attractiveness of provinces as low-cost elite providers. Where the majority just prefers federalism to unitary governance defines the minimum for q_{H}^*: $q_{\mathrm{H}}^{*\min}(\mu)$.[18] There is an upper bound on q_{H}^* as well. If the assigned services are too important, the majority demands a high value of q_{H}^* and, for any given tax rate, reduces the amount of resources that can be captured by the elite when adopting strategy ϕ_{H}. Given the protest costs that will result when a high capture strategy is chosen, there is a q_{H}^* above which high capture is no longer a credible choice for the elite. The resulting Assignment Constraint for the q^*-Regime is given by

[18] The details are given in the Appendix A.2.



Mbeki Assignment (FY 08/09) q_H^*=.88

Zuma Assignment (FY 09/10) q_H^*=.95

Coordinates for Points:

A*: μ = .055; q_H^* = .25
B*: μ = .192; q_H^* = .61
C*: μ = .192; q_H^* = .99

Figure 2.

$$q_H^{*max}(\mu) \geq q_H^{*} > q_H^{*min}(\mu). \tag{4}$$

Lemma 2 proves that when the q^*-Border and q^*-Assignment Constraints are satisfied, ϕ_H is a credible punishment.

Lemma 2. (Credible elite punishments in the q^*-regime). For political economies satisfying the q^*-Border and q^*-Assignment Constraints, the high capture strategy will be a credible punishment strategy whenever the majority adopts a revenue-maximizing redistributive tax rate. (Proof: See Appendix A.3)

The shaded area in Figure 2 shows the set of provincial borders (μ) and assignments (q_H^*) that satisfy the q^*-Border and q^*-Assignment Constraints for simple (median) majority rule for the South African economy in transition.

Comparing the coordinates of the shaded areas in Figures 1 (points *ABC*) and 2 (points *A*B*C**) for the South African political economy at the time of transition shows that the set of federal constitutions allowing credible elite punishments in the q^*-Regime is a subset of federal constitutions allowing credible elite punishments in the $\mathbf{q}$-Regime. This result holds generally as Lemma 3.

Lemma 3. (Feasible federal constitutions). When the Border and Assignment Constraints of Lemmas 1 and 2 are met, the set of feasible constitutions allowing democratic federalism is smaller in the q^*-Regime than the $\mathbf{q}$-Regime. The maximal size of the elite province is the same in both regimes. (Proof: See Appendix A.3)

Allowing the majority the right to choose the level of constitutionally assigned redistributive services, rather than having that level set exogenously, narrows the set of constitutions which can sustain credible elite punishments and therefore constitutional stage promises by the majority not to adopt maximal taxation.

4.3 Sustainable Democratic Federalism

Ensuring the feasibility of elite punishments does not by itself guarantee that a federal fiscal allocation will exist as a long-run equilibrium of the annual policy game, however. The cooperative allocation of democratic federalism will only survive if punishments for deviating to maximal taxation or high capture are sufficient to in fact discourage defections from democratic federalism. Formally:

Definition. (Sustainable federal allocations). A federal system will be sustainable with a tax less than τ_U and elite low capture if that allocation is a subgame-perfect Nash equilibrium for the infinitely repeated fiscal policy subgame.

Proposition 1 defines the conditions for when democratic federalism is an equilibrium, and thus sustainable, by specifying minimum and maximum bounds on central government tax rates and intergovernmental transfers. We do so for the case where majority and elite residents both play "grim trigger" strategies in an infinitely repeated policy game. Under the grim trigger strategy, the elite plays $\phi_L < \phi_H$ but were the majority to defect from democratic federalism and select τ_U, the elite would punish the majority by selecting ϕ_H forever.[19] Similarly, the majority plays $\tau_F < \tau_U$ but were the elite to

[19] It is useful to ask whether the elite could select a "tighter" threshold tax rate that will be less than τ_U for the implementation of its punishment strategy, ϕ_H. For elite punishments limited to a known upper rate of capture, as here, the answer is *no*. Announcing a lower threshold tax rate, say $\boldsymbol{\tau} < \tau_U$, for implementing ϕ_H would not be credible. Once the lower threshold $\boldsymbol{\tau}$ was crossed and ϕ_H imposed, no further punishment is possible. Given a fixed ϕ_H, the majority's optimal strategy is to raise taxes to τ_U. The elite's only credible threshold for its trigger strategy is therefore τ_U. It would be instructive to consider whether credible elite punishment "schedules" could be designed which might allow a more aggressive trigger strategy and thus a tighter upper limit for τ_F.

defect from the federal allocation and play ϕ_H, the majority would respond by playing τ_U forever. These grim trigger strategies are the most extreme form of punishment one player can impose on the other for defection in this game. These strategies give democratic federalism its best chance of survival.

The minimal tax rate ($\tau^{\min}$) defines the minimal redistribution required for the majority to find federalism with low capture as its preferred long-run outcome rather than defect to τ_U and run the risk of high capture. The maximal tax rate ($\tau^{\max}$) defines the maximal level of redistribution the elite will accept in a federalist system with low capture rather than defect to high capture and run the risk of a reversion to τ_U as a long-run equilibrium. For a subgame perfect equilibrium for the annual fiscal policy game, we require $\tau^{\max} \geq \tau_F > \tau_F{}^{\min}$, or equivalently from the budget constraint, $g^{\max} \geq g_F > g^{\min}$. Our central proposition uses the bounds $g^{\max} \geq g_F > g^{\min}$. Formally,

Proposition 1. (Sustainable democratic federalism). For either the *q*- or *q**-Regime and political economy satisfying the appropriate Border and Assignment Constraints, there exists a grim trigger strategy equilibrium in which democratic federalism is sustainable. In that equilibrium:

(1) The central government majority chooses a level of redistributive transfers bounded between a maximal grant acceptable to the elite and a minimal grant acceptable to the majority for appropriate elite and majority discount factors, and

(2) The elite province(s) adopts the low capture strategy. (Proof: See Appendix A.3.)

The exact specification of the bounds depends upon the discount factors of the majority and elite coalitions.

Comparative static properties follow directly from the Border and Assignment Constraints and from the specifications of the upper and lower bounds for redistributive services. Three seem worth stressing here. First, it is possible that no choice of μ will satisfy the Border Constraint, either because there are too few elite residents or because the cost advantage of elite production is so slight that credible high capture requires too many majority residents. Second, in the *q**-Regime, credible elite punishments may not be possible when the majority places a very high value on redistributive services. If so, federal institutions will be insufficient to protect elite interests. Finally, as either the majority or elite residents become less patient and their discount factor (δ) declines, the range of federally sustainable redistributions will be narrowed. It is possible for very impatient residents (particularly majority residents) that there will be no sustainable equilibrium level of redistributive transfers—in particular, the majority's minimally acceptable grant rises to exceed the maximal grant that the elite will pay: $g^{\min} > g^{\max}$.

5. Democratic Federalism and the South African Transition

5.1 Background

There is little doubt that at the time of the 1994 negotiations for the new constitution, the two important minority parties, the white elite National Party (NP) and the Zulu Inkata Freedom Party (IFP), needed credible assurances that their once favored economic positions would be protected in the new democratic regime. Both pressed for a federal solution with political control over at least one province, with each province promised significant fiscal powers; see Waldmeir (1997, Chapter 13). An agreement was reached in mid-April of 1994, codified as the Interim Constitution. The Interim Constitution drew the boundaries for nine new provinces, six to be controlled by the ANC, one by the IFP (KwaZulu Natal), and up to two by the NP (the rural Northern Cape and the urban Western Cape). Borders were drawn explicitly to have political majorities for the IFP and NP provinces. In addition, the Interim Constitution outlined fiscal assignment for redistributive services. K-12 education, health care, and the administration of welfare services and payments were specified as "concurrent" functions which may be managed by the national government alone (our de facto unitary governance), by the provinces alone (our democratic federalism), or shared.[20]

With the Interim Constitution in place, elections for the new President and the provincial legislatures went forward. Nelson Mandela was elected President, and the NP and IFP each won political control over one province, the Western Cape, and KwaZulu-Natal, respectively.[21] Although the Interim Constitution created independent provinces and the needed fiscal assignments, it left the details of federal governance largely unspecified. The final Constitution of the republic of South Africa (1996) approved in October, 1996 filled the gap. It maintained the provincial borders (Chapter 6: 103) and concurrent redistributive service assignments (Schedule 4, Part A) as specified in the Interim Constitution, but now assigned all important taxation, except for property taxation, as central government responsibilities unless permitted by an explicit act of Parliament (Chapter 13: 228).

Further, the new Constitution gives formal agenda-setting powers to the President (Chapter 4: 73) as well as the responsibility for supervising the provision of redistributive services (Chapter 5: 100-1b). The President may veto any parliamentary approved legislation (Chapter 4: 79). Subject to the constraints of democratic federalism, ANC party leadership has determined South Africa's fiscal policies; see Wittenberg (2006: 346).

[20] Shared provision would involve a preassigned level q_0 to be provided by the national government with the provinces then providing the difference between $\mathbf{q}$ or q^* and the q_0 on their own. The specified value of q_0 would operate in our model simply as a targeted lumpsum grant for redistributive services provided at a cost of $s_{\mathrm{m}}(q_0)$ in majority provinces and $s_{\mathrm{e}}(q_0)$ in elite provinces. The analysis here would go through completely.

[21] Provincial borders were originally drawn to facilitate an NP victory in the Northern Cape to accommodate the rural whites; see Muthien and Khosa (1998). Although the white landowners voted for the NP, the white farmhands failed to vote and the Northern Cape was won by the ANC by a small majority. It has remained an ANC province ever since.

Table 1 summarizes South Africa's post-Apartheid redistributive fiscal policies. The table lists aggregate central government revenues, total redistributive grants per capita for the country as a whole, for the elite-run Western Cape, and for the other majority-run provinces; provincial spending for assigned redistributive services in the elite ($s_e(q)$) and majority provinces ($s_m(q)$), and finally, the feasible level of redistributive services q for the elite and majority provinces measured in input units as years of public employee training per majority resident.[22]

We can see from Table 1 that provincial governments have been given a significant role in the provision of redistributive services, funded entirely by grants from the central government. Beginning in FY 1998/1999, and continuins to today the elite Western Cape receives on average 20% less in assigned services grants to provide a common level of redistributive services, consistent with their efficiency advantage. Finally, there is a clear break in the level of required redistributive services funded by these budgets in FY 2006/2007. The required level of redistributive service inputs begins to rise significantly. This date corresponds to the growing militancy by the left wing of the ANC and suggests a shift from moderating presidential politics to a possibly more redistributive majority rule politics.

The question we wish to answer is this: Has this level and allocation of redistributive public spending been sufficiently constrained so as to sustain democratic federalism in South Africa? That is, have the requirements of Lemmas 1 or 2 been satisfied, and if so, has the overall level of redistributive taxation and spending been consistent with the requirements of Proposition 1 for sustainable democratic federalism? To provide one set of answers to these questions, we have parameterized our model to the South African political economy at the time of the transitions.

5.2 Is Democratic Federalism Feasible in South Africa?

For democratic federalism to be a viable long-run constitution, it must first satisfy the Border and Assignment Constraints specified for the **q**-Regime (Lemma 1) or, for the later budgets after FY 2006/2007, the q^*-Regime (Lemma 2).

Crucial to the successful transition was that at least one important province—the Western Cape—be politically controlled by the elite.[23] The borders of

[22] We specify that input bundle by $q = a \cdot (X/M)$, where X/M is public employees (X) per adult majority resident (M) and where a is the measure of employee "productivity" equal to $a_e = 17$ years of education for the average elite public employees, $a_m = 14$ years of education for the average "trained" majority employee, and $a_u = 7$ years of education for the average "untrained" majority employee. For example, in FY 1995/1996 in the elite province, $q_e = 17(1/32) = 0.56$, whereas in the average majority province, $q_m = 14(1/32) = 0.44$. The number of employees per majority resident was set by national standards, but the quality of the employees—measured by years of schooling—-was much higher in the elite province. See Inman and Rubinfeld (2011) for complete details behind Table 1 as well as data relating to the basic grant available to the elite and the majority.

[23] The ANC has never won more than 45% of the vote in the Western Cape. Coalitions of the various elite opposition parties have won at least 51% of the vote; see www.elections.org.za. In the latest election 2009, the Democratic Alliance won 48% of the Western Cape and the "break-away" moderate party from the ANC called the Congress of the People (COPE) won 9%. The ANC won only 32% of the Western Cape vote.

Table 1. RSA Intergovernmental Transfers: Real (2000) Rand per Capita

Fiscal Year	Central government revenues	Redistributive grants national	Redistributive grants: Western Cape	$s_e(q)$ Western Cape	q_e Western Cape	Redistributive grants majority provinces	$s_m(q)$ Majority provinces	q_m Majority provinces
1995/1996	4237	2189	2923	1371	0.56	2119	1356	0.44
1996/1997	3938	2030	2587	1334	0.52	1978	1345	0.44
1997/1998*	3942	2000	2424	1250	0.49	1959	1332	0.43
1998/1999	4265	2154	2206	1398	0.55	2149	1709	0.55
1999/2000	4093	2108	2097	1368	0.54	2110	1674	0.54
2000/2001	6636	2242	2185	1455	0.57	2247	1778	0.58
2001/2002	5570	2302	2196	1494	0.59	2313	1826	0.59
2002/2003**	5178	1903	1720	1342	0.53	1923	1500	0.49
2003/2004	5630	2151	1896	1479	0.58	2180	1700	0.55
2004/2005	5610	2231	1941	1514	0.60	2264	1766	0.57
2005/2006	5962	2327	2011	1609	0.63	2363	1890	0.61
2006/2007	6764	2559	2186	1750	0.69	2603	2082	0.68
2007/2008	7760	2735	2293	1835	0.72	2787	2230	0.72
2008/2009	8381	3005	2522	2018	0.79	3063	2450	0.79
2009/2010	8113	3213	2710	2168	0.85	3273	2619	0.85

Sources. FY: 1995/1996 to 1997/1998: FFC, *The Allocation of Financial Resources Between the National and Provincial Governments: FY 1997/1998*, Tables 2, 3, 6b. FY 1998/1999 to 2009/2010: Minister of Finance, *Division of Revenue Bill, Various Years,* Part 4: Provincial Allocations.

Column definitions: For the purposes of this analysis, all allocations to KwaZula-Natal are included as part of the allocations to "Majority Provinces." Central Government Revenues = Total revenues per capita raised by central government taxation; Redistributive Grants = Total intergovernmental transfers per capita paid to the province(s), averaged over all provinces (National Average), for the Western Cape, and for all other (Majority) Provinces excluding the Western Cape; $s(q)$ = Assigned service grants per capita to fund 5–17 education, primary health care services for (lower income) citizens qualifying for medical assistance, and social security grants for the elderly, disabled, and children, for the Western Cape ($s_e(q)$) and the average for all other (Majority) Provinces ($s_m(q)$); q_e and q_m are estimates of the redistributive service bundle measured as public employee training years per majority adult resident provided in the Western Cape and all other (Majority) Provinces, respectively; see Inman and Rubinfeld (2011) for details.

*Data for FY 1997/1998 is based upon projected grants provided in the FFC, *The Allocation of Financial Resources Between the National and Provincial Governments: FY 1997/1998*, Table 6b.

**Beginning with the FY 2002/2003 Budget, the Department Finance adjusted the accounting procedures for funding of the provincial activity. There is therefore an unavoidable break in the data sequence. All financial data from FY 2002/2003 onward is recorded on a consistent basis.

the Western Cape were explicitly drawn to ensure elite political control over provincial politics and a sufficiently large share of majority residents as hostages so that elite high capture would be a credible punishment if the poor majority chose maximal taxation at the national level. The resulting share of the majority voting-age population residing in the elite province is estimated as $\mu = 0.184$.[24] For our specification of the South African political economy, this value of μ satisfies the required Border Constraints for both the **q**- and q^*-Regimes; see Figures 1 and 2.[25]

The Mandela presidency was arguably a **q**-Regime. During his tenure, the level of redistributive services was recommended by a constitutionally created independent commission (Chapter 13: 220–222) known as the Financial and Fiscol commission (FFC), forwarded to the legislature without change by President Mandela, and then approved without amendment by the ANC controlled legislature.[26] The FFC membership was appointed by Mandela and was equally divided between ANC and NP representatives. Commission decisions were typically made by unanimous agreements between the representatives of the two parties.

The FFC's recommended level of redistributive services was 1 teacher per 38 school-aged children, 3.5 preventive health care clinic visits a year for each majority adult and child, and 4500 (real 2000) Rand for each income eligible child, elderly, and disabled majority resident for social insurance transfers. Together, these targets required redistributive grants sufficient to pay for 0.038 public employees per majority resident or, for an average level of training of 14 years per employee, $q = 0.53$ public employee training-years per majority resident.[27] To fund this FFC recommended level of redistributive services, transfers of

[24] We define the actual value of $\mu = (M_e/M)$, where M_e is the majority adult population in the elite-run province and M is the total majority adult population. The actual voting outcomes over the past 14 years favors the non-ANC (elite) coalition by a small majority. Therefore, $N(\tau_F)/[M_e + N(\tau_F)] \geq 0.51$ must hold for the actual adult populations of the Western Cape, where $N(\tau_F)$ is the elite adult population in the elite province. The average elite adult population over this period in the Western Cape was $N(\tau_F) \simeq 4.8$ million. If so, then to meet the narrow majority voting outcomes observed in the actual voting data, $M_e = 4.6$ million must hold: $4.8M/[4.6M + 4.8M] = 0.51$. Finally, the total majority adult population in South Africa at the time of the transition was $M = 25$ million. Thus, we specify $\mu = (M_e/M) = 4.6M/25M = 0.184$.

[25] Figures 1 and 2 follow from our specification of the South African political economy available in the longer version of this article; see Inman and Rubinfeld (2011). The upper bound on the majority that can be assigned to the elite province, denoted μ^{max} in Figures 1 and 2 is specified as $\mu^{max} = N(\tau_U)/M = 0.192$, where $N(\tau_U)$ is the maximum elite population paying taxes even when the majority chooses the full redistributive tax rate. To specify this constraint for South Africa, we use the estimates by Gruber and Saez (2002) for the peak of the revenue hill for upper income residents in the United States; see Inman and Rubinfeld (2011) for details. Since avoiding taxation is easier than leaving the country, this estimate of μ^{max} is likely to provide a lower bound (i.e., smaller shaded area) to the feasibility of democratic federalism.

[26] For example, the 1998 budget proposals by the Finance Department to the legislature commented that "it's (FFC's) recommendations for the division of resources between the three spheres of government (that) form the basis of the current allocations" (1998 Budget Review, Department of Finance, as quoted in *FFC: A Ten Year Review*).

[27] See FFC, *The Allocation of Financial Resources Between the National and Provincial Governments*, *FY 1996/1997*, September 8, 1995.

1635 Rand/capita are needed in the majority provinces and 1347 Rand/capita in the Western Cape.[28] These estimates are very close to actual redistributive funding as reported in Table 1 for the budget years of the Mandela presidency (FY1995/1996 to FY2000/2001). Given $\mu = 0.184$, $\mathbf{q} \simeq 0.53$ falls within the set of feasible assignments satisfying Lemma 1 for a **q**-Regime; see Figure 1, shaded area ABC. Democratic federalism within a mandela presidency.

Matters became less certain under the leadership of Mandela's successor, Thabo Mbeki. The Mbeki budgets' levels of redistributive services were very near the Mandela/FFC recommendations until FY 2005/2006 at which time redistributive spending began a strong upward trend toward today's values of $q = 0.85$. This break in required redistributive services suggests a possible break in underlying political regimes as well, away from a strong president setting required redistributive services exogenously and toward a president increasingly responsive to the preferences of the majority controlled ANC. Concurrent political events culminating in the ouster of President Mbeki in favor of the populist candidate Jacob Zuma, first as head of the ANC (December, 2007), and then as President (September, 2008) strongly suggest such a regime change.[29] If so, South Africa's federal policies must now meet the requirements of Lemma 2 for a q^*-Regime. Our analysis suggests that the last official Mbeki budget in FY 2008/2009, arguably set by his response to majority opposition to his fiscal policies, does satisfy the requirements for Lemma 2. We estimate that for this budget $q^*_{\mathrm{H}} = 0.88$.[30] We conclude that for this budget democratic federalism remains feasible: $q^{*\max}_{\mathrm{H}}(\mu = 0.184) = 0.98 > 0.88 > 0.59 = q^{*\min}_{\mathrm{H}}(\mu = 0.184)$ (see Figure 2).

The last budget for which we have full data is FY 2009/2010 and corresponds to the first official budget from the new presidency of Jacob Zuma. The FY 2009/2010 budget implies $q^*_{\mathrm{H}} = 0.95$, just within the upper bound for redistributive services to meet the requirements of Lemma 2; see Figure 2. This is an 8% increase over the last Mbeki budget and a nearly 35% increase over what might

[28] From footnote 11, the required spending per majority adult to support $\mathbf{q} = 0.53$ is specified as $S(X/M) = S \cdot (\mathbf{q}/a) = 80{,}000\ (0.53/14) = 3030$ Rand/majority adult, where $S = 80{,}000$ Rand is the average salary of a public school teacher in 1999. The average ratio of majority adults to total population is 0.54 ($=25M/46M$), implying a required redistributive service grant *per capita* of 1635 Rand/capita ($= 3030\text{R} \times 0.54$). To fund $\mathbf{q} = 0.53$ in the elite province, $S\ (\mathbf{q}/a_{\mathrm{e}}) = 80{,}000\ (0.53/17) = 2494$ Rand/majority adult is required or approximately 1347 Rand/capita ($= 2494\text{R} \times 0.54$).

[29] President Mbeki began to feel pressure from the left wing of the ANC midway through his presidency, particularly on matters of public services for the poor. On this growing pressure from the left, see "Boost for Zuma's Leadership Campaign," *Financial Times* (September 21, 2007, p. 4).

[30] Figure 2 specifies the levels of majority demand for redistributive services under the assumption that the elite has adopted the high capture strategy. The value of the capture parameter ϕ affects the effective price for redistributive services. Thus, when evaluating the credibility of the high capture strategy, we do so conditional upon the majority's demand for q, given $\phi = \phi_{\mathrm{H}}$. The actual values of q that appear in Table 1, however, are the levels of redistributive services under the identifying assumption that democratic federalism is in place—that is, the capture parameter $\phi = \phi_{\mathrm{L}}$. This explains the difference between the values of q^*_{H} in Figure 2 for credible elite capture and the values of q reported in Table 1.

reasonably be seen as the last presidentially decided budget of FY 2005/2006. A slight increase in the majority's demands for redistributive services will push $q^*_{\mathrm{H}} > 0.98$. This level of demand for redistributive services moves equilibrium fiscal policies outside the feasible set for democratic federalism in Figure 2 (area $A^*B^*C^*$), undermines the ability of the elite to impose a credible high capture penalty, and leaves the door open to maximum redistribution. In this case, democratic federalism is no longer a feasible check on redistributive taxation.

5.3 Is South Africa's Federal Contract Sustainable?

Even if the requirements of Lemmas 1 and 2 are met and democratic federalism is *feasible*, it may not be *sustainable* by the requirements of Proposition 1. For sustainability, the parties to the constitution must be sufficiently far-sighted that they check their short-term inclinations to exploit the other party. Far-sighted players will have values of their discount factor (δ) near 1; short-sighted players nearer 0. The true values of the discount factors for majority and elite residents are not known. Karlan and Zinman (2008) estimate the rate of time preference of lower income South Africans at 200%, suggesting a discount factor of $\delta_{\mathrm{m}} = 0.33$. The real current rate of interest for South African treasury bonds is about 7%; assuming the elite chooses to save at that rate, then $\delta_{\mathrm{e}} = 0.93$. For those discount factors, democratic federalism can be sustainable for both the **q**- and q^*-Regimes for our specification of the South African political economy and the requirements of Proposition 1.[31]

5.4 The Future for Democratic Federalism in South Africa

Though democratic federalism has provided a valuable check on redistributive taxation in South Africa for the early years of its democracy and arguably offered the "inbuilt mechanism which makes it impossible for one group to suppress the other" that President Mandela thought so important for the transition to democracy, there is no guarantee that these federal institutions can continue to play this important role in the future. Although the majority has seen a near doubling of redistributive service inputs per majority resident since the start of the new democracy, current service levels still fall short of what is now provided to the typical elite family. If the majority were to demand parity in education and health care inputs to levels now available to the elites, q^* would need to rise to 1.14 and clearly move the fiscal equilibrium outside the set $A^*B^*C^*$ in Figure 2 needed for

[31] For the **q**-Regime, $g^{\mathrm{max}}(\delta_{\mathrm{e}} = 0.93; \mathbf{q} = 0.63) = 3300$ Rand/resident $> g^{\mathrm{min}}(\delta_{\mathrm{m}} = 0.33; \mathbf{q} = 0.63) = 3234$ Rand/resident $> g^{\mathrm{min}}(\delta_{\mathrm{m}} = 0.93, \mathbf{q} = \mathbf{.63}) = 3108$ Rand/resident. For the q^*-Regime, $g^{\mathrm{max}}(\delta_{\mathrm{e}} = 0.93; \mathrm{q_H}^* = .95) = 3304$ Rand/Resident $> g^{\mathrm{min}}(\delta_{\mathrm{m}} = 0.33; \mathrm{q_H}^* = .95) = 3261$ Rand/resident $> g^{\mathrm{min}}(\delta_{\mathrm{m}} = 0.93; \mathrm{q_H}^* = .95) = 3187$ Rand/resident. An impatient majority ($\delta_{\mathrm{m}} = 0.33$) will always demand more immediate redistribution than a patient majority ($\delta_{\mathrm{m}} = 0.93$) as is observed for our simulated values of g^{min}. See Inman and Rubinfeld (2011).

a feasible federal contract.[32] In this case, democratic federalism loses its ability to check redistributive taxation.[33]

If so, then checks on such majority demands for "full equality" will be needed. Two suggest themselves.[34] First, rather than assuming a monolithic majority, there may arise a less redistributive minority within the majority (ANC) party capable of holding the "radical" majority in check thereby keeping $q*$ within feasible set of Figure 2. Efforts by the elite to create an "instant middle class" within the ANC through affirmative action of corporate management may serve this role. Second, a president such as Mandela or an "early" Mbeki capable of winning majority party support without embracing maximal redistribution might be able to restore a **q**-Regime and still increase redistributive services above FY2010 levels; see Figure 1. The potential now exists within the current Parliament to propose and sustain such a nonmajoritarian redistributive budget.[35] The issue of course is whether President Zuma has both the inclination and ability to hold potentially "excessive" redistributive demands in check.[36] It is here, within ANC party politics and the specification of majority and presidential preferences, that the future of democratic federalism as an effective constraint on South Africa's redistributive fiscal policies will ultimately be decided.

[32] The typical elite resident is now provided with 1 teacher for every 20 students and 3.5 heath care visits per family member. Assuming a doubling of the current income transfer for children in poverty and disabled and elderly pensioners, then q_{m} would need to equal 1.14 public employee training years per majority residents. The ratio of $1.14/0.85 = 1.34$ implies a 34% shortfall in inputs for majority residents relative to the typical middle class (elite) resident.

[33] That the preferences of the majority can undo the stability of institutions was first stressed by Riker (1980) in his friendly critique of the then new institutional political science when he asked: Where do institutions come from? If the majority does not like the performance of an institution, they can change it. So too here.

[34] There is a third alternative worth mentioning but it does not appear likely within South Africa in the near term. This is the emergence of a strong centrist party within Parliamentary elections. This alternative would break the ANC's monopoly hold on fiscal policy and the likely outcome would be the median position on redistribution between that of the elite and the current ANC. At the moment, the main centrist party is the Democratic Alliance and they hold 17% of the seats in Parliament. In response to their loss of the ANC party apparatus in December of 2007, a group of ANC moderates broke away from the ANC to form a new party called the COPE. In the most recent parliamentary elections, COPE was third with 7.5% of the vote. The ANC still controls 65% of the seats in the new parliament.

[35] The ANC now controls 65% of Parliament's set. The major "elite party" known as the Democratic Alliance has 17% of the seats, whereas the breakaway moderate ANC party known as the COPE now has 7% of the seats. The anti-ANC IFP has 5% of the seats. What is required is an additional 4% of the legislature to form a blocking veto coalition. The President has the discretionary resources he needs to win over the required additional votes from the small conservative, moderate, and even radical parties. For how this might be done, see Fitts and Inman (1992).

[36] At least to date, Jacob Zuma seems so inclined. However, his current platform (shown as $q^*_{\mathrm{H}} = 0.95$ in Figure 2) has been strongly criticized by Julius Malema, leader of the ANC Youth League, who is advocating significantly more redistributive services paid for through the expropriation of white owned assets. Zuma and the older ANC leadership have so far been able to isolate Malema, labeling him an "anger young man." Malema has recently been assigned to take anger management classes; see C. Hunter-Gault, "Letter From South Africa," *The New Yorker* (July 5, 2010).

6. Summary and Extensions

Any nation state hoping to move from autocratic, elite rule to a functioning democracy with protected property rights, and to enjoy the economic benefits such a transition can confer, faces the challenge of providing credible assurances to the elite that their economic interests will not be fully exploited after the transition. Three strategies used in prior transitions—continued elite control of the military, elite veto power in the legislature, or the gradual extension of the franchise—are unlikely to be accepted by any repressed majority today.

South Africa's transition from apartheid to democracy suggests a fourth alternative, what we have called democratic federalism. If the economic elite is sufficiently large and geographically concentrated so that an elite controlled province can be established (our Border Constraint) and if the constitution assigns to the provinces, particularly the elite-run province, the right to provide important redistributive public services (our Assignment Constraint), then a federal democracy, even with universal suffrage and a majority controlled army and central legislature, can provide protection for elite economic interests. So constructed democratic federalism creates a "hostage game" played between the new majority's control of taxation and the elite's efficient provision of redistributive services that can constrain the majority's inclination to exploit the old elite. To date, democratic federalism appears to have so checked maximal redistribution in South Africa.

Democratic federalism, at least as specified here, may not be for everyone, however. Our Border Constraint requires a geographically concentrated elite to allow for an elite-run province. South Africa could meet this requirement by establishing elite provinces around either Johannesburg or Cape Town. The constitution chose Cape Town with the creation of the province of Western Cape. In contrast, early efforts to establish a federal democracy in then Rhodesia were undone by the reluctance of the rural elites spread throughout the country to accept minority status in all provinces; Barber (1967: Chapters 1–8). Our Assignment Constraint requires that the assigned redistributive services be efficiently provided by the elite minority. Recent efforts to establish a federal democracy in Sri Lanka failed for just this reason, as the Tamal minority had no comparative advantage in the provision of any majority valued government service.

One can imagine democratic federalism succeeding in Iraq, however. Like South Africa, there is a talented elite minority (the Sunni) concentrated in a major urban center (Baghdad) and a poor, largely rural majority now in control of the central government (the Shiite). A federal structure and provincial boundaries are now part of the interim Iraqi constitution; see Dawisha and Dawisha (2003) and Anderson and Stansfield (2005). What appears to be missing to date is a trusted leader, a "Mandela," capable of persuading the competing groups that such a federal compromise is in all parties' long-run interests.[37]

[37] Beyond case studies, our model of how federal institutions can facilitate the transition provides a foundation for including a "federal" variable in a larger empirical study of the transition into democracy as, for example, in Przeworski et al. (2000) or Boix (2003).

The central focus of our analysis has been the value of federal institutions to new democracies, but our same institutions have value in established democracies as well. This is reassuring. One does not need to introduce major constitutional reforms as the economy and democracy mature. In our analysis, the assignment of fiscal responsibilities centralizes the taxation of mobile tax bases to ensure the new majority that redistribution can occur, but decentralizes the provision of redistributive services to states or provinces to empower the old elite in the annual fiscal game that sets redistributive tax rates. Centralized taxation and provincial provision of redistributive services—education, health care, housing, day care, and the administration of income transfers—is the assignment of responsibilities recommended for established economies as well (see Musgrave 1959; Oates 1999).

As the democracy matures and rival political factions arise to challenge the monolithic poor majority, then provinces can come to serve two important stabilizing functions for established democracies. First, as Madison (1788) first argued in *Federalist* No. 51 and as de Figueiredo and Weingast (2005) showed quite generally, provincial governments can serve as an important check on central government powers. Provincial leaders represented in the central legislature can band together to check abuses of local rights and policies by a nationally elected executive. Second, provincial politics become useful "testing grounds" for new political leadership. In Myerson's (2006) analysis of political federalism, provincial governments provide elected politicians an opportunity to demonstrate their commitment to democratic rule and efficient government. Political competition within provinces and "yardstick" comparisons across provinces allows voters to identify and defeat corrupt or inefficient politicians. Those politicians that provide honest and efficient leadership at the provincial level develop a reputation that can lead to national election. Further, the threat of credibly honest and capable provincial leaders disciplines current national leaders.[38] Although our analysis shows how federalism helps the emergence of democracy, the analyses of diFigueiredo and Weingast and Myerson show how federalism contributes to long-run democratic survival.

Contemporary analyses of federal governments from Tiebout (1956) to Riker (1964) to Buchanan and Brennan (1980) to Weingast (1995, 2009) have focused on the virtues and difficulties of federal governance *within* established democracies. Our work has sought to answer a prior question: Can federal institutions facilitate the transition from autocracy *to* democracy, and if so, what should be the structure of those institutions? From our analysis, the answer is yes, provided federal institutions, perhaps as specified here,

[38] This was indeed the hope of the NP's President F. W. de Klerk. He expected the new ANC run governments to reveal their incompetence and perhaps corrupt behaviors and that such performance would stand in sharp contrast to the well-run Western Cape. "When asked whether he would miss the supremacy of the NP, (de Klerk) responded that he would not be out of office for long. He was not the only one who believed that liberation movements often fail at governing." C. Hunter-Gault, "Letter from South Africa," *The New Yorker* (July 5, 2010). On the potential for political competition within the South African provinces, see Hawker (2000) and Lodge (2005).

create provinces and assign fiscal responsibilities in ways that mutually empower and then benefit both the new majority and the old elite.

Appendix: Proofs

A.1. Model Specification and Variable Definitions

Demographics:

M = Adult (voting age) population.
M_e = Adult population residing in elite province(s).
$\mu = M_e/M$ = Share of adult population residing in elite province(s).
N = elite residents.

Income:

W = Income (exogenous) of majority residents.
Y = Income (exogenous) of elite residents.

Technology of Public Service Provision:

$q = a\cdot(X/M)$ = Quality adjusted (a) public employees (X) per majority resident (M).
a_e = Years of training of elite public employees.
a_m = Years of training of majority public employees.
a_u = Years of training of "untrained" public employees.

Costs of Public Service Provision:

S = Salary (uniform) paid to public employees.
$s_e(q) = S\cdot(X/M) = S\cdot(q/a_e)$ = Cost per q in elite province(s) using elite employees..
$s_m(q) = S\cdot(X/M) = S\cdot(q/a_m)$ = Cost per q in majority province(s) using majority employees.
$s_u(q) = S\cdot(X/M) = S\cdot(q/a_u)$ = Cost per q using untrained employees in unitary governance.
m = Share of unitary governance employees considered "untrained."
$s_F(q) = \mu\cdot s_e(q) + (1-\mu)\cdot s_m(q)$ = Average cost per q under federalism.
$s_U(q) = m\cdot s_m(q) + (1-m)\cdot s_u(q)$ = Average cost per q under unitary governance.

Government Budget Constraint:

g = Redistributive grant per majority resident $= g(\tau) = [\tau\cdot N(\tau) - Z]/M$;
τ = Redistributive tax rate per elite resident (N).
$N(\tau)$ = Elite residents paying the redistributive tax allowing for tax avoidance.
τ_U = Maximal redistributive tax rate.
Z = Outside (exogenous) payments ($Z > 0$) or transfers ($Z < 0$) to the redistribution budget.

Majority Preferences and Demands Under the q-Regime*:

$\upsilon(q) = \lambda \ln(q)$, $\lambda > 0$.
$q^*_{\mathrm{L}}(\mu, \lambda)$ = Demand for q when elite capture equals $\phi_{\mathrm{L.}}$
$q^*_{\mathrm{H}}(\mu, \lambda)$ = Demand for q when elite capture equals ϕ_{H}.
$q^*_{\mathrm{U}}(\lambda)$ = Demand for q when under unitary governance.

Majority Citizen Welfare:

$\omega_{\mathrm{e}}(\tau, \phi) = W + (1 - \phi_{\mathrm{L,\,H}})\cdot r_{\mathrm{e}}(\tau; q) + \lambda \ln(q)$ = Majority welfare in elite province.
ϕ_{L} = Low elite capture.
ϕ_{H} = High elite capture.
$r_{\mathrm{e}}(\tau; q) = [g(\tau) - s_{\mathrm{e}}(q)]$ = "Free" provincial revenues per majority resident.
$\omega_{\mathrm{m}}(\tau, \phi) = W + r_{\mathrm{m}}(\tau; q) + \lambda \ln(q)$ = Majority welfare in majority province.
$r_{\mathrm{m}}(\tau; q) = [g(\tau) - s_{\mathrm{m}}(q)]$ = "Free" provincial revenues per majority resident.
$\omega(\tau, \phi) = \mu \cdot \omega_{\mathrm{e}}(\tau, \phi) + (1 - \mu)\cdot\omega_{\mathrm{m}}(\tau, \phi)$ = "Average" majority welfare under federalism.
$\omega(U) = W + r(\tau_{\mathrm{U}}; q) + \lambda\cdot\ln(q)$ = Majority welfare under unitary governance.
$r(\tau_{\mathrm{U}}; q) = [g_{\mathrm{U}} - s_{\mathrm{U}}(q)]$ = "Free" Central Government Revenues per majority resident.

Elite Citizen Welfare:

$y(\tau; \phi_{\mathrm{L}}) = Y - \tau + \phi_{\mathrm{L}}\cdot r_{\mathrm{e}}(\tau; q)$ = Elite welfare in elite province with low capture.
$y(\tau; \phi_{\mathrm{H}}) = Y - \tau + \phi_{\mathrm{H}}\cdot r_{\mathrm{e}}(\tau; q) - \rho$ = Elite welfare in elite province with high capture.
$r_{\mathrm{e}}(\tau; q) = [g(\tau) - s_{\mathrm{e}}(q)]\cdot[M_{\mathrm{e}}/N(\tau)]$ = "Free" provincial revenues per elite resident.
ρ = Penalty per elite resident for adopting high capture.
$y(U) = Y - \tau_{\mathrm{U}}$ = Elite welfare under unitary governance.

A.2. Defining Border and Assignments Constraints

q-Border Constraint: Condition (i) for a credible elite punishment requires the elite to prefer the strategy ϕ_{H} whenever the majority defects from democratic federalism to administrative federalism with the central government setting taxes at the maximal tax rate, τ_{U}. In the **q**-Regime, this requires:

$$y(\tau_{\mathrm{U}}; \phi_{\mathrm{H}}) > y(\tau_{\mathrm{U}}; \phi_{\mathrm{L}}) \Leftrightarrow (\phi_{\mathrm{H}} - \phi_{\mathrm{L}})[g_{\mathrm{U}} - s_{\mathrm{e}}(\mathbf{q})][M_{\mathrm{e}}/N(\tau_{\mathrm{U}})] > \rho,$$

or

$$(M_{\mathrm{e}}/M) = \mu > \{\rho[N(\tau_{\mathrm{U}})/M]\}/\{(\phi_{\mathrm{H}} - \phi_{\mathrm{L}})[g_{\mathrm{U}} - s_{\mathrm{e}}(\mathbf{q})]\} \equiv \mu^{\min}(\mathbf{q}),$$

where μ is the fraction of majority residents who reside in the elite province. We use a strict inequality, assuming that the elite prefers to cooperate rather than defect, all else equal. Because condition (ii) for a credible elite punishment must also hold μ cannot be too large. Thus, $N(\tau_U) \geq M_e$ or dividing by M:

$$N(\tau_U)/M = \mu^{\max} \geq \mu = (M_e/M).$$

For high capture to be a credible punishment strategy for a given **q**, the constitutionally mandated population size of the elite province must satisfy the **q**-Border Constraint specified as

$$\mu^{\max} \geq \mu > \mu^{\min}(\mathbf{q}).$$

***q**-Assignment Constraint*: Condition (iii) for credible elite punishments requires that if the central government defects or punishes the elite, it continues to do so within the federal structure where provinces still have fiscal responsibilities—that is, within federalism with tax rates set at τ_U and not unitary governance. The binding constraint is the requirement that when the majority punishes the elite for defection, it does so using provinces rather than moving fully to centralized government provision.[39] For the majority's punishment strategy, condition (iii) requires:

$$\omega(\tau_U; \phi_H) > \omega(U) \Leftrightarrow s_U(\mathbf{q}) - [s_F(\mathbf{q}) - \phi_H \cdot \mu \cdot s_e(\mathbf{q})] > \phi_H \cdot \mu \cdot g_U,$$

which holds when **q** meets the constraint:[40]

$$\mathbf{q} > \mathbf{q}^{\min}(\mu; \phi_H) = (\phi_H \cdot \mu \cdot g_U)/[S \cdot \hat{a}(\mu; \phi_H)].$$

The constraint ensures that the majority's punishment strategy for an elite defection is not unitary governance. **q** cannot be set too high either. The maximum value of **q** that protects ϕ_H as a credible punishment strategy will be that **q** (= $\mathbf{q}^{\max}$) where the $\mu = \mu^{\min}(\mathbf{q})$ just holds for the constitutionally chosen value of μ. From the definition of $\mu = \mu^{\min}(\mathbf{q})$

$$\mathbf{q}^{\max}(\mu) = \{g_U \cdot (\phi_H - \phi_L) \cdot \mu - \rho \cdot [N(\tau_U)/M]\}/[(\phi_H - \phi_L)] \cdot \mu \cdot (S/a_e)].$$

The **q**-Assignment Constraint is defined by:

$$\mathbf{q}^{\max}(\mu) \geq \mathbf{q} > \mathbf{q}^{\min}(\mu; \phi_H).$$

q-Border Constraint*: In the q*-Regime, condition (i) for credible high capture by the elite requires:

[39] We also require that when the majority defects from the cooperative federal allocation, it defects to a regime still using provinces. This is needed so that the elite can still punish the majority for that defection. This requirement is specified as: $\omega(\tau_U; \phi_L) > \omega(U)$. We show in the full Technical Appendix available upon request that this requirement places less of a constraint on the minimal level of **q** than does the constraint needed for majority punishment for elite defection.

[40] From the definitions of $s_U(q)$, $s_F(q)$, and $s_e(q)$: $\hat{a}(\mu; \phi_H) = \mu \cdot [(1/a_m) - (1/a_e)] + (1 - m) \cdot [(1/a_u) - (1/a_m)] + (\mu \cdot \phi_H/a_e) > 0$.

$$y(\tau_U; q^*_H, \phi_H) > y(\tau_U; q^*_L, \phi_L) \Leftrightarrow \{\phi_H[g_U - s_e(q^*_H(\mu,\lambda))] - \phi_L\cdot[g_U - s_e(q^*_L(\mu,\lambda))]\}\cdot[M_e/N(\tau_U)] > \rho$$

or

$$(M_e/M) = \mu > \{\rho[N(\tau_U)/M]\}/\{\phi_H[g_U - s_e(q^*_H(\mu,\lambda))] - \phi_L\cdot[g_U - s_e(q^*_L(\mu,\lambda))]\} = \mu^{\min}(q^*_H),$$

where for pair of values of μ and q^*_H, there is an associated value of λ and thus of q^*_L which then allows us to specify a value for $\mu^{\min} = \mu^{\min}(q^*_H)$. Condition (ii) requiring elite political control again sets the upper bound, $\mu^{\max}$, defined as above. Together, the q^*-Border Constraint is specified as:

$$\mu^{\max} \geq \mu > \mu^{\min}(q^*_H).$$

q^*-Assignment Constraint: Again, condition (iii) for credible elite punishment must be met, now defined for majority chosen values of q^* as $\omega(\tau_U;\ \mu,\ q^*_H(\mu, \lambda),\ \phi_H) > \omega(U;\ q^*_U(\lambda))$ to ensure provinces survive the majority's decision to punish any defecting elite province.[41] From Table 1's specifications of $\omega(\tau_U;\ \mu,\ q^*_H(\mu, \lambda),\ \phi_H)$ and $\omega(U;\ q^*_U(\lambda))$, this requirement reduces to

$$[\upsilon(q^*{}_H(\mu,\lambda)) - p_H(\mu)\cdot q^*{}_H(\mu,\lambda)] - [\upsilon(q^*{}_U(\lambda)) - p_U\cdot q^*{}_U(\lambda)] > \phi_H\cdot\mu\cdot g_U,$$

where the left-hand side measures the difference between the consumer surplus earned by a typical majority resident under federalism with τ_U when the price of assigned services under federalism is $p_H(\mu)$ and that surplus earned by the majority resident under unitary governance when the price of a comparable service bundle under unitary governance is $p_{U.}$ Since $p_U > p_H(\mu)$, consumer surplus is greater under federalism. Because of elite capture, however, federalism also imposes an income loss $\phi_H\cdot\mu\cdot g_U$ on the average majority resident. The more important are assigned services to the majority ($\lambda\uparrow$), the larger becomes the gain in consumer surplus from moving to federalism from unitary governance. The value of λ where the inequality above just holds defines a minimal value for λ, denoted as $\lambda^{\min} = \lambda^{\min}(\mu)$. For each value of μ, there is an associated value of q^*_H that defines the minimal $q^*{}_H$ consistent with a credible elite punishment:

$$q^*_H(\mu,\ \lambda) > q^{*\min}_H(\mu) = q^*_H(\mu,\ \lambda^{\min}(\mu)).$$

As for the **q**-Regime here too there is an upper bound on majority demanded q consistent with a feasible federal allocation, now specified as an upper limit on λ. Given μ and the cost of high capture, ρ, there is a value of λ for which high capture is no longer a credible choice for the elite:

[41] The requirement that provinces survive the majority's decision to defect—$\omega(\tau_U;\ \mu,\ q^*_L(\mu,\ \lambda),\ \phi_L) > \omega(U;\ q^*_U(\lambda))$—is also met if the "tighter" requirement that provinces survive the majority's decision to punish is satisfied. This is shown in a full Technical Appendix available upon request].

$\mu^{\mathrm{max}} = \mu^{\mathrm{min}}(\lambda)$; $\lambda^{\mathrm{max}} = \lambda^{\mathrm{max}}(\mu)$. For each value of μ, define: $q_{\mathrm{H}}^{*\mathrm{max}}(\mu) = q_{\mathrm{H}}^{*}(\mu, \lambda^{\mathrm{max}}(\mu))$. Given μ, the q^*-Assignment Constraint is specified as follows:

$$q_{\mathrm{H}}^{*\mathrm{max}}(\mu) \geq q_{\mathrm{H}}^{*}(\mu, \lambda) > q_{\mathrm{H}}^{*\mathrm{min}}(\mu).$$

A.3. Specifying Feasible and Sustainable Democratic Federalism

We outline the proofs here. The algebraic details are provided in a Technical Appendix available from the authors upon request.

Lemma 1. **(**Credible elite punishments in the q-regime). For political economies satisfying the **q**-Border and **q**-Assignment Constraints, the high capture strategy will be a credible punishment strategy for the elite whenever the majority adopts a revenue-maximizing redistributive tax rate.

Outline of proof. To show sufficiency of the **q**-Border and **q**-Assignment Constraints for meeting the conditions for high capture to be a credible elite punishment, we first show that if the upper bound on the **q**-Assignment Constraint is met, then $g_{\mathrm{U}} - \rho/(\phi_{\mathrm{H}} - \phi_{\mathrm{L}}) \geq S\mathbf{q}/a_{\mathrm{e}} = s_{\mathrm{e}}(\mathbf{q})$ or $1 \geq \rho/(\phi_{\mathrm{H}} - \phi_{\mathrm{L}}) \cdot [g_{\mathrm{U}} - s_{\mathrm{e}}(\mathbf{q})]$. Multiplying both sides through by $N(\tau_{\mathrm{U}})/M$ then shows that $\mu^{\mathrm{max}} \geq \mu = M_{\mathrm{e}}/M$, and therefore $N(\tau_{\mathrm{U}}) \geq M_{\mathrm{e}}$, satisfying condition (ii) for a credible punishment. If the lower bound on the **q**-Border Constraint is met, then $\mu = M_{\mathrm{e}}/M > \{\rho[N(\tau_{\mathrm{U}})/M]\}/\{(\phi_{\mathrm{H}} - \phi_{\mathrm{L}})[g_{\mathrm{U}} - s_{\mathrm{e}}(\mathbf{q})]\}$, which implies $(\phi_{\mathrm{H}} - \phi_{\mathrm{L}})[g_{\mathrm{U}} - s_{\mathrm{e}}(\mathbf{q})][M_{\mathrm{e}}/N(\tau_{\mathrm{U}})] > \rho$. Adding $(Y - \tau_{\mathrm{U}})$ to both sides and rearranging terms gives $y(\tau_{\mathrm{U}}; \phi_{\mathrm{H}}) > y(\tau_{\mathrm{U}}; \phi_{\mathrm{L}})$, satisfying condition (i) for a credible elite punishment. Finally, from the lower bound of the assignment constraint, $\mathbf{q} > (\mu \cdot g_{\mathrm{U}} \cdot \phi_{\mathrm{H}})/[S \cdot \hat{a}(\mu)]$. Multiplying both sides by $S \cdot \hat{a}(\mu)$ and using the definitions for $s_{\mathrm{e}}(q)$, $s_{\mathrm{m}}(q)$, $s_{\mathrm{u}}(q)$, $s_{\mathrm{F}}(\mathbf{q})$, and $s_{\mathrm{U}}(\mathbf{q})$ will give $[s_{\mathrm{U}}(\mathbf{q}) - s_{\mathrm{F}}(\mathbf{q})] > \phi_{\mathrm{H}} \cdot \mu \cdot [g_{\mathrm{U}} - s_{\mathrm{e}}(\mathbf{q})]$. Adding $W + g_{\mathrm{U}} + \upsilon(\mathbf{q})$ to both sides and again rearranging terms and using the definitions of $\omega(\tau_{\mathrm{U}}; \phi_{\mathrm{H}})$ and $\omega(\mathrm{U})$, we have $\omega(\tau_{\mathrm{U}}; \phi_{\mathrm{H}}) > \omega(\mathrm{U})$. And since $\omega(\tau_{\mathrm{U}}; \phi_{\mathrm{L}}) > \omega(\tau_{\mathrm{U}}; \phi_{\mathrm{H}})$ for $\phi_{\mathrm{H}} > \phi_{\mathrm{L}}$, it also follows that $\omega(\tau_{\mathrm{U}}; \phi_{\mathrm{L}}) > \omega(\mathrm{U})$ as well. Thus, condition (iii) for a credible punishment is met. Necessity is shown in the complete Technical Appendix.

Lemma 2. (Credible elite punishments in the q^*-regime). For political economies satisfying the q^*-Border and q^*-Assignment Constraints, the high capture strategy will be a credible punishment strategy whenever the majority adopts a revenue-maximizing redistributive tax rate.

Outline of proof. For sufficiency, use the same argument as for Lemma 1 to show that if the upper bound of the q^*-Border Constraint is met, then $N(\tau_{\mathrm{U}}) \geq M_{\mathrm{e}}$ and condition (ii) for the a credible punishment is satisfied. If the lower bound of the q^*-Border Constraint is met, then by the argument presented above for Lemma 1, condition (i) for a credible punishment holds. Finally, to show that condition (iii) is met when the q^*-Assignment Constraint holds, note that because $\upsilon'(q) > 0$ and $\upsilon''(q) < 0$, the majority's preferences for q are single-peaked. If the assignment constraint is to hold for a given λ and $q_{\mathrm{H}}^{*}(\mu, \lambda)$ is the preferred level of q for those values of μ and λ, then any $q \neq q^*_{\mathrm{H}}(\mu, \lambda)$ gives the majority less utility than that obtained at $q^*_{\mathrm{H}}(\mu, \lambda)$. This is the case for the two values of q equal to either $q^{*\max}_{\mathrm{H}}(\mu)$ and $q^{*\min}_{\mathrm{H}}(\mu)$. By definition, $q^{*\max}_{\mathrm{H}}(\mu)$ and $q^{*\min}_{\mathrm{H}}(\mu)$ give a majority utility equal to that available under unitary governance—$\omega(U; q^*_{\mathrm{U}}(\lambda))$. Thus, $\omega(\tau_{\mathrm{U}}; \phi_{\mathrm{H}}) = \omega(\tau_{\mathrm{U}}; \mu, q^*_{\mathrm{H}}(\mu, \lambda), \phi_{\mathrm{H}}) > (\mathrm{U}; q^*_{\mathrm{U}}(\lambda)) = \omega(U)$[42]. Necessity is shown in the complete Technical Appendix.

Lemma 3. (Feasible federal constitutions). When the Border and Assignment Constraints of Lemmas 1 and 2 are met, the set of feasible constitutions allowing democratic federalism is smaller in the q^*-Regime than the **q**-Regime. The maximal size of the elite province is the same in both regimes.

Outline of proof. First, to show that $\mu^{\min}(q^*_{\mathrm{H}}) > \mu^{\min}(\mathbf{q})$ for common values $q^*_{\mathrm{H}} = \mathbf{q}$, we proceed by construction. Since $q^*_{\mathrm{H}}(\mu, \lambda) > q^*_{\mathrm{L}}(\mu, \lambda)$, it will be true that $0 > -\phi_{\mathrm{L}} \cdot [s_{\mathrm{e}}(q^*_{\mathrm{H}}(\mu, \lambda)) - s_{\mathrm{e}}(q^*_{\mathrm{L}}(\mu, \lambda))]$. Under the assumption that $q^*_{\mathrm{H}}(\mu, \lambda) = \mathbf{q}$, it will also be true that $(\phi_{\mathrm{H}} - \phi_{\mathrm{L}})\,[g_{\mathrm{U}} - s_{\mathrm{e}}(\mathbf{q}')] > (\phi_{\mathrm{H}} - \phi_{\mathrm{L}}) \cdot [g_{\mathrm{U}} - s_{\mathrm{e}}(q^*_{\mathrm{H}}(\mu, \lambda))] - \phi_{\mathrm{L}} \cdot [s_{\mathrm{e}}(q^*_{\mathrm{H}}(\mu, \lambda)) - s_{\mathrm{e}}(q^*_{\mathrm{L}}(\mu, \lambda))]$. Adding $[\phi_{\mathrm{L}}\, g_{\mathrm{U}} - \phi_{\mathrm{L}} g_{\mathrm{U}}]$ to the right-hand side and rearranging terms, implies that $\rho[N(\tau_{\mathrm{U}})/M]/\{\phi_{\mathrm{H}}[g_{\mathrm{U}} - s_{\mathrm{e}}(q^*_{\mathrm{H}}(\mu, \lambda))]\phi_{\mathrm{L}} \cdot [g_{\mathrm{U}} - s_{\mathrm{e}}(q^*_{\mathrm{L}}(\mu, \lambda))]\} = \mu^{\min}(q^*_{\mathrm{H}})\mu^{\min}(\mathbf{q}) = \{\rho[N(\tau_{\mathrm{U}})/M]\}/\{(\phi_{\mathrm{H}} - \phi_{\mathrm{L}})[g_{\mathrm{U}} - s_{\mathrm{e}}(\mathbf{q})]\}$ for common values of $q^*_{\mathrm{H}}(\mu, \lambda) = \mathbf{q}$. Second, when the Assignment Constraints hold, we know $[\upsilon(q^{*\min}_{\mathrm{H}}(\mu)) - p_{\mathrm{H}}(\mu) \cdot q^{*\min}_{\mathrm{H}}(\mu)] - [\upsilon(q^*_{\mathrm{U}}) - p_{\mathrm{U}} \cdot q^*_{\mathrm{U}}] = \phi_{\mathrm{H}} \cdot \mu \cdot g_{\mathrm{U}} = \mathbf{q}^{\min}(\mu) \cdot S \cdot \hat{a}(\mu; \phi_{\mathrm{H}})$, must hold from the definitions of $q^{*\min}_{\mathrm{H}}(\mu)$ and $\mathbf{q}^{\min}(\mu)$. Next, since $q^{*\min}_{\mathrm{H}}(\mu) \neq q^*_{\mathrm{U}}$ when demand curves slope downward, it will also be true that $[\upsilon(q^*_{\mathrm{U}}) - p_{\mathrm{U}} \cdot q^*_{\mathrm{U}}] > [\upsilon(q^{*\min}_{\mathrm{H}}(\mu)) - p_{\mathrm{U}} \cdot q^{*\min}_{\mathrm{H}}(\mu)]$ since q^*_{U} is optimal for the price $p_{\mathrm{U}.}$ Rearrange this expression and add $[p_{\mathrm{U}} \cdot q^*_{\mathrm{U}} - p_{\mathrm{H}}(\mu) \cdot q^{*\min}_{\mathrm{H}}(\mu)]$ to both sides. This gives $[p_{\mathrm{U}} - p_{\mathrm{H}}(\mu)] \cdot q^{*\min}_{\mathrm{H}}(\mu) > [\upsilon(q^{*\min}_{\mathrm{H}}(\mu)) - p_{\mathrm{H}}(\mu) \cdot q^{*\min}_{\mathrm{H}}(\mu)] - [\upsilon(q^*_{\mathrm{U}}) - p_{\mathrm{U}} \cdot q^*_{\mathrm{U}}] = \mathbf{q}^{\min}(\mu) \cdot S \cdot \hat{a}(\mu)$, using the step above. Finally, from the definitions of $p_{\mathrm{U}} = s'_{\mathrm{U}}(q)$, $p_{\mathrm{H}}(\mu) = s'_{\mathrm{F}}(q) - \phi_{\mathrm{H}}\mu . s_{\mathrm{e}}'(q)$, and $\hat{a}(\mu; \phi_{\mathrm{H}})$, we can show $[p_{\mathrm{U}} - p_{\mathrm{H}}(\mu)] \cdot q^{*\min}_{\mathrm{H}}(\mu) = [S \cdot \hat{a}(\mu; \phi_{\mathrm{H}})] \cdot q^{*\min}_{\mathrm{H}}(\mu)$. Thus, $[S \cdot \hat{a}(\mu; \phi_{\mathrm{H}})] \cdot q^{*\min}_{\mathrm{H}}(\mu) > \mathbf{q}^{\min}(\mu) \cdot [S \cdot \hat{a}(\mu; \phi_{\mathrm{H}})]$ from which it follows that $q^{*\min}_{\mathrm{H}}(\mu) > \mathbf{q}^{\min}(\mu)$. This completes the proof of Lemma 3.

Proposition 1. (Sustainable democratic federalism). For either the q- or q^*-Regime and political economy satisfying the appropriate Border and Assignment Constraints, there exists a grim trigger strategy equilibrium in which democratic federalism is sustainable. In that equilibrium:

(1) The central government majority chooses a level of redistributive transfers bounded between a maximal grant acceptable to the elite and

[42] As for the **q**-Regime, we will also require that the majority defection strategy even if the elite cooperates retains the use of provinces—that is, $\omega(\tau_{\mathrm{U}}; \phi_{\mathrm{L}}) > \omega(U)$. We show in the full Technical Appendix that this constraint is met if the requirement that provinces remain in place if majority punishes elite defection—that is, if $\omega(\tau_{\mathrm{U}}; \phi_{\mathrm{H}}) > \omega(U)$.

a minimal grant acceptable to the majority for appropriate elite and majority discount factors, and,

(2) The elite province(s) adopts the low capture strategy.

We prove existence of a democratic federalism as a subgame perfect Nash equilibrium for the **q**- and q^*-Regimes for at least some discount factor bounded as $0 < \delta \leq 1$. The proof moves in three steps.

Step 1. Specify the minimal grant (tax rate), the majority will accept and the maximal grant (tax rate) the elite will allow such that the majority and elite prefer democratic federalism when the other prefers democratic federalism. The minimal grant acceptable to the majority will be unambiguously positive, whereas the maximal grant that the elite will be pay will be less than that available from maximal taxation.

Step 2. Show that the elite's maximal grant is larger than the majority's minimal grant for either regime and for the potentially most favorable discount factor, $\delta = 1$:

$$\mathfrak{R}(\mathbf{q} \text{ or } \mathfrak{q}^*, \mathfrak{d} = 1) = \mathrm{g}^{\max}(\mathbf{q} \text{ or } \mathfrak{q}^*,\ 1) - \mathrm{g}^{\min}(\mathbf{q} \text{ or } \mathfrak{q}^*,\ 1) > 0.$$

If $\mathfrak{R}(\mathbf{q} \text{ or } \mathfrak{q}^*,\ \delta = 1) > 0$, then there is a economically feasible fiscal policy in the annual policy game that will sustain democratic federalism, at least for infinitely far-sighted majority and elite residents ($\delta = 1$).

Step 3. Show that the more general specification $\mathfrak{R}(\mathbf{q} \text{ or } \mathfrak{q}^*,\ \delta) = \mathrm{g}^{\max}(\mathbf{q} \text{ or } \mathfrak{q}^*,\ \delta) - \mathrm{g}^{\min}(\mathbf{q} \text{ or } \mathfrak{q}^*,\ \delta)$ is a continuous function of δ, implying there is a $\delta < 1$(though perhaps only slightly <1), where $\mathfrak{R}(\mathbf{q} \text{ or } \mathfrak{q}^*,\ \delta) > 0$ continues to hold.

The full details of the algebra for Steps 1–3 as well as the specifications for $g^{\max}(\mathbf{q}, \delta)$, $g^{\min}(\mathbf{q}, \delta)$, $g^{\max}(q^*, \delta)$, and $g^{\min}(q^*, \delta)$ are provided in a longer Technical Appendix available upon request.

References

Acemoglu, Daron, and James Robinson. 2001. "A Theory of Political Transitions," 91 *American Economic Review* 938–63.

Akerlof, George, and Rachel Kranton. 2005. "Identity and the Economics of Organizations," 18 *Journal of Economic Perspectives* 9–32.

Anderson, Liam, and Gareth Stansfield. 2005. "The Implications of Elections for Federalism in Iraq: Toward a Five-Region Model," 35 *Publius* 359–82.

Barber, James. 1967. *Rhodesia: The Road to Rebellion*. New York: Oxford University Press.

Bardhan, Pranab, and Dilip Mookherjee. 2006. "Pro-poor Targeting and Accountability of Local Governments in West Bengal," 79 *Journal of Development Economics* 303–27.

Besley, Timothy, and Stephen Coate. 1997. "An Economic Model of Representative Democracy," 112 *Quarterly Journal of Economics* 85–114.

Boix, Carles. 2003. *Democracy and Redistribution*. Chicago, IL: University of Chicago Press.

Buchanan, James, and Geoffrey Brennan. 1980. *The Power to Tax: Analytic Foundations of a Fiscal Constitution*. New York: Cambridge University Press.

Conley, John, and Akram Temimi. 2001. "Endogenous Enfranchisement When Groups' Preferences Conflict," 109 *Journal of Political Economy* 79–102.

Constitution of the Republic of South Africa. 1996. Act 108 of 1996.

Dawisha, Adeed, and Karen Dawisha. 2003. "How to Build a Democratic Iraq," 82 *Foreign Affairs* 36–45.

de Figueiredo, Rui, and Barry Weingast. 2005. "Self-Enforcing Federalism," 21 *Journal of Law, Economics, and Organizations* 103–35.

Dixit, Avinash. 1997. *The Making of Economic Policy: A Transaction-Cost Politics Perspective*. Cambridge, MA: The MIT Press.

Duggan, Mark. 2000. "Hospital Ownership and Public Medical Spending," 115 *Quarterly Journal of Economics* 1343–75.

Ferejohn, John, and Barry Weingast. 1992. "A Positive Theory of Statutory Interpretation," 12 *International Review of Law and Economics* 263–79.

Fitts, Michael, and Robert Inman. 1992. "Controlling Congress: Presidential Influence in Domestic Fiscal Policy," 80 *The Georgetown Law Journal* 1737–85.

Gordon, Nora. 2004. "Do Federal Funds Boost School Spending? Evidence from Title I," 88 *Journal of Public Economics* 1771–92.

Gruber, John, and Emmanuel Saez. 2002. "The Elasticity of Taxable Income: Evidence and Implications," 84 *Journal of Public Economics* 1–32.

Hawker, Geoffrey. 2000. "Political Leadership in the ANC: The South African Provinces, 1994-1999," 38 *The Journal of Modern African Studies* 631–58.

Inman, Robert P., and Daniel L. Rubinfeld. 2011. "Federal Institutions and the Democratic Transition: Learning from South Africa,"National Bureau of Economic Research. Working Paper 13733 (Revised).

Karlan, Dean, and Jonathan Zinman. 2008. "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts,"Working Paper, Yale University and Dartmouth College.

Lijphart, Arend. 1984. *Democracies: Patterns of Majoritarian and Consensus Government in Twenty-One Countries*. New Haven, CN: Yale University Press.

Lizzeri, Alessandro, and Nicola Persico. 2004. "Why Did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform'," 119 *Quarterly Journal of Economics* 707–65.

Lodge, Tom. 2005. "Provincial Government and State Authority in South Africa," 31 *Journal of Southern African Studies* 737–53.

Madison, James. 1788. "Federalist No. 51," in The Federalist Papers. New York: Bantam Books.

Musgrave, Richard. 1959. *The Theory of Public Finance*. New York: McGraw Hill.

Muthien, Yvonne, and Meshack Khosa. 1998. "Demarcating the New Provinces: A Critical Reflection on the Process," in Y. Muthien and M. Khosa, eds., *Regionalism in the New South Africa*. Brookfield, WI: Ashgate.

Myerson, Roger. 2006. "Federalism and Incentives for Success of Democracy," 1 *Quarterly Journal of Political Science* 3–23.

Oates, Wallace. 1999. "An Essay on Fiscal Federalism," 37 *Journal of Economic Literature* 1120–49.

Przeworski, Adam, Michael Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1960-1990*. New York: Cambridge University Press.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence From a Central Government Transfer Program in Uganda," 119 *Quarterly Journal of Economics* 679–709.

Riker, William. 1964. *Federalism: Origin, Operation, Significance*. Boston, MA: Little Brown.

Riker, William. 1980. "Implications from the Disequilibrium of Majority Rule for the Study of Institutions," 74 *American Political Science Review* 432–46.

Romer, Thomas, and Howard Rosenthal. 1979. "Bureaucrats vs. Voters: On the Political Economy of Resource Allocation by Direct Democracy," 93 *Quarterly Journal of Economics* 562–87.

Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Steytler, Nico, and Johann Mettler. 2001. "Federal Arrangements as a Peacemaking Device During South Africa's Transition to Democracy," 31 *Publius: The Journal of Federalism* 93–106.

Tiebout, Charles. 1956. "A Pure Theory of Local Public Expenditures," 60 *Journal of Political Economy* 415–24.

Waldmeir, Patty. 1997. *Anatomy of a Miracle: The End of Apartheid and the Birth of the New South Africa*. New York: W.W. Norton and Company.

Weingast, Barry. 1995. "The Economic Role of Political Institutions: Market-Preserving Federalism and Economic Development," 11 *Journal of Law, Economics, and Organization* 1–31.

Weingast, Barry. 2009. "Second Generation Fiscal Federalism: The Implications of Fiscal Incentives," 65 *Journal of Urban Economics* 279–93.

Williams, Katherine, and Charles O'Reilly III. 1998. "Demography and Diversity in Organizations: A Review of Forty Years of Research," 20 *Research in Organizational Behavior* 77–140.

Williamson, Oliver. 1983. "Credible Commitments: Using Hostages to Support Exchange," 73 *American Economic Review* 519–40.

Wittenberg, Martin. 2006. "Decentralization in South Africa," in Pranab Bardhan and Dilip Mookherjee, eds., *Decentralization and Local Governance in Developing Countries*. Cambridge, MA: MIT Press.