FEDERAL INSTITUTIONS AND THE DEMOCRATIC TRANSITION:
Learning from South Africa

by

Robert P. Inman and Daniel L. Rubinfeld

Draft: May 15, 2006

ABSTRACT

We present a political economy model of a peaceful transition from autocracy to democracy, motivated by, and calibrated to, South Africa at the time of its own successful transition from white, elite rule under apartheid to today’s multi-racial democracy. The primary barrier faced by South Africa in its transition is symptomatic of a central barrier to peaceful transitions generally: How can the emerging majority credibly promise not to economically exploit the once ruling elite? South Africa found its answer in a “democratic federalism” constitutional system. The constitution creates an annual policy game where the new majority and the old elite each credibly control one policy instrument (fiscal policy) of central importance to the other. The majority-controlled central government controls general tax policy, while provinces, at least one of which is politically controlled by the old elite, are assigned responsibility for providing important redistributive services. In effect, the federal constitution creates a policy “hostage game” between the new majority and the old elite in which each can check policy abuses by the other. We show that for plausible specifications of the political economy, this annual policy game has a stable, less-than-fully exploitative stationary equilibrium that the elite may prefer to current autocratic rule. The model is calibrated to the South African political economy at the time of its transition and “predicts” the actual structure of redistributive fiscal policy now in place. Between forms of democracies, the federal regime is preferred to unitary governance by both the elite and the new ANC majority. At the time of the transition, this federal democracy is shown to be unambiguously preferred to autocracy by the then oppressed majority and conditionally preferred by the then ruling elite. For the elite, the move to democracy means higher tax rates but also higher rates of economic growth; democracy will be preferred to apartheid, if the elite’s (real) rate of time preference is less than their rate of return from the transition.

Robert P. Inman is the Richard K. Mellon Professor, Finance and Economics, the Wharton School, University of Pennsylvania, Philadelphia, PA 19104, inman@wharton.upenn.edu

Daniel L. Rubinfeld is the Robert L. Bridges Professor of Law and Professor of Economics, University of California, Berkeley, CA 94720, rubinfeld@law.berkeley.edu

The work here has benefitted significantly from comments by seminar participants at Berkeley, Cornell, NYU, and Penn. Particular thanks are due to Carolyn Ballay, Grant Long, and Jon Stott for extraordinarily valuable research assistance.
What if the poor, on the ground of their being a majority, proceed to divide among themselves the possessions of the wealthy – will not this be unjust? . . . (We may reply in return) . . . Whenever a majority of any sort, irrespective of wealth or poverty, divides among its members the possessions of a minority, that majority is obviously ruining the state. Aristotle, *The Politics of Aristotle*, Ernest Barker (ed.), p. 122.

Poverty is the cause of the defects of democracy. This the reason why measures should be taken to ensure a permanent level of prosperity. This in the interest of all classes, including the prosperous themselves; and therefore the proper policy is to accumulate any surplus revenue in a fund, and then to distribute this fund in block grants to the poor. Aristotle, *The Politics of Aristotle*, Ernest Barker (ed.), p. 269.

I. Introduction

Democratic governance is clearly preferable to its alternative, autocratic rule. While the importance of democracy as a pre-condition for improved economic growth remains an open research question, there is little doubt that stable democratic institutions provide basic protections for essential human and economic rights and opportunities to participate, both directly and indirectly, in the management of societal affairs.\(^1\) There appear to be significant international advantages to stable democracies as well. Democracies are less likely to initiate international aggressions and more likely to support free trade agreements.\(^2\) Given the case for democratic rule, the issue then becomes how best to facilitate the transition from autocracy to democracy.

The paper presents a model of the political economy of democratic transition. The motivation is drawn from the recent history of South Africa’s successful (to this point) transition from a white, elite autocracy to a multi-racial, majority-rule democracy, as described very briefly in Section II. As Aristotle stressed in *The Politics*, the barriers faced by South Africa in its move to democracy are symptomatic of transitions barriers generally. The barriers arise because the once oppressed majority that is destined to dominate democratic national politics may not be able to credibly promise the current ruling elites that they

---

\(^1\) On democracy and economic growth see the debate between Acemoglu, Johnson, and Robinson (2001) and Glaeser, et. al. (2004). On democracy as an institution to promote economic fairness and personal liberties, see Dreze and Sen (1989), Rodrik (1999), and Shleifer et. al. (2004). For evidence on democracy and political participation, see Verba and Nie (1972), Frey and Stutzer (2000).

\(^2\) On democracy and war, see Bueno de Mesquita, et. al. (1999). On democracy and trade, see Mansfield, Milner, and Rosendorff (2002).
will not be economically exploited in the new regime. If there are not sufficient guarantees and the current elites have military power, they may opt to resist rather than support the transition.

The challenge is to find a democratic constitution that can credibly offers such protections. Section III responds to this challenge by showing that under well-defined conditions a federal constitution – defined as system of government in which a nationally elected central government and locally elected provincial governments each hold significant fiscal powers for the financing and provision of essential public services – can credibly provide the needed economic protections. With a federal constitution in place, the transition to democracy can work, even when the unitary form of governance – rule by a single national majority – might fail to facilitate the transition.

The formal analysis is then applied to understand South Africa’s own peaceful transition to democracy. In Section IV, we find that democratic federalism offers a sustainable and amendment-proof constitution for the new South Africa. Section V shows the resulting federal constitution was the economically preferred alternative at the time of South Africa’s transition, both for the country’s new majority and for a majority of the country’s once ruling elite as well. Section VI concludes, placing this work within the wider study of endogenous institutions.

II. South Africa’s Transition: A Brief History

The more than forty years of apartheid rule in South Africa began in 1948 with the election of the white National Party (NP) to run the national government. In August, 1962 Nelson Mandela and other opposition African National Congress (ANC) leaders were arrested. By the fall of 1986, it was apparent to the moderate leadership in both the National Party and the ANC that, after years of violent conflicts, negotiations towards full South African democracy was the only way forward. In September 1989, having to choose among the alternatives of renewed repression and possible war or the status quo or a move towards compromise and a democratic South Africa, NP leader F.W. De Klerk found the latter alternative the most
As a condition for majority Black rule, de Klerk and the National Party insisted on constitutional protections for the civil liberties and property of the white minority.

Following negotiations that began in earnest in 1990 the ANC accepted in principle a constitutional bill of rights that protected individual civil and property rights. Lacking a sufficient number of adequately trained Black civil servants, the ANC also accepted the National Party’s demand to employ current white bureaucrats. More problematic was the need to ensure private ownership of firms and to prevent excessively high redistributive taxes.

The National Party proposed the devolution of taxing and spending powers to nine provincial governments, at least two of which would be controlled by the NP. The ANC in turn proposed four provinces, with all taxing powers centralized and all provinces effectively controlled by an ANC majority. At this point a third party entered the constitutional negotiations: the Inkatha Freedom Party (IFP) representing the Zulu nation, led by Chief Mangosuthu Buthelezi. Chief Buthelezi demanded an Inkatha-controlled province containing the historic lands of the Zulu nation to be called KwaZulu-Natal, and the devolution of significant spending powers to the provinces. The commitment of significant governmental transfers to the provinces was crucial for Buthelezi, since this would maintain the lucrative homeland payments he had enjoyed under apartheid. Given the NP’s ability to sustain apartheid and Inkatha’s threat to boycott the planned 1994 elections, the ANC agreed in principle to the NP’s and the IFP’s demands for nine provinces with significant spending responsibilities and sizeable central government transfers.

The interim constitution signed in the winter of 1994 codified this in principle agreement, but decisions as to how to actually assign taxing and spending powers and the level of intergovernmental

---

transfers were postponed until after the 1994 election. By any measure, the election was a success. Voting was completed without incident. Nelson Mandela was elected President and the ANC won 62.7 percent of the national vote. The National Party polled 20.4 percent nationally, but as planned, the NP won a majority of provincial legislative seats in the new province of the Western Cape with 53 percent of the provincial vote. Similarly, and as planned, Chief Buthelezi and the Inkatha Freedom Party won 51 percent of the provincial vote in KwaZulu-Natal. Unexpectedly, however, the rural (white farmer) Northern Cape fell to the ANC in a close election. The six remaining provinces went decisively to the ANC.

With elections completed, attention turned to drafting a final constitution. The interim constitution created a Financial and Fiscal Commission (FFC) charged with the task of assigning tax and spending responsibilities to the national and provincial governments and designing a formula for intergovernmental transfers capable of ensuring adequate funding for provincial government services. The composition of the Commission was equally divided between ANC and NP representation, with two additional representatives from KwaZulu-Natal. The Commission proposed an assignment of all important taxing powers to the central government, but delivery and spending responsibility for important redistributive services – K-12 education, primary health care, and social security (welfare) – became the responsibility of the new provinces. Without taxing powers, however, funding for these redistributive services required centrally decided intergovernmental transfers. The Commission’s recommendations were incorporated into the final constitution, which was unanimously approved on October 11, 1996.

Though the exact details of the final constitution – namely the size of provinces, their exact responsibilities, and the final level of intergovernmental transfers – apply only to South Africa, we argue that the constitution’s fundamental features – provinces, shared responsibilities, and transfers – apply more broadly. Section III offers a general argument for federalism as a constitutional form to facilitate the democratic transition. Sections IV and V then show how the South African constitution stands as an application of our central proposition.
III. Federal Institutions and the Democratic Transition

A. Overview: This section specifies the conditions under which an appropriately designed federal constitution might facilitate the peaceful transition from autocracy to democracy. The underlying model is a sequential game with two stages. In the first constitutional stage, a constitution is chosen requiring the joint approval of both the elite minority and the poor majority – that is, both parties have a veto over the new constitution. The constitution can specify either a unitary, or fully centralized, democracy with a single national democratic government setting all policies, or a federal, or partially decentralized, democracy where policy responsibilities are shared between the national government and constitutionally created provinces. In the case of a federal democracy, the constitution will specify provincial boundaries, also assign tax, spending, and regulatory responsibilities between central and provincial governments and perhaps set a minimal level of service provision that each province must provide. If no, or minimal, taxing powers are given to the provinces, then the constitution must allow for intergovernmental transfers from central government revenues to the provinces. Both unitary and federal constitutions will allow amendments, subject to approval by a constitutionally required (possibly, super-) majority.

In the second policy stage, the parties play an annual policy game in which the central government controlled by the poor majority chooses tax rates, spending, and regulations and, in a federal democracy, a level of intergovernmental transfers (possibly zero) to the provinces. Provincial governments are elected by a majority of residents in each province; some provinces may be politically controlled by the once ruling elite. Provinces then allocate their resources to their constitutionally-assigned responsibilities. Elite-run provinces may choose to meet, or not, their constitutional responsibilities. In particular, the elite may allocate some of their assigned intergovernmental transfers meant for redistributive services to elite services or to elite tax relief, allocations we call “elite capture.” Elite capture is not possible in unitary democracies.

The sequential game is solved through backward induction. The second-stage policy game is specified for fiscal policies, in particular, the financing and provision of redistributive (targeted) goods and
services—for example, education, health care, housing, land reform, and income transfers. When solving the policy game under \textit{unitary democracy}, policies will be set a nationally elected poor majority so as to maximize the welfare of a typical majority citizen. The majority will impose a national tax on the once ruling minority elite, the proceeds of which will be allocated to provide redistributive services and transfers to the poor majority through a national bureaucracy. The tax on the elite can be interpreted broadly as a tax on labor income, capital, firm profits (in the extreme, nationalization), or land (in the extreme, land expropriation). In response to national taxation, the elite are free to leave the country and take all their human capital with them plus any of their capital income or land rents the constitution allows. Given elite mobility, there will be an elite tax rate which maximizes redistributive revenues. We assume that the majority will choose this revenue maximizing, fully exploitative, tax rate under unitary democracy. It is this possibility that discourages the elite from embracing unitary democracy.

In the case of a federal constitution, the second stage policy game has the central government, again controlled by the poor majority, setting the national tax rate with the proceeds allocated in whole or in part to intergovernmental transfers to the provinces. Provincial governments, some of which are politically controlled by the old elite, will then allocate their revenues to constitutionally assigned redistributive services or to elite-only services via elite capture. Given the fact of elite capture, the poor majority will prefer unitary democracy unless there is some compensating reason to use provinces as providers of redistributive services. In our model there is. Elite residents are the efficient providers of such services, but these efficiencies are only available in elite-run provinces; in majority-run provinces or under unitary democracy, elite providers take their talents as teachers, doctors, nurses, and contractors to the private sector—or they leave the country.

It is useful to distinguish two versions of federal democracy. The first, \textit{democratic federalism}, has the central government set its national tax rate at less than the fully exploitative, revenue-maximizing tax rate. The second, \textit{administrative federalism}, has the central government set the national tax rate at the revenue-maximizing tax rate, but continue to allocate the proceeds of that tax to provinces as
intergovernmental transfers. Under administrative federalism, provinces are still valuable to the ruling poor majority as low cost providers of redistributive services, at least in the elite-run provinces. Even though national redistributive tax rates are maximal, elite residents prefer administrative federalism to unitary democracy since elite capture is still possible. Of course, the elite’s first choice would be democratic federalism where national tax rates are below maximal tax rates and elite capture is also possible. Which form of federalism actually obtains is decided by the poor majority and is an endogenous outcome of the second-stage policy game.

Only constitutions which are sustainable (i.e., self-enforcing) in the second-stage policy game will be considered by the ruling elite and the future poor majority as credible constitutions when playing the first-stage constitutional game. Given the assumption below that the elite loses control of the army with the transition, the unitary constitution is always sustainable. We therefore focus our attention on finding a sustainable federal constitution which we specify along two, or perhaps, three dimensions. The first dimension is provincial borders, specified here by the share of the majority population (μ) initially assigned to live with elite residents in what we will call the “elite province(s).” The second dimension is the constitutional assignment of provincial service responsibilities. At a minimum, assignment specifies the services that provinces must provide. For this paper, we index assigned services by a parameter λ which measures how important assigned services are for majority resident welfare; K-12 education has a high λ while street lighting has a relatively low λ. Constitutional assignment might also include a third dimension, one that specifies a level of productive inputs per majority resident (q) to produce the assigned services – for example, many U.S. state constitutions specify a minimal level of educational spending which all local school districts must provide. The first-stage then selects a preferred constitution – values of μ, λ, and perhaps q – from among the set of all possible values of (μ, λ, q) that generate politically sustainable constitutions.

Six assumptions underlie the analysis. First, the poor majority bears sufficiently large costs in lost economic and democratic rights under autocracy that it prefers any democratic constitution, whether unitary
or federal, to autocracy. Second, the oppressed majority does not have sufficient military strength to defeat the current autocratic regime and to then unilaterally impose a constitution. Thus, the majority and the ruling elite must both agree to a new democratic constitution; otherwise the autocratic regime and all associated conflicts and costs will remain. Third, once a democratic constitution is in place, the current ruling elite turns over control of its military to the new majority. As a result, neither elite secession nor a military coup d’état restoring the elite to power is possible. Fourth, the civil and political rights of the elite will be protected in the democratic regime; those not convicted of crimes against humanity during autocracy will be able to exit the country if they so choose. Fifth, the majority and the elite are fully informed as they negotiate the transition to democracy. Sixth, negotiators for the elite and the majority wish to negotiate a constitution which best protects the long-run (dynastic) interests of their average constituents.

B. The Annual Policy Game: In a democratic economy, each of the $M$ “poor” majority residents earn $W$ a year, while each of the $N$ “elite” residents earn $Y > W$. Initially, there are $N_0$ elite residents who are free to leave the country, and may choose to do so depending on the tax rate imposed by the ruling majority. Elite residents pay a nationally chosen per-capita tax, $\tau$, with the proceeds redistributed to majority residents only as a per capita grant, $b$, and/or as inputs $q$ to provide majority-valued public goods.\footnote{This redistributive national tax rate should be interpreted as the incremental national tax rate paid in addition to all other taxes used to fund such pure public goods as national defense, public infrastructure, and domestic security.} If allowed, provincial taxes are used to finance private goods and are thus treated for simplicity as an expenditure paid for from each province’s residents’ “after-redistributive-tax-and-transfer” incomes.

Given the possibility of exit, the equilibrium number of elite taxpaying residents depends not only on the initial elite population, but also the redistributive tax burden per elite resident. For simplicity we combine the elite’s two responses to high taxation, leaving the country or curtailing taxable activities, into
a single relationship of the form: \( N(\tau) = N_0 - \beta \tau \). Aggregate redistributive tax revenues are given by \( \tau N(\tau) = N_0 \tau - \beta \tau^2 \), implying a redistribution maximizing tax rate of \( \tau_0 = N_0/2\beta \), the rate that will be chosen under unitary democracies.

Revenues raised from redistributive taxation will be allocated to provide public services and a per-capita basic grant. The constitution will require a common and uniform level of public services inputs \( q \) to be provided to all majority residents specified as \( q = a(X/M) \) where \( (X/M) \) is public employees \( (X) \) per majority resident \( (M) \) and \( a \) is employee productivity measured by years of training. There are three classes of public employees: elite providers with \( a_e \) years of training, majority providers with \( a_m \) years of training, and untrained majority providers with \( a_u \) years of training, where, \( a_e > a_m > a_u \). The input bundle then provides assigned public services to the typical majority resident specified as \( v = v(q) \); we assume \( v'(q) > 0 \) and \( v''(q) < 0 \).

The central government is assumed to mandate a common civil service wage \( S \) paid equally to all public employees. The cost per majority resident of providing the common input bundle \( q \) is \( s(q) \), which is inversely related to worker training: \( s_e(q) < s_m(q) < s_u(q) \). The central government’s budget constraint is specified generally as:

\[
s(q) + b = g(\tau) = [\tau N(\tau) - Z]/M,
\]

where \( g(\tau) \) is aggregate redistributive spending available for each majority resident, and \( Z > 0 \) is a payment to any special interest groups initially capable of blocking the transition. \( Z < 0 \) can be interpreted as

----------------------------------

8 A more general specification might define redistributive revenues as \( \tau N(\tau) Y(\tau) \), where population is specified by \( N(\tau) = n_0 - n_1 \tau \) and taxable income by \( Y(\tau) = Y_0 - \psi \tau \); see Haughwout, et. al. (2004) for estimates of \( N(\tau) \) and Gruber and Saiz (2002) for estimates of \( Y(\tau) \).

9 To ease notation, we assume a common wage across all public employees regardless of skill level. The common wage assumption could be justified by simply assuming all public employees face the same private market alternative, that is, observed employee skill differentials are unique to providing redistributive public services. Separate wages for the three sets of workers could be allowed as long as there remains a public sector cost advantage to using elite employees.

10 \( s_e(q) = S'(X/M) = S'(q/a_e) \) using elite providers; \( s_m(q) = S'(q/a_m) \) using majority providers; and \( s_u(q) = S'(q/a_u) \) using untrained providers.
international aid, where such assistance allows higher grants and services for the new majority and lower
taxes on the elite, both of which facilitate transition.

Under unitary democracies, redistributive taxes are raised centrally, the special interest payment is
made, and funds are then distributed by a central government bureaucracy directly to the majority citizens,
first through the central government’s provision of valued services and then via the per-capita grant. The cost
of providing $q$ using a single unitary government, $s_u(q)$, is a weighted average of the cost associated with
trained and untrained providers.\footnote{Specifically, $s_u(q) = m \cdot s_m(q) + (1 - m) \cdot s_e(q)$, where $m$ is the
share of majority residents serviced by trained majority providers. We are assuming either that well-trained,
elite public providers choose not to provide redistributive services under a majority-run unitary democracy,
or if they do, they are no more productive than the average majority provider, perhaps allocating a portion of
their work day to private clinics and schools for elite residents.} The redistributive grant to each majority resident
under unitary democracy will be $b = g(\tau) - s_u(q)$. Institutionally, unitary democracies can be established directly by the
constitution (no provinces) or exist \textit{de facto} under a federal constitution by simply allowing the central government to
establish a separate bureaucracy using majority providers and sending no grant monies to the provinces. In
the \textit{de facto} case, provinces exist in name only.

Under federal democracies, redistributive taxes are raised centrally, the special interest payment is
made, and all remaining funds are distributed as a uniform intergovernmental transfer of $g(\tau)$ per majority
resident from the central government to the provinces. From this uniform grant, provinces must provide
assigned services, after which all remaining funds are distributed by the province as a per-majority person
grant. Residents in majority provinces receive their public service input bundle at a cost of $s_m(q)$ and a basic
grant $b = g(\tau) - s_m(q)$. Majority residents residing in an elite controlled province will also be given $q$, now
provided by elite public employees, at a cost of $s_e(q)$, and in principle, a basic grant of $b = g(\tau) - s_e(q)$. The
input bundle $q$ is uniform across all provinces, set either at the constitutional stage (“exogenous $q$”) or, if
outside enforcement is not possible, then chosen by the majority controlled central government (“endogenous
$q$”). If they wish, we assume the majority can costlessly enforce $q$ in both majority and elite controlled
provinces.

What the central government cannot control is how the elite province spends its basic grant: \( b = g(\tau) - s_e(q) \). Elite provinces may re-allocate some of the basic grant from majority to elite residents, most likely through the provision of provincial services (roads, parks, security) in elite residential neighborhoods. The basic grant actually received by majority residents in elite provinces will therefore be \( (1 - \varphi)[g(\tau) - s_e(q)] \), where \( \varphi \) is the rate of capture by the elite. The rate of capture will be one of two values: \( \varphi^L \) and \( \varphi^H \). The lower value (\( \varphi^L \)) is set by the ability of the majority residents in the elite province to organize costly protests when capture is observed, perhaps by observing service inequities between elite and majority neighborhoods. Once organized, these protests are assumed to impose a cost \( \rho \) on each elite resident. The elite provincial leadership will therefore push \( \varphi \) to the point where capture is just observed, \( \varphi^L \). The upper value \( \varphi^H \) is set by the ability of majority residents to exit the elite province for a majority run province.\(^{12}\) In the analysis that follows we assume that some capture is always possible before local protests occur, so that \( 0 < \varphi^L < \varphi^H \leq 1 \). We assume that \( \varphi^L \) and \( \varphi^H \) are common knowledge but not contractible; in particular, the outcome \( \varphi = 0 \) is not legally enforceable.\(^{13}\) Capture is not a problem in majority controlled provinces.

We can now detail the incidence of public budgets for majority and elite residents in unitary and federal democracies. This final allocation defines the payoffs of the annual policy game, shown in Table 1. Under democratic federalism, \( F \), there are two cases to consider, one in which the elite leadership of the elite-
run province captures only the lower value $\varphi^L$ of the basic grant, and a second case where the elite leadership chooses to capture the upper value, $\varphi^H$. The welfare of a majority resident in an elite province will be:

$$\omega_*(F; \varphi^{L \lor H}) = W + (1 - \varphi^{L \lor H})[g_e - s_e(q)] + \lambda \cdot u(q),$$

depending on whether the elite chooses low or high capture, where $[g_e - s_e(q)]$ is the basic grant received by a majority resident, and $\lambda \cdot u(q)$ is the typical majority resident’s utility from the uniform provision of the public service input bundle $q$. The input bundle $q$ provides a public service $u(q)$ valued by the majority at rate $\lambda > 0$. Constitutional assignment of services to the provinces is equivalent to selecting a value of $\lambda$ for the stage-two policy game; important services (education) have a high $\lambda$, less important services (street lighting) a low $\lambda$. Since $u'(q) > 0$ and $u''(q) < 0$, majority preferences are single-peaked in $q$. Welfare for majority residents in a majority-controlled province is specified similarly, except there is no risk of capture:

$$\omega_m(F) = W + [g_m - s_m(q)] + \lambda \cdot u(q).$$

The average majority resident’s welfare is a weighted average of the welfare of majority residents residing in elite province(s) and the welfare of majority residents residing in majority province(s), where the weights are the share of all majority residents residing in the elite province, $\mu = (M_e/M)$, and the majority province(s), $(1 - \mu) = [1 - (M_e/M)]$:

$$\omega(F; \varphi^{L \lor H}) = \mu \omega_*(F; \varphi^{L \lor H}) + (1 - \mu)\omega_m(F).$$

The shares are exogenous to the policy game, set by provincial boundaries and decided at the first constitutional stage. Majority resident welfare under democratic federalism is therefore:

$$\omega(F; \varphi^{L \lor H}) = W + g_e[1 - \varphi^{L \lor H}\mu] - s_e(q) + \varphi^{L \lor H}\mu \cdot s_e(q) + \lambda \cdot u(q), \quad (1L, 1H)$$

---

14 Majority residents living in the elite province will only leave that province if $\varphi$ chosen by the elite provincial leadership exceeds the upper bound, $\varphi^H$, in which case all of $M_e$ residents exit. Since this will mean no grants to the elite province and thus no capture, the elite leadership will keep $\varphi \leq \varphi^H$ and the majority population stays at $M_e$. Thus we assume $\mu = (M_e/M)$ is given in the policy stage, having been set at the constitutional stage by the drawing of provincial lines.
where \( s_F(q) \) is the national average cost of service input provision under federalism.\(^\text{15}\) The majority prefers low capture.

Under *administrative federalism*, \( A \), the central government’s poor majority selects an equilibrium tax rate of \( \tau_U \) and a redistributive grant equal to \( g_U \), just as they would under a unitary system. However, because \( q \) and \( b \) are still administered by the provinces, elite capture is possible. The weighted average welfare for the typical majority resident under administrative federalism becomes:

\[
\omega(A; \varphi^{L, or H}) = W + g_U [1 - \varphi^{L, or H} \mu] - s_F(q) + \varphi^{L, or H} \mu \cdot s_e(q) + \lambda \cdot \upsilon(q),
\]

Here too, the majority prefers low capture.

Under *unitary democracy*, \( U \), simple majority rule will ensure the national redistributive tax rate will be set at its maximal rate (\( \tau^U \)). In unitary democracy there are no provinces, no elite capture, and no elite providers of mandated services; \( q \) is provided by the central government only. The average majority resident’s welfare is therefore given by:

\[
\omega(U) = W + [g_U - s_U(q)] + \lambda \cdot \upsilon(q),
\]

where \( W \) is the majority resident’s market wage, \([g_U - s_U(q)]\) is the basic grant received by the majority resident, and \( \lambda \cdot \upsilon(q) \) is the typical majority resident’s utility from valued services.

Under democratic federalism, the elite residents are assumed to reside only in elite-run provinces. They pay redistributive taxes to the central government and receive from their elite-run provincial governments any captured intergovernmental transfers not allocated to constitutionally mandated or centrally chosen service inputs \( q \). Elite residents do not receive services from funded inputs, \( q \). Their public services are funded separately from elite incomes available after paying redistributive taxation. The elite resident's

\(^\text{15}\) \( s_F(q) = \mu \cdot s_e(q) + (1 - \mu) \cdot s_m(q) \). This specification of costs implicitly assumes the number of trained majority public employees is sufficient to provide services to the all majority residents not in the elite province: \( m \geq 1 - \mu \). An extension to allow for the use of untrained majority employees in federalism would be straightforward, but we do not pursue it here. Important later is the additional plausible assumption that \( s_F(q) > s_e(q) \) for all \( q \), and from the linear nature of our technology, that \( s_F(q) > s_m(q) \) as well. Sufficient for this to hold is that \( [1 - \mu]/\mu > [a_m/a_e] \). Together then, \( 1 \geq [(1 - m)/\mu] > [a_m/a_e] \) constrains the relationship of majority labor supply to relative labor productivity. We assume the constraints are met.
welfare depends on the central government’s choice of \( \tau \), the mandated or centrally chosen level of \( q \), and the elite’s choice of \( \phi \).

Under **democratic federalism**, \( \tau_f < \tau_U \), and:

\[
y(F; \phi^f) = Y - \tau_f + \phi^f[g_e - s_e(q)]\big/(M_e/N(\tau_f)),
\]

(4L)

for low capture, and:

\[
y(F; \phi^h) = [Y - \rho] - \tau_f + \phi^h[g_e - s_e(q)]\big/(M_e/N(\tau_f)),
\]

(4H)

for high capture, when majority residents of the elite province imposes protest costs \( \rho \). Under **administrative federalism**, \( \tau = \tau_U \), and:

\[
y(A; \phi^l) = Y - \tau_U + \phi^l[g_U - s_e(q)]\big/(M_e/N(\tau_U)), \text{ and,}
\]

(5L)

\[
y(A; \phi^h) = [Y - \rho] - \tau_U + \phi^h[g_U - s_e(q)]\big/(M_e/N(\tau_U)),
\]

(5H)

for the elite resident’s incomes under low and high capture, respectively. Finally, under **unitary democracy** without provinces (and thus no capture) the elite resident’s welfare is simply his or her after-tax income,

\[
y(U) = Y - \tau_U.
\]

(6)

Clearly, for any rate of capture, elite residents prefer democratic federalism to administrative federalism and administrative federalism to unitary governance: \( y(F; \phi^f) > y(A; \phi^l) > y(U) \) and \( y(F; \phi^h) > y(A; \phi^h) > y(U) \). Whether the elite prefers low or high capture will depend upon the net gain per elite resident of moving from low to high capture compared to the “protest penalty” \( \rho \) imposed when the high capture strategy is adopted.

Table 1 summarizes the pay-offs in any single year of the fiscal policy game for typical majority and elite residents, given the strategies of the majority-run central government and the elite-run provincial governments. Fiscal strategies in any given year are assumed to be chosen simultaneously. The central government can use provinces and set \( \tau = \tau_f < \tau_U \) (Strategy, \( F \)), use provinces but set \( \tau = \tau_U \) (Strategy, \( A \)), or ignore provinces, create a unitary democracy, and set \( \tau = \tau_U \) (Strategy, \( U \)). The elite run provinces can
adopt either a low (Strategy, $\varphi^L$) or high (Strategy, $\varphi^H$) rate of capture.\textsuperscript{16} Cells (1) and (4) define pay-offs under democratic federalism, cells (2) and (5) under administrative federalism, and cells (3) and (6) under unitary governance.

[INSERT TABLE 1]

C. Feasible Democratic Federalism: For democratic federalism to be a feasible candidate constitution, it must provide Pareto superior allocations to those available under either administrative federalism or unitary democracy. This is indeed the case in our specification of the fiscal policy game, since democratic federalism (Cell (1)) ensures greater potential private goods consumption for both majority and elite residents than is possible under either administrative federalism (Cells (2) or (5)) or under unitary democracy (Cells (3) and (6)). Because $\tau_E < \tau_U$, democratic federalism has more elite residents contributing to national income and lower per (elite) person fixed costs both in providing valued public services to majority residents and in meeting fixed special interest payments. In addition, there are no potentially wasted protest costs (as with administrative federalism) nor is there a need to use untrained providers to supply (as with unitary democracy). (See the Appendix, Section 1.)

Whether the economic benefits of democratic federalism will be realized depends on the constitutional ground rules that dictate the play of the annual policy game. Were the fiscal policy game to be played only once, or with a sufficiently high discount rate so that future payoffs became unimportant, democratic federalism would be a dominated strategy for the majority, always less preferred than either administrative federalism or unitary governance.\textsuperscript{17} Fortunately, the policy game is played repeatedly. Under appropriately designed constitutional rules, unending repeat play will allow the elite to punish the majority

\textsuperscript{16} In principle the elite has a third strategy: set $q = 0$ and capture all of $g$. We rule out this strategy by our assumption of full information. As soon as the elite ceases to provide $q$, the majority knows and the majority ceases to pay $g$. Thus stealing all of $g$ by setting $q = 0$ is not possible.

\textsuperscript{17} For common values of $Z, M, q,$ and $\mu$, $\tau_U > \tau_F$ implies $g_U > g_F$, and therefore $\omega(A; \varphi^L) > \omega(F; \varphi^L)$ and $\omega(A; \varphi^H) > \omega(F; \varphi^H)$. If $\varphi^L$ or $\varphi^H$ are large enough, then $\omega(U) > \omega(A; \varphi^L) > \omega(F; \varphi^L)$ and $\omega(U) > \omega(A; \varphi^H) > \omega(F; \varphi^H)$ as well.
whenever the majority defects from democratic federalism. For the policy game in Table 1, the elite’s punishment strategy will be to increase capture from $\varphi^i$ to $\varphi^H$.

To effectively deter majority defection from democratic federalism, however, the elite’s threat to adopt the high capture strategy must be credible. For high capture to be a credible threat, (i) provinces must remain as viable fiscal jurisdictions, even when the majority defects from democratic federalism; (ii) the elite must be a political majority in at least one province; and (iii) the elite majority must prefer $\varphi^H$ to $\varphi^L$ when the majority defects. Formally:

**DEFINITION: CREDIBLE PUNISHMENT.** The high capture strategy will be a credible punishment strategy when:

(i) The poor majority prefers provinces and administrative federalism as their defection alternative, i.e., $\omega(A; \varphi^H) > \omega(U)$;

(ii) The elite is a political majority in at least one province, i.e., $N(\tau,U) \geq M_e$; and,

(iii) The elite prefers the high capture strategy to low capture when the poor majority defects to administrative federalism, i.e., $y(A; \varphi^H) > y(A; \varphi^L)$.

The three requirements for a credible elite punishment will be met when the fiscal constitution creates an elite province of sufficient majority population for the high capture strategy to be an effective threat and the level of assigned services are such that provinces are valued by the majority as service providers. We call the population constraint the constitutional *Border Constraint* and the choice of assigned services and, when enforceable, the chosen input levels the constitutional *Assignment Constraint*. When both the border and assignment constraints are met, the elite has a credible punishment for excessive majority taxation and democratic federalism is a feasible constitution. Proposition 1 then shows that for at least one set of Nash strategies – the grim trigger strategies – democratic federalism is not only feasible but also sustainable.

We specify the border and assignment constraints for credible elite punishments for two alternative political economy regimes. The first, called the *q-Regime*, assumes the constitution’s border and assignment constraints are specified for an constitutionally chosen and *exogenously enforced* level of q that holds under
each of the three forms of democracy. In the q-Regime, the constitution (i) sets provincial borders (μ), (ii) assigns the services to be provided in the province (λ), and (iii) mandates and enforces an input level (q) to provide those services. Successful enforcement of the input bundle is assumed for the q-Regime, perhaps emerging from a consensus belief in the ethical validity of the constitutionally assigned service standard.\(^{18}\)

The second regime, called the \(q^*-\text{Regime}\), is arguably the more realistic environment and here we allow the level of the input bundle q for assigned services to be endogenously chosen by the majority-run central government depending upon the fiscal incentives implicit in each form of democracy. In the \(q^*-\text{Regime}\), the constitution (i) selects the provincial borders (μ) and (ii) assigns services (λ), but the constitutional mandate setting q is dropped. The majority-run central government now decides the level of service inputs as \(q^*\).\(^{19}\) We show below that a credible elite punishment, and thus a federal democracy, may be more difficult to sustain in the endogenous \(q^*\) environment.

1. \textit{Border and Assignment Constraints in the q-Regime}: For elite punishment to be credible, the elite-run province must willingly adopt the high capture strategy when the majority defects from F to A. This will be case when the net fiscal gain to an elite resident exceeds the costs the resident bears from provincial unrest when they adopt high capture. Given borders (μ) and assignment (λ), and an exogenously enforced common value of q, high capture is preferred when:

\[
y(A; \varphi^H) > y(A; \varphi^L) \iff (\varphi^H - \varphi^L)[g_u - s_e(q)]\left[\frac{M_e}{N(J_u)}\right] > \rho,
\]

or when:

\[
(M_e/M) = \mu > \{\rho[N(\tau_u)/M]\}/\{(\varphi^H - \varphi^L)[g_u - s_e(q)]\} = \mu^{\text{min}}(q), \tag{7}
\]

---

\(^{18}\) See for example, Ackerman (1991; Chapter 1) who argues one role for a country’s Supreme Court is to remind citizens of this consensus belief and to clarify its application to policy conflicts.

\(^{19}\) In Section IV we present evidence that South Africa may have evolved from a q-Regime (1996-2000) in which President Mandela “enforced” the level of q negotiated in the original, bilateral, NP-ANC constitutional bargain to a \(q^*\)-Regime (2001-2006) where today, under President Mbeki, the ANC majority sets the input level to provide constitutionally assigned services.
where $\mu$ is the fraction of the nation’s majority residents who reside in the elite province.\footnote{We require strict inequality because the elite are assumed to prefer to cooperate rather than defect, all else equal.} If $v^H$ is to be credible, there must be enough majority residents in the elite province so that the additional resources captured using $v^H$ exceed the costs imposed by local political protests when capture is high. The more majority residents “assigned” to the elite province the greater the revenues per elite resident from high capture. In effect, the majority residents assigned to the elite province act as “hostages” in the annual fiscal policy game.

But $\mu$ cannot be too large. For the high capture threat to be credible, the elite must still be a political majority in their province. Thus $N(\tau_U) > M_e$, or:

$$N(\tau_U)/M = \mu^\text{max} \geq \mu = (M_e/M).$$  \hspace{1cm} (8)

For high capture to be a credible punishment strategy with an exogenous common $q$, the constitutionally mandated population size of the elite province must satisfy the $q$-Border Constraint specified as:

$$\mu^\text{max} \geq \mu > \mu^\text{min}(q).$$  \hspace{1cm} (9q)

The border constraint defines the population size of the elite province as $N(\tau) + M_e$.\footnote{Federal constitutions specify $\mu$ by a map of provincial borders at the time of the adoption of the constitution; for South Africa, see Muthien and Khosa (1998).}

The elite can only punish through high capture if they control a political jurisdiction with budgetary responsibilities for redistributive services and basic grants. If the central government defects from democratic federalism, it must defect to the alternative where provinces still have fiscal responsibilities — that is, administrative federalism must be preferred to unitary governance. This will be the case when $\omega(A; \phi^i) > \omega(U)$, or equivalently, when:

$$s_i(q) - [s_i(q) - \phi^i \cdot \mu \cdot s_i(q)] > \phi^i \cdot \mu \cdot g_i,$$  \hspace{1cm} (10)
where the LHS of the inequality represents the additional expenditures needed to provide $q$ under unitary governance and the RHS represents the savings in less capture by adopting unitary governance. When elite capture ($\varphi^H$) is low and/or when the relative cost advantage of administrative federalism is high ($a_i^1, a_u^1, m^1$), then administrative federalism will be preferred to unitary governance. Given the rate of capture and the relative costs of providing $q$, this inequality holds when the constitutionally mandated level of $q$ is large enough, defined by:  

$$q > q^{\min}(\mu) = (\varphi^U \cdot g_U) / (S \cdot \hat{a}(\mu)).$$  

Raising $q$ reduces the basic grant available for capture and therefore makes administrative federalism relatively more attractive. Setting $q > q^{\min}$ ensures administrative federalism is the majority’s preferred defection strategy.

But the exogenously enforced, uniform $q$ cannot be too large. As $q$ increases, the net returns to capture may fall below the amount needed for the elite to find the high capture punishment strategy a preferred response to a majority defection from $F$ to $A$. The maximum common $q$ that protects $\varphi^H$ as a credible punishment strategy will be that $q (= q^{\max})$ for which the two border constraints just hold – that is, where $\mu^{\max} = \mu^{\min}(q^{\max})$. Together these lower and upper bounds define the constitutional $q$-Assignment Constraint for redistributive services as:

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

When the political regime allows the exogenous enforcement of a common value of $q$ for all forms of democracy, then if both the $q$-Border and $q$-Assignment Constraints are met the elite will be in a position, under democratic federalism, to credibly punish the majority for any deviations from $\tau_F$ to $\tau_U$. The result is

---

22 We assume $s_i(q) > s_F(q)$; see fn. 15. A strict inequality is required since we assume, all else equal, the majority prefers unitary democracy. From the definitions of $s_i(q)$, $s_F(q)$, and $s_i(q)$: $\hat{a}(\mu) = \mu \cdot [(1/a_m) - (1/a_i)] + (1 - m) \cdot [(1/a_u) - (1/a_m)] + (\mu \cdot \varphi^U/a_v) > 0$.

23 Using eqs. (7) and (8) above, $q^{\max} = [g_U(\varphi^H - \varphi^U) - p] / (\varphi^H - \varphi^U)(S/a_v)$. All elements defining $q^{\max}$ are exogenous in the second-stage policy game. For all $q \leq q^{\max}$, $\mu^{\max} \geq \mu^{\max}$ as required for the border constraint. See Figure 1.
summarized in Lemma 1.

**LEMMA 1: CREDIBLE PUNISHMENTS IN THE EXOGENOUS 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 (centralizing) redistributive tax rate. (Proof: See Appendix.)

Together the q-Border and q-Assignment Constraints define a set of possible constitutional values of μ and q where the elite can credibly threaten to punish the majority when they deviate from a federal tax rate of \( \tau_\ell < \tau_\mu \); see Figure 1. (Figure 1 is calibrated to the South African political economy; see Section IV.) The q-Border Constraint requires values of μ that lie above the \( \mu_{\text{min}}(q) \) curve and below the \( \mu_{\text{max}} \) line. The q-Assignment Constraint requires values of the common q that lie to the right of the \( q_{\text{min}}(\mu) \) curve and at or to the left of \( q_{\text{max}} \). The shaded area shows all values of μ and an exogenously enforced q where the two constraints are jointly satisfied. Finally, in the q-Regime, the Border and Assignment constraints hold for any basket of assigned services – that is, for all values of \( \lambda \), whether for a low \( \lambda \) (street lighting) or for a high \( \lambda \) (education or health care). This will not be the case in the q*-Regime, however. When q cannot be enforced, the constitution’s choice of assigned services (\( \lambda \)) becomes crucial.

There is no guarantee that both constraints can be met. First, if the lower bound for the border constraint – \( \mu_{\text{min}}(q) \) – rises, then the lowest feasible value for μ – denoted \( \mu_{\text{min}}(q_{\text{min}}) \) in Figure 1 – may rise above \( \mu_{\text{max}} \), and the shaded area in Figure 1 may disappear. A credible elite punishment is no longer possible, and democratic federalism is not sustainable. The number of majority “hostages” needed to make high capture credible is greater than the number of elite residents. If so, elite political control of the province is lost. More hostages will be needed as the penalty for high capture increases (\( R \)) or as high capture returns relatively less revenues either because \( \varphi^H \) falls, or \( \varphi^L \) increases, or because \( g_U \) is relatively low. Maximal grants, \( g_U \), will be low when either the initial pre-democracy ratio of elite to majority residents is small or
when the elite population has a high propensity to exit as tax rates rise.\textsuperscript{24} As Albert Hirschman (1970) first noted: Democracy often requires loyalty.

Second, though the constitution requires a level of \( q \) services to be provided by federal or unitary governments, there is no assurance that the majority controlled central government will comply with the constitutional mandate. It is possible that the majority’s preferred level of \( q \) – denoted as \( q^* \) – might fall outside the feasible set in Figure 1 that allows credible elite punishments. When \( q \) is endogenously chosen by the majority, the border and assignment constraints supporting credible elite punishments must be re-specified. Feasible federal constitutions are still possible, but now new \( q^* \)-Border and \( q^* \)-Assignment Constraints are needed.\textsuperscript{25}

2. \textit{Border and Assignment Constraints in the \( q^* \)-Regime:} As for the \( q \)-Regime, the majority must continue to prefer \( A \) to \( U \) when it defects and if it does defect to \( A \), the elite-run province must willingly adopt the high capture strategy. To ensure these conditions will be met, we define new border and assignment constraints, conditional on the majority’s chosen levels of \( q \) in each democratic regime. We specify the majority’s preferred level of \( q \) by equating their marginal benefits of providing \( q \) to the marginal cost of that provision in each of the three democratic regimes, conditional upon the elite’s choice of capture:

\[
(F, \varphi^I): \lambda \cdot v'(q) = p_e(\mu) = s_e'(q) - \varphi^I \cdot \mu \cdot s_e'(q) \rightarrow q^* = q^*(\mu, \lambda; \varphi^I) = q^*_e(\mu, \lambda)
\]

\[
(F, \varphi^II): \lambda \cdot v'(q) = p_u(\mu) = s_u'(q) - \varphi^{II} \cdot \mu \cdot s_e'(q) \rightarrow q^* = q^*(\mu, \lambda; \varphi^{II}) = q^*_u(\mu, \lambda)
\]

\[
(A, \varphi^I): \lambda \cdot v'(q) = p_e(\mu) = s_e'(q) - \varphi^I \cdot \mu \cdot s_e'(q) \rightarrow q^* = q^*(\mu, \lambda; \varphi^I) = q^*_e(\mu, \lambda)
\]

\[
(A, \varphi^{II}): \lambda \cdot v'(q) = p_u(\mu) = s_u'(q) - \varphi^{II} \cdot \mu \cdot s_e'(q) \rightarrow q^* = q^*(\mu, \lambda; \varphi^{II}) = q^*_u(\mu, \lambda)
\]

\textsuperscript{24} For the linear exit relationship \( N(\tau) = N_0 - \beta \cdot \tau, g_U = [\tau_U - (.5N_0 - Z)]/M \), where \( \tau_U = .5N_0/\beta \). Low values of the initial elite to majority population – \( N_0/M \) – and high values of the exit parameter – \( \beta \) – both lower \( g_U \).

\textsuperscript{25} We ignore the possibility that provincial boundaries (\( \mu \)) may be endogenous. For any value of \( q \), the specification of \( \omega(F) \) implies the majority will prefer either a value of \( \mu = \mu^{\text{max}} \) or a value of \( \mu^{\text{min}}(q) \) to the constitutional choice of \( \mu \), depending upon whether the constant cost savings from moving a few more poor residents into the elite province exceeds (is less than or equal to) the average income lost from such a relocation: \( \mu^{\text{max}} \) is preferred if \( [s_u(q^*) - s_e(q^*)] > \varphi^I [g_U - s_e(q^*)] \) and \( \mu^{\text{min}}(q^*) \) preferred otherwise. In the case of South Africa \( \mu^{\text{max}} \) is preferred. The constitutionally chosen value of \( \mu = .184 \) is very close.
\[(U): \lambda \cdot v'(q) = p_u = s_u'(q) \rightarrow q^* = q^*_u(\lambda)\]

The unit prices of q in the two federal regimes, \(p_L(\mu)\) and \(p_H(\mu)\), equal the marginal production costs \(s_v'(q)\) minus the income benefits to majority residents that higher values of q provide because of reduced captured revenues. These prices fall as \(\mu\), the share of majority residents benefitting from low cost elite providers, increases.\(^{26}\) Thus, \(\partial q^*_L/\partial \mu > 0\) and \(\partial q^*_H/\partial \mu > 0\). Majority benefits from q rise as \(\lambda\) rises. Thus \(\partial q^*_L/\partial \lambda > 0\) and \(\partial q^*_H/\partial \lambda > 0\) as well. Finally, because the prices for q under low or high capture are similar for the \(F\) and \(A\) constitutions, the majority funds the same level of q in these regimes, denoted as \(q^*_L(\mu, \lambda)\) and \(q^*_H(\mu, \lambda)\), respectively. Under unitary governance, there is no capture. As \(s_v'(q) > s_v'(q)\) and \(\varphi^H > \varphi^L\), therefore \(p_U > p_L(\mu) > p_H(\mu)\) for the relevant case of \(\mu > 0\). Since we assume \(v''(q) < 0\), therefore too, \(q^*_U(\lambda) < q^*_L(\mu, \lambda) < q^*_H(\mu, \lambda)\). The border and assignment constraints can now be specified conditional on \(\mu\) and \(\lambda\) and their associated q*’s.

As before, the border constraint follows from our requirement that the elite province must adopt the high capture strategy when the majority defects from \(F\) to \(A\). Now, however, allowance must be made for the fact that q is chosen by the central government majority, conditional on the decision by the elite to adopt low or high capture. High capture will be preferred when:

\[y(A; q^*_H, \varphi^H) > y(A; q^*_L, \varphi^L) \rightarrow \{\varphi^H[g_U - s_\lambda(q^*_H(\mu, \lambda))] - \varphi^L[g_U - s_\lambda(q^*_L(\mu, \lambda))]\} \cdot \frac{M_U}{N(\tau_U)} > \rho,\]

or when:

\[(M/M) = \mu > \frac{\rho [N(\tau_U)/M]}{\{\varphi^H[g_U - s_\lambda(q^*_H(\mu, \lambda))] - \varphi^L[g_U - s_\lambda(q^*_L(\mu, \lambda))]\} = \mu^{\text{min}}(q^*_L), \quad (13)\]

where \(\mu\) is again the fraction of the country’s majority residents who reside in the elite province and now must exceed \(\mu^{\text{min}}(q^*_L)\).\(^{27}\) The upper bound for the border constraint remains the same as in the q-Regime. For high capture to be a credible punishment strategy with endogenous q*, the constitutionally mandated

\(^{26}\) This follows directly from the definitions of \(p_L(\mu)\) and \(p_H(\mu)\) and the facts that \(a_e > a_m\) and \(1 > \varphi^L\).

\(^{27}\) We are assuming that high capture is profitable so that: \(\varphi^H[g_U - s_\lambda(q^*_H(\mu, \lambda))] - \varphi^L[g_U - s_\lambda(q^*_L(\mu, \lambda))] > 0\). This constraint places a mild upper (absolute value) bound on the majority’s price elasticity of demand for assigned goods, generally 2 or higher.
population size of the elite province must now satisfy the \( q^*-\text{Border Constraint} \) specified as:

\[
\mu_{\text{max}} \geq \mu > \mu_{\text{min}}(q^*). \tag{9q^*}
\]

To ensure the elite has a credible punishment when the majority adopts high taxation, the majority must prefer to defect to administrative federalism rather than to a unitary democracy. In the \( q^* \)-Regime, this will be true for any \( \mu > 0 \) if and only if:

\[
\omega(A; \mu, q^*_l(\mu, \lambda), \varphi^l) > \omega(U; q^*_u(\lambda)).
\]

Majority preferences are single-peaked in \( q \) as \( v'(q) > 0 \) and \( v''(q) < 0 \). Given single-peaked preferences in \( q \), the majority will prefer administrative federalism to unitary governance when there exists a best \( q^*^*(\mu, \lambda) \) bounded between the values \( q^*_{\text{min}}(\mu, \lambda) \) and \( q^*_{\text{max}}(\mu, \lambda) \), where \( q^*_{\text{min}}(\mu, \lambda) \) and \( q^*_{\text{max}}(\mu, \lambda) \) are the two values of \( q^* \) – one less than \( q^*_l(\mu, \lambda) \) and the other greater than \( q^*_l(\mu, \lambda) \) – that are just sufficient to equate majority welfare under administrative federalism to the welfare threshold set by unitary governance.\(^{28}\) As with exogenous \( q \), however, there may be a tighter upper constraint on the upper value of \( q^*_{l}(\mu, \lambda) \), defined by that \( q_{l}^*_{\text{max}}(\mu, \lambda) \) where the border constraints just hold: \( \mu_{\text{max}} = \mu_{\text{min}}(q^*_{l}). \) Assuming \( q_{l}^*_{\text{max}}(\mu, \lambda) \) defines the relevant upper bound for \( q^*_{l} \), then the \( q^*\)-\text{Assignment Constraint} is specified as:\(^{29}\)

\[
q_{l}^*_{\text{max}}(\mu, \lambda) > q^*_{l}(\mu, \lambda) > q_{l}^*_{\text{min}}(\mu, \lambda). \tag{12q^*}
\]

Given provincial borders (\( \mu \)), meeting the \( q^*\)-\text{Assignment Constraint} requires the right selection of services for provincial assignment: Should we assign high \( \lambda \) (education) or low \( \lambda \) (street lighting) services to provinces? The importance of selecting the right bundle for provincially assigned public services is seen most easily when we note that \( \omega(A; \mu, q^*_l(\mu, \lambda), \varphi^l) > \omega(U; q^*_u(\lambda)) \) as:

\[
[u(q^*_l(\mu, \lambda)) - p^l(\mu, q^*_l(\mu, \lambda))] - [u(q^*_u) - p^u] > \varphi^l - \mu \cdot q^*_u. \tag{14}
\]

The LHS of the inequality measures the difference between the consumer surplus earned by a typical

\(^{28}\) If \( \omega(A; \mu, q^*_l(\mu, \lambda), \varphi^l) \leq \omega(U; q^*_u(\lambda)) \) for a given \( \mu \) and \( \lambda \), then there is no \( q^*_{l} \) that can satisfy the Assignment constraint; the bounds on \( q^*_l(\mu, \lambda) \) are not defined.

\(^{29}\) If \( q^*_{l\text{max}}(\mu, \lambda) \) is the tighter constraint, then the \( q^*\)-Assignment Constraint becomes \( q^*_{l\text{max}}(\mu, \lambda) > q^*_l(\mu, \lambda) > q^*_l(\mu, \lambda) > q^*_u(\mu, \lambda). \)
majority resident under administrative federalism when the price of assigned services is $p_8(\mu)$ and the surplus earned by the majority resident under unitary democracy when the price of a comparable service bundle is $p_U$. Since $p_U > p_8(\mu)$, consumer surplus is larger under administrative federalism, shown as the shaded area in Figure 2.\(^{30}\) (Calibrated to the South African political economy; see Section IV.) Because of elite capture, however, administrative federalism also imposes an income loss $q^* \cdot \mu \cdot g_U$ on the average majority resident.

To meet the $q^*$-Assignment Constraint, assigned services – indexed by $\lambda$ – must generate a gain in consumer surplus from adopting administrative federalism rather than unitary governance that exceeds the income loss of elite capture. From Figure 2, the gain in consumer surplus grows when assigned services are highly valued – so $\lambda \cdot v'(q)$ shifts outward and $q^*_8(\mu)$ and $q^*_U$ are large – and when elite provinces have a strong production advantage over unitary governance so $[p_U - p_8(\mu)]$ is large as well. When these conditions are met, provinces are important to the majority and worth the “cost” of capture. South Africa’s decision to assign K-12 education, primary health care, and the administration of income transfers to provinces seems consistent with the $q^*$-Assignment Constraint. We can now state Lemma 2.

**Lemma 2: Credible Punishments in the Endogenous 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 (centralizing) redistributive tax rate. (Proof: See Appendix.)

The shaded area in Figure 3 (again, calibrated for the South African political economy) shows the set of provincial borders ($\mu$) and chosen levels of assigned services ($q^*_8$) that jointly satisfy the requirements of Lemma 2 for credible elite punishment. The rising solid lines within the shaded area represent the portions of the demand curves $q^*_8(\mu, \lambda)$ that satisfy the $q^*$-Border Constraint, with higher values of $\mu$ raising

---

\(^{30}\) Public finance economists will recognize the shaded area in Figure 2 as the total burden of a tax, approximated by the sum $[p_U - p_8(\mu)] \cdot q^*_U + .5 \cdot [p_U - p_8(\mu)] \cdot [q^*_8(\mu) \cdot q^*_U] + .5 \cdot [p_U - p_8(\mu)] \cdot [q^*_8(\mu) + q^*_U]$. When $\cdot 5 \cdot [p_U - p_8(\mu)] \cdot [q^*_8(\mu) + q^*_U] > \phi^* \cdot \mu \cdot g_U$, the $q^*$-Assignment Constraint is met, from which the intuition above follows directly.
along a demand curve and higher values of \( \lambda \) shifting the demand curve to the right. For example, at \( \mu^* = .18 \), representing 18 percent of the majority population assigned to the elite province, the majority would prefer an input bundle \( q^*_L = .53 \) if the assigned service utility weight \( \lambda \) were equal 2840. Higher utility weights slide the demand curve to the right, and for \( \mu^* = .18 \), might lead demands for \( q^*_L \) outside the feasible set, such as \( q^*_L = .65 \) when \( \lambda = 3487 \). Each point within the shaded area represents one combination of provincial borders (\( \mu \)) and assigned services (\( \lambda \)) that satisfies the \( q^*_\text{-Border} \) and \( q^*_\text{-Assignment} \) Constraints.

As in the q-Regime, however, there is no guarantee that both constraints can be jointly satisfied. Here too, the set of feasible federal constitutions shrinks as \( \mu^{\text{min}}(q^*_L) \) rises towards \( \mu^{\text{max}} \) and for the same reasons as obtained in the q-Regime: a rising penalty for high capture (\( \rho \)), declining relative return (\( \varphi^H - \varphi^E \)) to high capture, or a low value of \( g_u \) because of low initial elite populations or high elite exit.

3. Comparing Regimes: Figure 3 illustrates that between the q- and \( q^* \)-Regimes, losing the ability to enforce a particular level of redistributive services may significantly constrain the set of feasible federal constitutions. When the majority cannot credibly promise to provide a constitutionally mandated level of \( q \), tighter bounds on the other two dimensions of the federal constitution – borders and assignments – will be needed to ensure the constitution can be enforced. First, in the \( q^* \)-Regime, when the majority defects from \( F \) to \( A \), the elite’s decision to punish and adopt \( \varphi^H \) must be consistent with the majority’s decision to adopt \( q^*_H(\mu, \lambda) \). For any initial value of \( q^*_L = q_i \), it must be true that high capture generates less revenues for the elite in the \( q^* \)-Regime, since \( \varphi^H - s_e(q^*_H(\mu, \lambda)) < \varphi^E - s_e(q) \) as \( q^*_H(\mu, \lambda) > q^*_L = q \). To cover the protest costs (\( \rho \)) borne by each elite resident when \( q^H \) is chosen, more majority residents must be assigned as “hostages” to the elite province. Thus for each value of \( q^*_L = q \) in Figure 3, \( \mu \) must rise in the \( q^* \)-Regime. This is shown as the \( \mu^{\text{min}}(q^*_L) \) curve everywhere above the \( \mu^{\text{min}}(q) \).

Second, to ensure the majority does defect only as far as \( A \), and not to \( U \), the welfare gain from staying in \( A \) must just exceed the benefits saving what had been capture – \( \varphi^E - \mu \cdot g_u \) – by moving to \( U \). In the q-Regime, the welfare gain in \( A \) is the savings in costs from using provinces to provide the exogenous level
of q; see the LHS of eq. (10). These savings are represented by a rectangle, such as the heavily shaded area in Figure 2. If \( q^\star \cdot g_U \) just equals this heavily shaded rectangle, then \( q^{\min}(\mu) \) would be represented by \( q^\star(\lambda) \) in Figure 2. In the \( q^\star \)-Regime, however, the majority can choose q. By selecting \( A \) over \( U \) the majority gains not just the cost savings (the rectangle) of \( A \), but they also gain consumer surplus by being able to buy more q (a triangle). Because \( p_L(\mu) < p_U \), they do. Thus \( q_L^\star > q^{\min}(\mu) \). Where the area of consumer surplus just equals \( q^\star \cdot g_U \) defines \( q_L^{\min}(\mu, \lambda) \); thus \( q_L^{\min}(\mu, \lambda) > q^{\min}(\mu) \). For any choice of \( \mu \), the minimal q, or equivalently the value of assigned services \( \lambda \), must be higher in the \( q^\star \)-Regime. This is shown in Figure 3 as the \( q_L^{\min}(\mu, \lambda) \) curve everywhere to the right of the \( q^{\min}(\mu) \).

Together, the \( q^\star \)-Regime’s induced upward shift of the Border Constraint and rightward shift of the Assignment Constraint shrink the set of borders (\( \mu \)) and assignments (\( \lambda \), or equivalently, \( q^\star_L \)) that permit feasible federalism. We summarize this result as:

**Lemma 3: Feasible Federalism with Endogenous Service Levels.** When the constitution loses the ability to enforce the level of assigned services (q), then for credible elite punishments and feasible federalism, either elite provinces specified by \( \mu \) must be larger or provincially assigned services specified by \( \lambda \) must be more highly valued, or both. (Proof: See Appendix.)

Lemma 3 provides a formal foundation for a useful intuition: Democratic federalism is likely to survive as a viable institution only in societies first, with enough loyal elite residents, and second, with enough low-cost elite providers of public services valued highly by the majority. At the time of its transition, South Africa appears to have met this test.

**D. Sustainable Democratic Federalism:** Ensuring the feasibility of democratic federalism does not guarantee the institution will survive as a long-run equilibrium of the annual policy game. Though both the majority and the elite credible punishments if the other deviates from the federal agreement, democratic federalism only survives if those punishments are sufficient to discourage defections from the strategy pair \((F, q^\star)\) in Table 1. Formally:

**Definition: Sustainable Democratic Federalism.** Democratic federalism will be a sustainable constitutional regime for the infinitely repeated policy game if the strategy pair \((F, q^\star)\) is a
subgame-perfect Nash equilibrium.

For the majority, the strategy $F$ requires a choice of the central government’s tax rate be a value of $\tau_F < \tau_U$, and its grants to the provinces be $g_F < g_U$. For subgame perfection, this strategy choice $(\tau_F, g_F)$ must allocate the surplus available under democratic federalism to the poor majority and the current elite in such a way that both prefer democratic federalism to either administrative federalism or unitary democracy in all budget years. To ensure the elite are better off under democratic federalism taxes cannot be too high; the majority’s fiscal choices must satisfy $\tau_U > \tau_F^{\max} > \tau_F$ or equivalently, $g_U > g_F^{\max} \geq g_F$, where $\tau_F^{\max}$ is just sufficient to ensure the elite will continue to play $\varphi^1$ under democratic federalism. To ensure that the poor majority are better off under democratic federalism, total redistributive spending cannot be too low; the majority’s fiscal choices must satisfy $g_F > g_{\min}$ or equivalently, $\tau_F > \tau_{\min}$, where $g_{\min}$ is the intergovernmental transfer just sufficient to ensure that the majority prefers democratic federalism. For subgame perfection, therefore, the majority chosen tax rate and total intergovernmental transfers under federalism must satisfy $\tau_U > \tau_F^{\max} > \tau_F > \tau_F^{\min}$, or equivalently, $g_U > g_F^{\max} > g_F > g_{\min}$. In the analysis that follows, we focus on the bounds $g_U > g_F^{\max} > g_F > g_{\min}$, specified separately for the $q$-Regime and the $q^*$-Regime.

Proposition 1 specifies $g_F^{\max}$ and $g_F^{\min}$ for the $q$- and $q^*$-Regimes for one pair of subgame perfect strategies and then proves that there exists a majority preferred fiscal choice $(\tau_F, g_F)$ and an elite capture decision $(\varphi^1)$ satisfying those bounds. The strategy choices considered here are grim trigger strategies defined as:

**DEFINITION: GRIM TRIGGER STRATEGIES.** The poor majority and the elite are said to play grim trigger strategies at each stage of the policy game if:

(i) along the equilibrium path, the majority-run central government selects $(\tau_F, g_F)$ and the elite-run province selects $\varphi^1$; and,

(ii) off the equilibrium path, were the majority central government to have selected $(\tau_U, g_U)$ the elite would play $\varphi^0$ in all subsequent budget periods, and where, were the elite province to have selected $\varphi^0$, the majority would play either administrative federalism ($A$) selecting $(\tau_U, g_U)$, or unitary democracy ($U$) selecting $\tau_U$ in all subsequent budget periods.
These grim trigger strategies are the most extreme form of punishment one player can impose on the other for defection in this game; if democratic federalism cannot be sustained under these strategies, it cannot be sustainable under any other feasible strategies.\textsuperscript{31} Were the majority to defect and select \( \tau_u \), the elite would punish by selecting the high capture strategy forever. Were the elite to defect and adopt a high capture strategy, then the majority-run central government would punish by selecting \( \tau_v \) and using administrative federalism forever, or by selecting \( \tau_u \), and using unitary democracy forever. For any provincial size (\( \mu \)) and assignment (\( q \)) satisfying the border and assignment constraints, the majority prefers to punish an elite defection (\( \varphi^U \)) using unitary democracy when \( \omega(U) > \omega(A; \varphi^U) \) in the \( q \)-Regime and \( \omega(U; q^*_U(\lambda)) > \omega(A; \mu, q^*_U(\mu, \lambda), \varphi^U) \) in the \( q^* \)-Regime, and to punish using administrative federalism when the inequalities are reversed. We can now state our main result.

**PROPOSITION 1: SUSTAINABLE DEMOCRATIC FEDERALISM:** For elite provincial size and redistributive assignment satisfying the Border and Assignment Constraints, there exists a grim trigger strategy equilibrium where democratic federalism is sustainable. For majority and elite residents with discount factor \( 0 < \delta \leq 1 \), that equilibrium is characterized by:

1. For majority punishment strategy \( U \), the majority chooses \( g_F \) within the bounds:
   \[ g_U > g^\text{max}(U) > g^\text{max}(U) > g^\text{min}(U) > g^\text{min}(U) ; \]

2. For majority punishment strategy \( A \), the majority chooses \( g_F \) within the bounds:
   \[ g_U > g^\text{max}(A) > g^\text{max}(A) > g^\text{min}(A) > g^\text{min}(A) ; \] and,

3. The elite province adopts \( \varphi^U \). (Proof: See Appendix.)

A complete specification for each value of \( g^\text{max} \) and \( g^\text{min} \) for the \( q \)- and \( q^* \)-Regimes is given in the Appendix.

Whether \( U \) or \( A \) is the majority’s preferred punishment, sustainable grants are unambiguously higher

\textsuperscript{31} In this infinitely repeated Prisoners’ dilemma game, the players’ payoffs in the non-cooperative Nash equilibrium of the annual stage game are also their reservation payoffs. In this special case, the grim trigger strategy is equivalent to Abreu’s (1988) strongest credible punishment strategy; see Gibbons (1992, p. 99). Softer punishment strategies might be considered, as the grim trigger strategy in this game, while subgame perfect, is not weakly renegotiation proof; see Farrell and Maskin (1989). Agreeing to let “bygones be bygones,” the elite and majority can adopt a strategy pair that allows them to move from punishment forever back to cooperation in time. If so, defection becomes less costly making federalism harder to maintain; the range \([g^\text{max}, g^\text{min}] \) of feasible federal fiscal contracts gets smaller. Thus grim trigger strategies give federalism its best chance for adoption in this game with full information.
in the q*-Regime than the q-Regime. Intuitively, when the q- and q*-Regimes have identical allocations in
democratic federalism, given the majority the right to set q at preferred levels when they do defect makes defection more attractive for the majority and less so for the elite. This raises the minimal transfers needed
to hold the majority in democratic federalism and allows a higher tax rate on the elite before they will be
tempted to defect. Both effects favor the majority and hurt the elite by moving the range of constitutionally
acceptable intergovernmental transfers towards more redistribution. Not only does the q*-Regime make
democratic federalism more difficult to achieve in the annual policy game (Lemma 3), it also makes it
potentially more costly for the elite to accept at the constitutional stage (Proposition 1).

It is also important to understand the role of impatience for constitutional agreement, measured here
by the discount factor \( \delta \), which may differ for the two groups. Lower values of \( \delta \) represent greater
impatience. The full specification of Proposition 1 in the Appendix show that a more impatient elite implies
lower values for \( g^{\max}_*(U), g^{\max}(U), g^{\max}_*(A), \) and \( g^{\max}(A) \), while a more impatient majority will imply larger
values for \( g^{\min}_*(U), g^{\min}(U), g^{\min}_*(A) \) and \( g^{\min}(A) \); see the Appendix. Importantly, if both parties are very
impatient, we cannot rule out the possibility that \( g^{\min}_*(\bullet) > g^{\max}_*(\bullet) \) in all cases and that there will be no
equilibrium transfer that is capable of sustaining democratic federalism and also agreeable to both parties
at the constitutional stage. Successful democratic constitution writing requires at least some commitment
to the long-run by the residents of the new democracy.

Finally, while any value of \( g_e \) satisfying Proposition 1 will be consistent with a sustainable federal
democracy, a choice must be made. Constitutional negotiators might choose to divide the surplus earned in
democratic federalism equitably between the majority and the elite by setting \( g_e \) midway between \( g^{\min} \) and
\( g^{\max} \). But such a choice must be enforced exogenously. The central government’s majority clearly prefers
\( \tau_e = \tau^{\max} \) and \( g_e = g^{\max} \). It is therefore useful therefore to explore the likely determinants of \( g^{\max} \).

In the q-Regime either \( g^{\max}(U) \) or \( g^{\max}(A) \) will result or in the q*-Regime either \( g^{\max}_*(U) \) or \( g^{\max}_*(A) \)
is the outcome. Which value of \( g^{\max} \) sets the upper bound on transfers will depend upon whether unitary
democracy or administrative federalism is the majority’s preferred punishment strategy for elite deviations to high capture. Comparing \( \omega(U) \) to \( \omega(A; \varphi^U) \) in the q-Regime and \( \omega(U; q^*_U(\lambda)) \) to \( \omega(A; \mu, q^*_U(\mu, \lambda), \varphi^U) \) in the q*-Regime shows that unitary democracy will be the preferred punishment when high capture is costly (\( \varphi^U = 1 \)) or when the cost disadvantage of leaving provincial for unitary government provision of redistributive services is small. This cost disadvantage will be small when the majority has a relatively well-trained cadre of majority loyal public providers \( (a^U, a^m) \) sufficient to meet most all of the majority’s service needs \( (m^U) \). Corollary 1 provides a useful comparative static result:

**COROLLARY 1**: \( g^{\max}(U) > g^{\max}(A) \) and \( g^{\max^*}(U) > g^{\max^*}(A) \). For federal constitutions satisfying required Border and Assignment Constraints, there exists Grim Trigger Strategy Equilibria for majority punishment strategies of either unitary democracy or administrative federalism forever, where (i) democratic federalism is sustainable, and (ii) \( g^{\max}(U) > g^{\max}(A) \) in the q-Regime and \( g^{\max^*}(U) > g^{\max^*}(A) \) in the q*-Regime. (Proof: See the Appendix)

When the majority has a large and well-trained pool of public employees it will be able to negotiate, or set in annual budgets, a relatively high level of redistributive grants. This bargaining advantage has the downside, however, of discouraging the elite from accepting democratic federalism at the time of the transition. Democratic federalism is most likely an jointly acceptable, and therefore sustainable, constitution when both parties to the negotiations are needed for democracy’s long-run success. When the majority offers peace and the elite provides efficiency in the production of valued public services, then democratic federalism, and perhaps democracy itself, has a chance. Sections IV and V present evidence for South Africa as a case on point.

**IV. Is South Africa’s Federal Democracy Sustainable?**

There is little doubt that South Africa’s transition from autocracy to democracy would not have occurred were the majority ANC and the minority NP and IFP negotiators not willing to accept an interim federal constitution with at least one politically protected province for each minority party, and with each province promised significant, but not fully autonomous, fiscal powers. It was only after this agreement was
reached on April 19, 1994 that democratic elections went forward.\textsuperscript{32}

The resulting 1994 Interim Constitution left many details unspecified, however. To set the constitutional rules for fiscal policy, the interim constitution created the Financial and Fiscal Commission (FFC) with equal representation from the ANC and old ruling National Party.\textsuperscript{33} With equal representation on the commission, both parties to the new democracy could block constitutional recommendations by the FFC. In fact, all FFC recommendations were unanimously supported by the commission members. The FFC recommended a fiscal assignment that allocated all important taxing powers to the central government and the three important redistributive services – K-12 education, primary health care, and social security (welfare) grants – to the provinces. Without significant taxing powers, the provinces required centrally funded intergovernmental transfers. The FFC recommended that the aggregate level of grants be set centrally by the Ministry of Finance in consultation with the FFC and that the allocation of total grants across provinces be set by an FFC-specified funding formula that ensured adequate financing for provincially provided redistributive services and a “significant” basic grant for general administration and other provincial services. The FFC’s recommendations were unanimously approved as part of the final 1996 Constitution.

Table 2 summarizes the resulting system of federal finance. The table reports the total central government tax revenues per capita, total redistributive grants per capita, the allocation of total grants to the elite (Western Cape) and majority (Other) provinces, provincial grants per capita needed to fund assigned services in the elite and majority provinces, and provincial basic grants per capita to fund other provincial

\textsuperscript{32} Waldmeir (1997, Chapter 13).

\textsuperscript{33} The interim constitution required the FFC to be composed of one representative appointed by the executive committees of each of the nine provinces and nine at-large representatives appointed by Nelson Mandela. No members of the FFC were to hold party office nor be candidates for public office and all were to be recognized experts in either economics, accounting, public administration, or taxation. President Mandela used his appointments to ensure a numerical balance between ANC and NP representation on the commission. Similarly balanced commissions were established to draft the other major sections of the final constitution, those on human rights, legislative and presidential powers, courts and administration justice, and public administration.
services in the elite and majority provinces.\textsuperscript{34}

Four conclusions are evident from Table 2. First, provincial governments have been given a significant and growing role in the provision of the important redistributive services in South Africa, funded entirely by grants from the central government. Second, the Western Cape receives approximately 20 percent less in assigned services grants as required by the relative differences in public employee productivity. Third, while most (\textasciitilde{75} percent) of the allocated grant monies are targeted towards meeting assigned redistributive services, basic service grants are significant and do allow for capture. Fourth, the initially high level of basic grants given to the Western Cape have been gradually reduced to levels more consistent with majority provinces, particularly beginning in FY 2003/04. In the early years of the democracy, there was a conscious decision by the FFC to ease the transition into the new democracy for the elite province. It was feared that large initial cuts in discretionary spending in the Western Cape would lead to sudden lay-offs of white public employees, an important political ally of the National Party at the time of democratic negotiations.\textsuperscript{35}

While funding patterns are consistent with the description of the democratic federalism, the evidence in Table 2 only begins the analysis of our fundamental question: Is the South African federal constitution sustainable? To answer this question we evaluate whether the requirements of Lemmas 1 and 2 and Proposition 1 are met by the South African political economy. We do so in five steps. First, we specify the model’s required economic and demographic parameters for the South African economy at the time of the transition; see Table 3. Second, we translate the new constitution’s border and assignment rules into estimates for the border and assignment constraints. Third, we confirm that the constitution’s initial constraints, denoted as $\mu$ and $q^{FFC}$, meet Lemma 1’s conditions for a credible elite punishment strategy.


\textsuperscript{35} Waldmeir (1997, p. 228).
Fourth, using Proposition 1 and our calibration of the South African political economy, we compute the range of aggregate grants consistent with sustainable democratic federalism. We assume \( g_e \) equal to the maximal redistributive grant consistent with democratic federalism. Given \( q^{FFC} \) we then compute the constitution’s required service grants for elite and majority provinces, and then finally, the level of basic grants for elite and majority provinces. Predicted grants are reported in Table 4. While total predicted and actual intergovernmental grants are quite close – and consistent with sustainable democratic federalism – predicted grants needed to fund \( q^{FFC} \) are well below actual service grants for all fiscal years FY 2001 and beyond. The results are suggestive of a shift in the enforcement of assigned services, from a \( q^- \) to a \( q^* \)-Regime. It is suggestive that the FY 2001 budget was the first decided by a post-Mandela ANC central government. Fifth, we therefore estimate the ANC’s revealed preference for redistributive public goods under the assumption that South Africa is now in a \( q^* \)-Regime and compute \( q^*_L = q^{ANC} \). We then examine whether \( \mu \) and \( q^{ANC} \) satisfy the \( q^* \)-Border and \( q^* \)-Assignment Constraints of Lemma 2. They do, barely. Against the requirements of Lemmas 1 and 2 and Proposition 1, we conclude that South Africa’s new federal constitution is currently sustainable.

A. Calibration: Table 3 summarizes our best estimates of our model’s required economic and demographic parameters for South Africa at the time of its political transition.\[36\] Using South African Census data for 1991 and 2001, we estimate the initial (pre-tax) elite and majority adult (voting) age populations as \( N_0 = 9.6 \) million and \( M = 25 \) million residents and elite and majority pre-tax incomes as \( Y = 86,000 \) (real 2000) Rand/Elite Adult (~$22,000 USD in 1996) and \( W = 9,700 \) (real 2000) Rand/Majority Adult (~$2,500 USD), respectively. No South African data are available to estimate \( \beta \), the elite’s response to redistributive taxation. Given this uncertainty we adopt a conservative estimate for \( \beta = .00015 \), one making federalism
The calibration for \( \phi^e \) was chosen to imply a plausible peak to the revenue hill from the taxation of elite residents. Setting \( \hat{\beta} = .00015 \) implies a revenue-maximizing redistributive tax rate per elite resident of 32,000 Rand, or approximately 37 percent as the maximal share of pre-tax elite income that can be allocated to fiscal redistribution. The protest costs borne by elite residents from engaging in high capture was set at .02 of elite resident income or \( \rho = 1720 \) (real 2000) Rand/Elite Resident; .02 is our estimate of the cost as a percent of income imposed on elite residents’ by Black unions’ political protests during the latter stages of apartheid.

We specify public employee productivity to be proportional to years of training and use as our estimates of \( a_e, a_m, \) and \( a_u \) the years of schooling of white (elite) teachers (= 17 years), certified Black and Colored teachers \( (a_m = 14 \) years), and uncertified Black teachers \( (a_u = 10 \) years), respectively.\(^{39}\) The fraction of the majority population that can be serviced by trained majority public employees \( (m) \) was computed knowing the eligible majority population and the number of trained majority teachers and primary care medical providers: \( m = .85.\(^{40}\) \)

We adopt as the estimate of the common public employee salary \( S \) the average salary for teachers in the Western and Eastern Capes in 2001.\(^{41}\) Given the estimates of \( a_e, a_m, a_u, S, \) and \( m \), the costs per majority

\(^{37}\) The calibration for \( \hat{\beta} \) was chosen to imply a plausible peak to the revenue hill from the taxation of elite residents. Setting \( \hat{\beta} = .00015 \) implies a revenue-maximizing redistributive tax rate per elite resident of 32,000 Rand, or approximately 37 percent as the maximal share of pre-tax elite income that can be allocated to fiscal redistribution. We estimate the required tax rate on individual personal incomes needed to fund central government expenditures other than redistributive aid as approximately 30 percent; Financial and Fiscal Commission (1998), Tables 2.1 and 2.6 and Development Bank of South Africa (1995), Table 60. Assuming taxation for redistributive aid becomes an added burden for the elite, the choice of \( \hat{\beta} \) implies a revenue maximizing tax rate for the entire central government budget of approximately 67 percent \( (= .67 = .37 + .30) \). As a possible point of comparison, Gruber and Saez (2002, p. 26) estimate the revenue maximizing tax rate on taxable personal income for U.S. households at 71 percent.

\(^{38}\) See Reinikka and Svensson (2003, 2004).

\(^{39}\) Fiske and Ladd (2004), Tables 6.6 and 6.7.

\(^{40}\) Fiske and Ladd (2004), Table 6.7 for teachers, and the Development Bank of South Africa (1995), Table 26 for health care professionals. For details of this estimate, see our Data Appendix, available upon request.

\(^{41}\) Fiske and Ladd (2004), Tables 6.5 and 6.6.
resident of providing any level of redistributive services using alternative providers are given in Table 3.

To ensure the cooperation of the Zulu Chief Buthelezi and his Inkatha Freedom Party at the time of the interim constitution, special payments were implicitly promised to the new KwaZulu-Natal province. We assume the amount paid equaled the payments to the original KwaZulu homeland under the old apartheid regime: \( Z = 600 \text{ Million (Real 2000)} \text{ Rand.} \)

The discount factor \( \delta \) is specified to equal \((1/1+r)\) where \( r \) is the rate of time preference for elite and majority residents equal to the real rate of return on ten year South African bonds, averaged over the period 1996-2006; \( r = .03.^{42} \) Thus \( \delta = .97. \)

Finally, we compute the revenue maximizing redistributive tax as \( \tau_u = N_u/2\beta = 32,000 \text{ (real 2000)} \text{ Rand/Elite Adult,} \) the implied tax-paying elite population \( N(\tau_u) = 4.8 \text{ million elite adults,} \) and the maximal grant per majority adult \( g_u = [(\tau_u \cdot N(\tau_u) - Z)/M = 6,120 \text{ (real 2000)} \text{ Rand/Majority Adult.} \) The maximal redistributive tax and grant are expected to be chosen by the majority under unitary democracy.

**B. The Constitution’s Original Border and Assignment Constraints:** The 1996 Constitution created two provinces meant to accommodate the old ruling elite, the urbanized Western Cape centered around Cape Town and the rural Northern Cape.\(^43\) Only the Western Cape has remained in elite political control. While the constitutional allocation of the voting age population was sufficient to ensure elite control of the Northern Cape, white voter turnout has always been too low to give an elite victory. Thus we consider the Northern Cape a majority province. A value of \( \mu = .184 \text{ allocating no more than 18.4 percent of the majority population into the Western Cape is sufficient to ensure continued political control of this province by the elite.}^{44} \)

---

\(^{42}\) Source: [www.reservebank.co.za](http://www.reservebank.co.za); Series KB2003J.

\(^{43}\) See Muthien and Khosa (1998).

\(^{44}\) In the 2004 Western Cape elections, the ANC won 45 percent of the vote, while a four-party coalition of elite parties won 51 percent of the vote; see [www.elections.org.za](http://www.elections.org.za). We therefore define \( \mu = (M_j/M) \) so that \( N(\tau_j)/[M_j + N(\tau_j)] = .51. \) \( N(\tau_j) \) is estimated to be only slightly larger than 4.8 Million since the equilibrium value of \( \tau_j \) is shown to be very
The 1996 Constitution empowered the FFC to set the constitution’s standards for redistributive services. The FFC’s minimum requirements were at least 1 teacher per 38 school-aged children, 3.5 preventive health care clinic visits a year for each majority adult and child for health care, and 4500 (real 2000) Rand for each eligible (elderly, disabled, child in poverty) majority resident for social security transfers. Together, these targets require redistributive grants sufficient to fund the equivalent of .038 public employees per majority resident. Given \( q = a\left(\frac{X}{M}\right) \), and assuming an average value of \( a = 14 \) (years), we estimate the FFC’s minimum mandated service level to be \( q^{\text{FFC}} = 14\cdot(0.038) = .53 \).

C. Is the Elite Punishment Strategy Credible? Figure 1 shows the set of all border and assignment requirements of Lemma 1 required to ensure the elite’s punishment strategy of high capture (\( \varphi^H \)) is credible under the assumption that South Africa is in a \( q \)-Regime. The original constitutional standard of \( q^{\text{FFC}} \) were set by the FFC, a constitutional commission with equal ANC majority and NP elite representation. The FY 1996-2000 budgets submitted by the Finance Ministry under the (presumed) directive of President Mandela respected the FFC recommendations, holding to the \( q^{\text{FFC}} \) constraint; see Table 2.

The fully shaded area in Figure 1 shows all constitutionally chosen values of \( \mu \) and \( q \) that will satisfy Lemma 1’s border and assignment constraints for South Africa’s calibrated political economy. The vertical \( \mu \) axis of Figure 1 defines the border requirements for the new constitution. The upper bound for \( \mu \) is the maximal share of the majority population that can be allocated to the province to ensure that the elite is close to \( \tau_q \); see Table 4. Thus \( M_e = 4.6 \text{ Million} \), and \( \mu = (M_e/M) = 4.6M/25M = .184 \).

The value of \( (X/M) \) is estimated from FFC targets and the assumption that (1) each majority adult has one child requiring .026 education professionals per majority resident; (2) each medical professional can provide 3.5 visits to each of 500 majority residents a year requiring .002 health care professionals per majority resident; and (3) that approximately 17 percent of the majority population qualifies for some form of income assistance for an average spending per majority resident of 765 Rand per year or, in fiscally equivalent units of a public employee, about .0095 public employees per majority resident. Together the mandates require funding sufficient to pay for .0375 public employees per majority resident.

We can offer only anecdotal evidence on President Mandela’s role here. The authors served as advisors to the FFC from 1994-1999 and to the Finance Ministry from 1996-1999. It is our impression that during this period the Finance Ministry accepted the FFC recommendations and forwarded those funding requests, largely unaltered, to Parliament. The ANC dominated Parliament then approved the Ministry’s budgets with no changes.
always in political control: \( \mu^{\text{max}} = N(\tau_e)/M = 4.8m./25m. = .192 \). The lower bound for \( \mu \) shown as the line \( \mu^{\text{min}}(q) \) defines the share of majority population that must be allocated to the elite province to ensure enough “hostages” so that the elite’s choice of \( \varphi^H \) is a credible punishment. The horizontal axis \( q \) represents the required level of public inputs to meet constitutional assignment, measured as the number of public employee training years per majority resident. The upper bound for redistributive assignment is specified as \( q^{\text{max}} = .74 \). The lower bound for assignment is specified the line \( q^{\text{min}}(\mu) \). Assuming a Mandela enforced \( q \)-Regime, Figure 1 shows the original specification of the South African constitution as \( \mu = .184 \) and \( q^{\text{FFC}} = .53 \) met the border and assignment requirements of Lemma 1 for credible elite punishment and therefore for feasible democratic federalism.

**D. Fiscal Policies and Sustainable Federalism**: Proposition 1 states that a federal constitution satisfying a political economy’s border and assignment constraints can be sustained when intergovernmental transfers are bounded as \( g^{\text{max}} \geq g_F > g^{\text{min}} \), where the precise specification of the bounds depends upon the majority’s preferred punishment strategy. When calibrated to the South African political economy, we estimate the majority’s preferred punishment when the elite defects is to move to unitary governance.\(^{48}\)

Table 4 reports the estimated values for the bounds on aggregate redistributive transfers \( g^{\text{max}}(U) \) and \( g^{\text{min}}(U) \) when unitary governance is the majority’s preferred punishment. Assuming the ANC majority chooses the maximal level of redistributive spending consistent with sustainable federalism, then \( g_F = g^{\text{max}}(U) = 3365 \) Rand per capita. For the constitution assignment of \( q^{\text{FFC}} = .53 \) and the cost specifications in Table 3, we estimate assigned service grants for the elite province, \( s_e(q^{\text{FFC}}) \), to be 1372 Rand per capita and the assigned service grants in the majority provinces, \( s_m(q^{\text{FFC}}) \), to be 1665 Rand per capita; see Table 4. The provinces’ unconstrained basic grants revenues, \( b_e \) and \( b_m \), are the differences between total grants and

---

\(^{47}\) See fn. 24 above.

\(^{48}\) Unitary governance is the preferred punishment when \( \omega(U) \geq \omega(A; \varphi^H) \). For our calibration of the South African economy, \( \omega(U) = \mu \varphi^H [g_F - s_e(q^{\text{FFC}})] = (.18)(.85)[6120 - 2494] = 555 > 276 = [3210 - 2934] = [s_e(q^{\text{FFC}}) - s_F(q^{\text{FFC}})] = \omega(A; \varphi^H). \)
mandated service grants and equal 1993 Rand per capita in the elite province and 1700 Rand per capita in the typical majority province; see Table 4.

[INSERT TABLE 4 HERE]

Two trends are evident when we compare predicted (Table 4) and actual (Table 2) intergovernmental transfers. First, total transfers in Table 2, below those predicted in Table 4 for sustainable democratic federalism in the Mandela years, are now very close to predicted $g^{\text{max}}(U)$. Second, while the constitutional service assignment ($\lambda$) has remained unchanged since 1996 – still K-12 education, basic health care, and administration of social service grants – predicted service grants assuming $q^{\text{FFC}} = .53$ are well below Table 2's actual funding for assigned services. A trend break between predicted and actual assigned service funding occurs between FY 2000 and FY 2001, the last budget from President Mandela and the first budget under newly elected President Mbeki. The break is suggestive of a regime shift, from a Mandela enforced $q$-Regime at $q^{\text{FFC}} = .53$, to an Mbeki majority rule $q^*$-Regime, where $q$ is allowed to reflect ANC preferences for the assigned services. From the cost specifications in Table 3 and Table 2's actual transfers for mandated services to elite and majority provinces, we can estimate the ANC’s revealed preferred service level: $q^{\text{ANC}} = .65$ (s.e. = .07) for the sample period FY2001-2006.\(^{49}\)

Assuming the Mbeki presidency is a $q^*$-Regime, we must re-evaluate the feasibility of democratic federalism against the requirements of Lemma 2. To do so requires an estimate of $\lambda$ for the constitution’s currently assigned provincial services. We specify $\lambda \cdot u(q) = \lambda \cdot \ln(q)$. For budget years FY2001-2006, we estimate $\lambda = 3487$ (s.e. = 416).\(^{50}\) Given $\lambda = 3487$, we compute the $q_{L \min}(\mu, \lambda)$ and $\mu^{\text{min}}(q^*_L)$ schedules

---

\(^{49}\) For each of the six Mbeki fiscal years, we can estimate a separate value for $q$ that can be funded by the per capita grants reported in Table 2, adjusted upward by dividing by the ratio of majority adults to total residents ($= .55$) to be comparable to the costs of $q$ per majority adult resident. In each fiscal year, $q_e = s_e(q)/4706$.55 and $q_m = s_m(q)/5714$.55, where $s_e(q)$ and $s_m(q)$ are as reported in Table 2. The average value of $q$ over the six observations is .65 (s.e. = .07).

\(^{50}\) The additive specification of public service preferences as $\lambda \cdot u(q) = \lambda \cdot \ln(q)$ implies a zero income elasticity of demand (from additive preferences) and a unitary (-1) price elasticity of demand (from $\ln(q)$): $q = \lambda / p$, where $p$ is the unit cost of $q$. Assuming the data in Table 2 is from a democratic federal regime then $p = p_L(\mu) = s_L(q) \cdot \Phi^\mu \cdot s_L(q)$. Thus $\lambda = p_L(\mu) = s_L(q) - \Phi^\mu \cdot s_L(q) = (1-\mu) s_m(q) + \mu (1-\Phi^\mu) s_e(q)$. For FT 2001-2006, we have six observations for
which, along with $\mu^{\text{max}} = .192$, define the set of borders ($\mu$) and assignments ($\lambda$, equivalently $q^*_L$) which can satisfy the $q^*$-Border and $q^*$-Assignment constraints of Lemma 2; see Figure 3. President Mbeki’s administration has preferences for assigned provincial services which lie just outside Lemma 2’s feasible set for a credible elite punishment. But not significantly so. The upper bound for the $q^*$-Assignment constraint is $q^*_L^{\text{max}} = .64$, within the confidence interval for our estimate of $q^{\text{ANC}} = .65$ (s.e. = .07). Though pushing the limits of sustainable federalism, both with respect to the growth in $g^\text{max}_F$ and the chosen level of redistributive services, we conclude that President Mbeki has maintained South Africa’s constitutional commitment to democratic federalism.

V. Governance and the Economic Value of Transition

Did South Africa’s elite make the correct economic decision when they abandoned apartheid for democratic federalism? Under democratic federalism, the elite pays significant redistributive taxes. Under apartheid, the elite paid lower redistributive taxes, but incurred the military costs of running a repressive regime, and in addition, bore potentially significant economic costs from international sanctions and domestic labor unrest. Table 5 provides estimates of the economic costs to the elite of the apartheid regime.

---

51 It is worth exploring how sensitive the conclusions are to the calibration choices in Table 3. Conditional on the values of all other parameters, increasing either $a_m$ from 10 to 13.5 years of training or the share $m$ of majority residents covered by trained public employees from .85 to .98 will so lower the cost of unitary governance that the elite will be unable to rationally prevent the majority’s adoption of unitary governance. That is, $g^{\text{min}}(U) > g^{\text{max}}(U)$, and there will be no values of $g_F$ that can satisfy the minimal bargaining demands of both the majority and the elite. Both $a_m$ and $m$ increase if the ANC expands the training of majority teachers, nurses, and doctors, but to date the ANC has continued to rely upon white elite providers of public services; see Fiske and Ladd (2004, Chapter 10).

Reducing the discount factor $\delta$ from .97 to .89 also threatens the long-run viability of the federal compact. For $\delta = .89$ ($\tau = .11$), the majority demands an increase in current year redistribution which exceeds the willingness of the elite to contribute, and again, $g^{\text{min}}(U) > g^{\text{max}}(U)$ and no federal agreement can be sustained. The fact that South Africa has maintained fiscal discipline at current world interest rates, suggests $\tau < .11$.

Finally, holding fixed all other parameters, increasing $\rho$ above 2950 Rand, narrowing the gap between high and low capture so that $(\phi^H - \phi^L) < .38$, raising the exit elasticity to $\beta > .00025$, or lowering the initial population ratio $(N_0/M) < .30$ will shrink the shaded areas of feasible federalism in Figures 1 and 2 to zero. In these cases, Lemmas 1 or 2 cannot hold for any values of $\mu$, $\lambda$, and $q$. In such cases, elite punishment is not credible and democratic federalism cannot be sustained. For South Africa, however, our estimated values of $\rho$, $(\phi^H - \phi^L)$, $\beta$, and $(N_0/M)$ are comfortably within the range needed for credible elite punishments.
Table 6 compares the elite’s potential improvement in lifetime economic welfare under unitary and federal democracies over their economic prospects from maintaining apartheid. As an economic matter alone, we find the elite made the correct decision in opting for democratic federalism, and indeed, they would have preferred unitary governance to apartheid too.

[INSERT TABLE 5 HERE]

Apartheid was maintained in two ways, armed repression and the geographic isolation of Blacks in homeland territories. Both were expensive. Table 5 reports the average annual expenditures on military, police and justice (prison) services for the apartheid years following the 1976 Soweto massacre, generally regarded as a defining (radicalizing) moment in majority resistance, both violent and non-violent, to the apartheid regime. Average military spending per elite resident over this period was 3,510 Rand (~$600 2000 USD) per year. Homeland territories were rural and provided residents with only subsistence agriculture. By 1985, two-thirds of South Africa’s Blacks lived in one of the ten homelands; the remaining third lived on white owned farms or in the segregated townships outside the major cities. Basic public services to the homelands were provided by subsidies from the white central government at an average annual cost per elite resident of 2,740 Rand (~$450 2000 USD) per year. Together the fiscal burden of military and homeland spending was 6,250 Rand per year or a bit more than 7 percent of the average elite resident’s annual income of 86,000 Rand – significant, but still a good deal less than the estimated 40 percent tax burden on elite incomes under democratic federalism; see Table 5.

The real costs for the elite under apartheid came from the damage that international sanctions and domestic labor unrest imposed on economic growth; see Table 5. Beginning in 1976, and motivated by the Soweto massacre, the international community implemented a series of increasingly stringent sanctions on imports to, and exports from, South Africa. They remained in place until the end of 1993, when the transition and the election of President Mandela seemed assured. Their combined effect, measured by the variable SANC (= 1 for 1976-1993; 0 otherwise), was to reduce economic growth by an average of 1.55
percent per annum (Table 5, col.1). The channel through which sanctions lowered growth was the trade channel; Table 5, cols.2-5. More important was the threat of domestic labor unrest measured by the emergence in 1985 of a militant trade union federation called the Congress of South African Trade Unions (COSATU). Over the period of COSATU militancy (= 1 for 1985-1993; 0 otherwise), economic growth fell an additional 2.1 percent per annum (Table 5, col.1), from the joint adverse effects of labor unrest on investment and trade (Table 5, cols. 2-5). Overall, the South African economy grew at an annual rate of 2.43 percent during the peaceful years of apartheid (1950-1975), then declined to .88 percent when only sanctions were in force (1976-1984), and finally fell to -1.2 percent as both sanctions and COSATU discouraged trade and investment (1985-1993). With the election of Nelson Mandela as President, growth became positive again, achieving an average annual rate of .67 in the first six years of democracy (1994-2000). Since 2000, we estimate the economy has been growing at an average annual rate of 2.5 percent.\(^{53}\) Moving from negative growth to an annual growth rate of 2 percent or more will have sizeable long-run implications for economic welfare. It is here that the real benefits to the elite of leaving apartheid are realized.

Table 6 computes the long-run benefits from leaving apartheid, inclusive of the net differences in elite taxation under the three relevant regimes: apartheid, unitary democracy, and federal democracy.\(^ {54}\) The annual fiscal costs of apartheid have been estimated above as 6,250 Rand per year. In democracy, the annual fiscal costs will be 32,000 Rand per elite resident from redistributive taxation under unitary governance and

\(^{52}\) Like other country growth regressions, we find trade openness (\(\text{OPEN} = \text{exports} + \text{imports} \text{ as a percent of GDP}\)) is an important determinant of growth; Table 5, columns (2) and (3). Sanctions had a significant adverse effect on trade openness’ Table 5 column (4). Sanctions began with the United Nations arms embargo in 1976 just after Soweto. Beginning in 1978, the international capital markets refused to make long-term loans to the apartheid government (Waldmeir, 1997, p. 23). In 1985, Chase Manhattan Bank withdrew all short-term lending and other major banks then followed (Waldmeir, 1997, p. 56), and the EC and the British Commonwealth countries imposed trade and financial sanctions. In 1986, the U.S. Congress enacted the Comprehensive Anti-Apartheid Act, banning U.S. investments and loans to South Africa and imports from South Africa. All sanctions remained in place until 1993. See also Hufbauer, Schott, and Elliott (1990; pp. 221-248).

\(^{53}\) Source: South African Reserve Bank, www.reservebank.co.za, Time Series KBP6244J.

\(^{54}\) We consider only three regimes for South Africa since, for our calibration, the majority prefers unitary democracy to administrative federalism as the strategic alternative to democratic federalism; see Table 4.
31,175 Rand per elite resident from redistributive taxation (net of capture) under federal governance.\(^{55}\) Post apartheid, there are no homeland payments but there annual expenditures for military, police and justice averaging 39.1 Billion (2000) Rand per year for the years 1994-2006, an additional fiscal burden on the typical elite resident of 8,145 Rand per year.\(^{56}\) These higher annual fiscal burdens under democracy – 40,145 under unitary governance and 39,320 Rand under federal governance vs. 6,250 Rand under apartheid – are eventually offset, however, by the democracy’s higher rate of economic growth. For elite residents, the present value benefit of democracy’s growth dividend discounted at \(r = .03 (\delta = .97)\) just begins to dominate the present value cost of democracy’s higher taxes after 22 years (year 2018) under unitary, and after 21 years (year 2017) under federal governance; see Table 6. Of course, the majority benefits immediately.

For elite residents, the long-run economic welfare from the transition can be sizeable. After 70 years, or roughly two adult generations, the net gains for a representative elite adult from leaving autocracy for democracy (\(\Delta NPV\) in Table 6) is 2.197 Million Rand (\(\approx 370,000\) (2000) USD) under democratic federalism and 2.179 Million Rand (\(\approx 363,000\) (2000) USD) under unitary governance. The elite prefers democratic federalism to unitary governance since by managing their own province, the elite can and does capture back a portion of their redistributive tax payments to the central government. The internal rates of return are .118 for the transition to democratic federalism and .115 for the transition to a unitary democracy; see Table 6.

---

\(^{55}\) Estimated at the time of the democratic transition. In both unitary and federal democratic equilibriums, there will be approximately 4.8 million adult elite residents and 25 million adult majority residents. Taxes paid to the central government by each elite resident will be 32,000 Rand to support \(g_U\) and \((25/4.8)\times6119 = 31,870\) to support \(g_F\); see Table 4. Under democratic federalism, however, the elite are able to capture .2 of the difference between \(g_F\) and \(s_j(q^{\text{PC}})\) or .20\([g_F - s_j(q^{\text{PC}})][M_j/N(\tau_j)] = .20[6119 - 2494](4.6/4.8) = 695\) Rand/Elite Resident. The net tax burden will therefore be 31,870 - 695 = 31,175 Rand.

\(^{56}\) Source: Table 5. Elite population declines in post-apartheid South Africa to 4.8 million in our calibrations. Thus, the burden elite resident becomes 8,145 Rand (= 39,100 Million Rand/4.8 Million Elite Adults). If one views homeland transfers as a “protection expenditure” – in effect, the cost of containing Blacks in segregated areas – then the total costs of protection under apartheid should be measured as military spending plus homeland spending or (Table 5) 60 Billion Rand per year (= 33.7 Billion + 26.3 Billion). In this light, democracy provides an annual “peace dividend” of about 20.9 Billion Rand (= 60 Billion - 39.1 Billion). In our framework these funds are being allocated to redistributive spending for the majority.
Only elite residents with a real rate of time preference greater than 12 percent ($\delta < .89$) would have found it rational to have rejected the transition.57

Finally, the federal form of governance is economically preferred by both the elite and majority residents of South Africa to unitary governance; see Table 6. From the perspective of Table 1, sustainable democratic federalism (Cell 1) is pareto preferred to its non-cooperative, strategic alternative of unitary governance (Cells 3 or 6). From the perspective of a democratic transition, however, both forms of governance are preferred to apartheid. We conclude, democratic federalism was a sufficient institution for South Africa’s transition, but from an economic perspective alone, it was not a necessary institution.

VI. Concluding Remarks

Our analysis of the South African transition suggests at least four lessons for those seeking to move from autocracy to democracy. First, a peaceful transition requires that both the old ruling elite and the new majority willingly accept the new constitution. For elite approval, the new constitution must offer credible, long-run protection of elite resources in a self-enforcing way, without reliance on outside military intervention or unspecified moral suasion. We have specified the economic and demographic conditions for when one form of governance, called democratic federalism, might successfully provide such protections. The South African constitution provides one example and perhaps a role model for other emerging democracies.

Second, for democracy to be attractive to autocracy’s elite, costs at least as large as democracy’s anticipated redistributive burden must be imposed upon the elite in the autocratic regime. Those costs may

57 The keys to democracy high rate of return for South Africa was in removing international sanctions and politically active Black unions as impediments to economic growth. Two counterfactuals are worth considering. Absent international sanctions but assuming continued domestic political resistance by COSATU, elite net economic returns from the transition would have only become positive in a present value sense ($\Delta NPV = 0$) after 32 years from the start of democracy. The implied internal rate of return for accepting democracy over apartheid falls to .086. Absent domestic political resistance but assuming continued international sanctions, the elite breaks even ($\Delta NPV = 0$) following the transition only after 43 years from the start of democracy, and the internal rate of return of the transition falls further to .063. These calculations suggest international sanctions alone might not have been enough to push white South Africa to accept an ANC majority-run democracy. Indeed, comments by the NP leadership at the time of the transition suggest they saw international sanctions more as a nuisance than a significant economic hardship; see Waldmeir (1997; p. 134).
come from military spending to prevent a majority coup d’état, from the economic losses following majority boycotts and strikes, or from the burden of international economic sanctions. By credibly controlling democracy’s redistributive burden, democratic federalism lowers the threshold for required costs in autocracy and thereby makes a peaceful transition to democracy more likely.

Third, our analysis makes clear exactly what is needed for the elite to peacefully check the majority’s desire to impose maximal redistributive taxation in the new democratic regime. Having given up the army, and thus the threat of a counter coup d’état, the elite needs another means to penalize the majority if they deviate to maximal redistribution. In our analysis, that penalty is found in the elite’s ability to deny the majority access to high productivity, low cost elite providers of important redistributive public services, for example, education, health care, and social services. What the federal constitution provides is an institutional structure for organizing elite providers into a blocking cartel by creating an elite province hiring elite providers with the monopoly right to provide redistributive services to majority residents in the province. Using provincial government has the additional advantage of providing a legal channel through which to distribute the cartel’s captured surplus to elite residents, perhaps most plausibly as extra services targeted to elite neighborhoods. In this way it is very useful to have the elite concentrated geographically. An elite province created for a landed aristocracy spread throughout the country may be difficult to administer. Practically speaking, sustainable democratic federalism may require a significant, “first-world” urban center.

Finally, while democratic federalism can facilitate the transition to democracy, it is a fragile constitutional form. South Africa, for example, is close to the point where the elite, as a political minority in national politics, may no longer be able to discourage maximal redistributive taxation through its ability to control provincial fiscal policies. Its threat to deny transfer incomes to its province’s majority residents by a “high capture” strategy may no longer be credible. If so, the federal constitution’s protections of elite resources will be lost and, at least with respect to redistributive economic policy, unitary governance
becomes the *de facto* constitutional form. Other than by re-capturing government through its own military coup d’état, the elite’s only hope will be to find new political allies within the majority, perhaps as part of new, locally based political parties or through alliances with high growth majority controlled provinces. In this regard, our work here on federalism as a means to facilitate the democratic transition complements the classic work by Riker (1964) and the recent analysis of de Figueiredo and Weingast (2005) on sustaining the federal compact in established democracies.
REFERENCES


TABLES

TABLE 1: THE POLICY GAME

TABLE 2: RSA INTERGOVERNMENTAL TRANSFERS

TABLE 3: POLITICAL ECONOMY OF SOUTH AFRICA

TABLE 4: RSA INTERGOVERNMENTAL TRANSFERS: PREDICTED

TABLE 5: THE ECONOMIC BURDEN OF APARTHEID

TABLE 6: NET ECONOMIC GAINS FROM THE DEMOCRATIC TRANSITION
**TABLE 1: THE POLICY GAME**

*MAJORITY PAYOFFS per MAJORITY CITIZEN*

<table>
<thead>
<tr>
<th>STRATEGIES</th>
<th>MAJORITY</th>
<th>MAJORITY</th>
<th>MAJORITY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>STRATEGY F: (τ_F; PROVINCES &gt; 1)</td>
<td>STRATEGY A: (τ_U; PROVINCES &gt; 1)</td>
<td>STRATEGY U: (τ_U)</td>
</tr>
<tr>
<td>ELITE</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>STRATEGY: φ^L</td>
<td>(\omega(F; \varphi^L) = W + g_F[1 - \varphi^L - \mu] - s_F(q) + \varphi^L \cdot \mu \cdot s_F(q) + \lambda u(q))</td>
<td>(\omega(A; \varphi^L) = W + g_U[1 - \varphi^L - \mu] - s_F(q) + \varphi^L \cdot \mu \cdot s_F(q) + \lambda u(q))</td>
<td>(\omega(U) = W + g_U - s_U(q) + \lambda u(q))</td>
</tr>
<tr>
<td>ELITE</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>STRATEGY: φ^H</td>
<td>(\omega(F; \varphi^H) = W + g_F[1 - \varphi^H - \mu] - s_F(q) + \varphi^H \cdot \mu \cdot s_F(q) + \lambda u(q))</td>
<td>(\omega(A; \varphi^H) = W + g_U[1 - \varphi^H - \mu] - s_F(q) + \varphi^H \cdot \mu \cdot s_F(q) + \lambda u(q))</td>
<td>(\omega(U) = W + g_U - s_U(q) + \lambda u(q))</td>
</tr>
</tbody>
</table>

*ELITE PAYOFFS per ELITE CITIZEN*

<table>
<thead>
<tr>
<th>STRATEGIES</th>
<th>MAJORITY</th>
<th>MAJORITY</th>
<th>MAJORITY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>STRATEGY F: (τ_F; PROVINCES &gt; 1)</td>
<td>STRATEGY A: (τ_U; PROVINCES &gt; 1)</td>
<td>STRATEGY U: (τ_U)</td>
</tr>
<tr>
<td>ELITE</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>STRATEGY: φ^L</td>
<td>(y(F; \varphi^L) = Y - \tau_F + \varphi^L[g_F - s_F(q)] / [M_F/N(\tau_F)])</td>
<td>(y(A; \varphi^L) = Y - \tau_U + \varphi^L[g_U - s_F(q)] / [M_U/N(\tau_U)])</td>
<td>(y(U) = Y - \tau_U)</td>
</tr>
<tr>
<td>ELITE</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>STRATEGY: φ^H</td>
<td>(y(F; \varphi^H) = [Y - \rho] - \tau_F + \varphi^H[g_F - s_F(q)] / [M_F/N(\tau_F)])</td>
<td>(y(A; \varphi^H) = [Y - \rho] - \tau_U + \varphi^H[g_U - s_F(q)] / [M_U/N(\tau_U)])</td>
<td>(y(U) = Y - \tau_U)</td>
</tr>
<tr>
<td>FISCAL YEAR</td>
<td>All Central Government Revenues</td>
<td>( g_f ) National Average</td>
<td>( g_f ) Western Cape</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>1995/96</td>
<td>4237</td>
<td>2189</td>
<td>2923</td>
</tr>
<tr>
<td>1996/97</td>
<td>3938</td>
<td>2030</td>
<td>2587</td>
</tr>
<tr>
<td>1997/98</td>
<td>(3942)*</td>
<td>(2000)*</td>
<td>(2424)*</td>
</tr>
<tr>
<td>1998/99</td>
<td>4265</td>
<td>2154</td>
<td>2206</td>
</tr>
<tr>
<td>1999/00</td>
<td>4093</td>
<td>2108</td>
<td>2097</td>
</tr>
<tr>
<td>2000/01</td>
<td>6636</td>
<td>2242</td>
<td>2185</td>
</tr>
<tr>
<td>2001/02</td>
<td>5570</td>
<td>2302</td>
<td>2196</td>
</tr>
<tr>
<td>2002/03</td>
<td>5674</td>
<td>2430</td>
<td>2254</td>
</tr>
<tr>
<td>2003/04</td>
<td>5928</td>
<td>2640</td>
<td>2207</td>
</tr>
<tr>
<td>2004/05</td>
<td>6495</td>
<td>2952</td>
<td>2570</td>
</tr>
<tr>
<td>2005/06</td>
<td>(6787)*</td>
<td>(3061)*</td>
<td>(2694)*</td>
</tr>
</tbody>
</table>

NOTES TO TABLE 2

\textsuperscript{1} \textit{COLUMN DEFINITIONS:} Central Government Revenues = Total revenues per capita raised by central government taxation; \( g_F \) = Total intergovernmental transfers per capita paid to the province(s), averaged over all provinces (National Average), for the Western Cape, and for all other provinces excluding the Western Cape (Ave. Other Provinces); \( s(q) \) = 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 (\( s_c(q) \)) and the average for all other provinces (\( s_m(q) \)); and \( b \) = “basic grant” per capita to fund all other provincial services and is defined as \( b = g_F - s(q) \).

TABLE 3: POLITICAL ECONOMY OF SOUTH AFRICA

DEMOGRAPHICS and INCOMES

\[ N_0 = 9.6 \text{ Million Elite Adults} \]
\[ M = 25 \text{ Million Majority Adults} \]
\[ Y = 86,000 \text{ (Real 2000) Rand/Elite Adult} \]
\[ W = 9,700 \text{ (Real 2000) Rand/Majority Adult} \]
\[ \beta = .00015 \]

ELITE CAPTURE:

\[ \varphi^L = .20 \text{ (Rate of Capture per Rand of Basic Grant)} \]
\[ \varphi^H = .85 \text{ (Rate of Capture per Rand of Basic Grant)} \]
\[ \rho = 1720 \text{ (Real 2000) Rand/Elite Adult (Protest Cost if } \varphi \geq \varphi^H) \]

REDISTRIBUTIVE SERVICE TECHNOLOGY

\[ a_e = 17 \text{ (Years of Training; Elite Public Employee)} \]
\[ a_m = 14 \text{ (Years of Training; Trained Majority Public Employee)} \]
\[ a_u = 10 \text{ (Years of Training; Untrained Majority Public Employee)} \]
\[ m = .85 \text{ (Share of Majority Residents Serviced by Trained Majority)} \]

REDISTRIBUTIVE SERVICE COSTS

\[ S = 80,000 \text{ (Real 2000) Rand/Public Employee (Average Uniform Salary)} \]
\[ s_e(q) = S/a_e = (80,000/17)q = 4706 \cdot q \text{ (Real 2000) Rand/Majority Adult} \]
\[ s_m(q) = S/a_m = (80,000/14)q = 5714 \cdot q \text{ (Real 2000) Rand/Majority Adult} \]
\[ s_u(q) = S/a_u = (80,000/10)q = 8000 \cdot q \text{ (Real 2000) Rand/Majority Adult} \]
\[ s_t(q) = m \cdot s_m(q) + (1 - m) \cdot s_u(q) = 6057 \cdot q \text{ (Real 2000) Rand/Majority Adult} \]

SPECIAL INTEREST PAYMENTS

\[ Z = 600 \text{ Million (Real 2000) Rand (Compensation for Lost Homeland Payments)} \]

DISCOUNT FACTOR

\[ \delta = .97 \]

UNITARY DEMOCRACY’S FISCAL POLICIES

\[ \tau_U = 32,000 \text{ (Real 2000) Rand/Elite Adult (Revenue Maximizing Tax)} \]
\[ N(\tau_U) = .5 \cdot N_0 = 4.8 \text{ Million Elite Adults (Elite Adults When } \tau = \tau_U) \]
\[ g_U = \left[ \tau_U \cdot N(\tau_U) - Z \right]/M = 6120 \text{ (Real 2000) Rand/Majority Adult (Maximal Grant)} \]

† Source: Data Appendix available upon request.
**TABLE 4: RSA INTERGOVERNMENTAL TRANSFERS: PREDICTED†**
(Transfers per Majority Adult Resident, Real 2000 Rand; *Transfers per Capita, Real 2000 Rand††*)

<table>
<thead>
<tr>
<th>VALUE</th>
<th>MAJORITY PUNISHMENT</th>
<th>( g_{\text{min}}(U) )</th>
<th>( g_{\text{max}}(U) = g_F )</th>
<th>( s_e(q^{\text{FFC}}) )</th>
<th>( b_e )</th>
<th>( s_m(q^{\text{FFC}}) )</th>
<th>( b_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>“BEST” ESTIMATES*</td>
<td>Unitary</td>
<td>6035; 3319</td>
<td>6119; 3365</td>
<td>2494; 1372</td>
<td>3625; 1993</td>
<td>3028; 1665</td>
<td>3091; 1700</td>
</tr>
</tbody>
</table>

† COLUMN DEFINITIONS: MAJORITY PUNISHMENT is the preferred long-run punishment strategy of the majority if the elite chooses to adopt the high \((\varphi^*)\) capture strategy; \( g_{\text{min}} \) = minimal intergovernmental transfer per majority adult resident \((\text{per capita})\) that will be accepted by the majority under democratic federalism, where unitary democracy is the majority’s credible punishment option; \( g_{\text{max}} \) = Maximal intergovernmental transfer per majority adult resident \((\text{per capita})\) that will be paid by the elite under democratic federalism, where unitary democracy is the majority’s credible punishment option; \( s_e(q^{\text{FFC}}) \) = Intergovernmental transfer per majority adult resident \((\text{per capita})\) required to provide the constitutionally assigned level of public services per majority resident in the elite province; \( b_e \) = Basic grant per majority adult resident \((\text{per capita})\) in the elite province to fund other provincial services, defined as \( b_e = g_{\text{max}} - s_e(q^{\text{FFC}}) \); \( s_m(q^{\text{FFC}}) \) = Intergovernmental transfer per majority adult resident \((\text{per capita})\) required to provide the assigned level of public services per majority resident in majority provinces; \( b_m \) = Basic grant per majority adult resident \((\text{per capita})\) in majority provinces to fund other provincial services, defined as \( b_m = g_{\text{max}} - s_m(q^{\text{FFC}}) \).

†† TRANSFERS PER CAPITA (Italics): The calibration model estimates intergovernmental transfers per majority adult resident. For purposes of comparisons to Table 2, intergovernmental transfers per capita are computed as (Transfers per majority adult)-(Majority Adults/Population), where (Majority Adults/Population) = (25 m./45.5m) = .55 for 2002/2003 population.

* CALIBRATION “BEST” ESTIMATES : Demographic and economic parameters as specified in Table 3. Constitutional Border Constraint: \( \mu = .184 \); Constitutional Assignment Constraint: \( q^{\text{FFC}} = .53 \).
TABLE 5: THE ECONOMIC BURDEN OF APARTHEID

ANNUAL MILITARY, POLICE AND JUSTICE EXPENDITURES: 1977-1993

33.7 Billion (2000) Rand/Year; 3,510 Rand/Elite Adult per Year

ANNUAL HOMELAND PAYMENTS: 1985-1993

26.3 Billion (2000) Rand/Year; 2,740 Rand/Elite Adult per Year

IMPACT OF SANCTIONS AND LABOR UNREST ON GROWTH: 1950-2000

<table>
<thead>
<tr>
<th>INDEPENDENT VARIABLES</th>
<th>GROWTH RATE (1)</th>
<th>GROWTH RATE (2)</th>
<th>GROWTH RATE (3)</th>
<th>OPEN (4)</th>
<th>INV (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.431 (.355)*</td>
<td>-6.938 (1.539)*</td>
<td>-6.933 (1.556)*</td>
<td>52.038 (.531)*</td>
<td>14.777 (.487)*</td>
</tr>
<tr>
<td>OPEN</td>
<td>-</td>
<td>.137 (.037)*</td>
<td>.140 (.040)*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>INV</td>
<td>-</td>
<td>.147 (.069)*</td>
<td>.137 (.085)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DEM</td>
<td>-1.765 (.818)*</td>
<td>-</td>
<td>-.188 (.941)</td>
<td>-3.909 (1.241)*</td>
<td>-7.142 (1.139)*</td>
</tr>
<tr>
<td>SANC</td>
<td>-1.554 (.659)*</td>
<td>-</td>
<td>-</td>
<td>-12.198 (.998)*</td>
<td>-203 (.916)</td>
</tr>
<tr>
<td>COSATU</td>
<td>-2.099 (.772)*</td>
<td>-</td>
<td>-</td>
<td>-5.466 (1.176)*</td>
<td>-6.366 (1.079)*</td>
</tr>
<tr>
<td>R²(Adj)</td>
<td>.345</td>
<td>.345</td>
<td>.344</td>
<td>.869</td>
<td>.600</td>
</tr>
</tbody>
</table>

* Significant at the 5% level; standard errors within parentheses.

58 Source: South African Department of Information, Perskor, South Africa, Various Years.


60 Dependent variable is South Africa’s annual real rate of growth of GDP per capita. Independent variables include: OPEN (exports plus imports as a percentage of GDP), INV (gross investment as a share of GDP), DEM (1 for the years 1994-2000; 0 otherwise), SANC (1 for the years 1976-1993; 0 otherwise), and COSATU (1 for the years 1985-1993; 0 otherwise). Source: GROWTH RATE, OPEN, and INV are from the Penn World Tables, 6.1 and correspond to the PWT variables GRGDPCH, OPENK, and KI, respectively. DEM, SANC, and COSATU are defined in the text.
### TABLE 6: NET ECONOMIC GAINS FROM THE DEMOCRATIC TRANSITION†

<table>
<thead>
<tr>
<th></th>
<th>YEARS UNTIL $\Delta$NPV ≥ 0</th>
<th>$\Delta$NPV 70 YEAR HORIZON</th>
<th>DEMOCRACY’S INTERNAL RATE OF RETURN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAJORITY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unitary</td>
<td>0</td>
<td>367,888 Rand</td>
<td>∞</td>
</tr>
<tr>
<td>Federal</td>
<td>0</td>
<td>375,679 Rand</td>
<td>∞</td>
</tr>
<tr>
<td>ELITE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unitary</td>
<td>22 Years</td>
<td>2,179,985 Rand</td>
<td>.115</td>
</tr>
<tr>
<td>Federal</td>
<td>21 Years</td>
<td>2,197,120 Rand</td>
<td>.118</td>
</tr>
</tbody>
</table>

†COLUMN DEFINITIONS: Years until $\Delta$NPV ≥ 0 are the number of years until the net present value gains in after-tax incomes in moving from apartheid to each form of democracy just exceeds zero. $\Delta$NPV for a 70 year horizon is the net present value gains in after-tax income in moving from apartheid to each form of democracy. Democracy’s Internal Rate of Return is defined by that discount rate where net present value of after-tax incomes under democracy are just equal to the net present value of after-tax incomes under apartheid so that $\Delta$NPV for the full 70 year horizon is just equal to 0.

SPECIFICATION OF $\Delta$NPV: For majority residents, $\Delta$NPV is computed as the discounted present value of the difference between after-transfer incomes under either unitary or federal democracies and after-transfer incomes under apartheid. Annual incomes before transfers for majority residents are: $W_{ Dt} = W_0(1 + g_D)^t$ under unitary or federal democracy and $W_{ At} = W_0(1 + g_A)^t$ under apartheid, where $W_0 = 9,700$, $g_D$ is the growth rate under democracy specified as an incremental increase from .0067 to .025 over the ten years from 1996-2005 and as .012 for all years beyond 2006 and $g_A = - .012$; see text. The estimates of NPV for majority residents includes the income equivalent value of receiving redistributive services in democracy, specified as $q_{FFC} = .53$ as specified by the FFC for years 1996-2000 (the transition or “Mandela” years) and then $q_{U} = .62$ under unitary democracy and $q_{ANC} = .65$ under federal democracy for all (ANC) years thereafter. Annual transfers are set equal to actual transfers (Table 2) for the first four years of democracy (Mandela years) and then $g_{U} = 6120$ and $g_{f} = 6119$ thereafter. For elite residents, $\Delta$NPV is computed as the discounted present value of the difference between after-tax incomes under either unitary or federal democracies and after-tax incomes under apartheid. Annual incomes before taxation for elite residents are: $Y_{ Dt} = Y_0(1 + g_D)^t$ under unitary or federal democracy and $Y_{ At} = Y_0(1 + g_A)^t$ under apartheid, where $Y_0 = 86,000$ Rand and $g_D$ and $g_A$ are specified as above. Annual (real) taxes include taxes for homeland payments under apartheid, for redistributive transfers under democracy, and for military, police, and justice expenditures as required by both regimes (Table 5). A discount rate of .03 ($\delta = .97$) is used for the computation of $\Delta$NPV.
FIGURES

FIGURE 1: Feasible Federalism for the $q$-Regime

FIGURE 2: Surplus with Administrative Federalism

FIGURE 3: Feasible Federalism for the $q^*$-Regime
Figure 1: Feasible Federalism for the $q$-Regime

- $q^\text{min}(\mu)$
- $\mu^*) = 0.184$
- $\mu^\text{min}(q^{\text{min}}) = 0.10$
- $q^\text{min}(\mu) = 0.23$
- $q^\text{min} = 0.53$
- $q^\text{max} = 0.74$
- $\mu^\text{max} = 0.192$
Figure 2: Surplus with Administrative Federalism

\[ p(m) = 5363 \]

\[ p(\mu) = 6057 \]

\[ 0.58 = q^*(\lambda) \]

\[ q^*(\mu, \lambda) = 0.65 \]
Figure 3: Feasible Federalism for the $q^*$-Regime