FEDERALISM AND SOUTH AFRICA'S DEMOCRATIC BARGAIN:

THE ZUMA CHALLENGE

by

Robert P. Inman and Daniel Rubinfeld*

ABSTRACT

South Africa's transition from apartheid to democracy stands as one of the past century's most important political events. The major hurdle to the transition was for the poor majority ANC to provide a credible promise not to exploit the full economic resources of the then ruling economic elite. The new constitution adopted a form of federal governance that has the potential to provide such protections by specifying an annual policy game where the new majority and the minority elite each control one policy instrument of importance to the other. Provided the majority is sufficiently patient and not "too demanding" in their preferences for redistribution the game has a stable equilibrium with less than maximal redistributive taxation. Our analysis makes these restrictions on preferences precise. The new, more radical ANC and the Zuma presidency challenge this equilibrium.

* Inman is the Richard Mellon Professor, Finance and Economics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104. Rubinfeld is the Robert L. Bridges Professor, Law and Economics, The Law School, University of California, Berkeley, 94720. The analysis has benefitted from presentations to our colleagues at Berkeley, Cornell, Michigan, NYU, Penn, Stanford, and Wesleyan. Particular thanks are due to Grant Long and Jon Stott for extraordinary research assistance.

Federalism and South Africa's Democratic Bargain: The Zuma Challenge

> *First, peaceful democratic change in a plural society requires a consociational pattern of decision-making. Secondly, among the many ways of implementing consociational principles, federalism is a particularly promising method and deserve special attention.* Arend Lijphart, "Consociation and Federation: Conceptual and Empirical Links," *Canadian Journal of Political Science*, Vol 12 (September, 1979), pp. 514-515.

> *Look at the prominent people around him. If some of the things they say come to pass then we will be facing a calamity such as 'We need free education' (which Mr. Zuma called for last week). How are you going to pay for it without nationalizing the mines?* Saki Macozoma, a leading member of the African National Congress, commenting on the presidential candidacy of Jacob Zuma (*Financial Times*, December 17, 2007, p.).

South Africa's transition from apartheid to a truly multi-racial democracy stands as one of the significant political events of the last century. The transition was peacefully negotiated, the democratic bargain is still holding, and despite still high rates of unemployment, the average South African resident, both black and white, is economically better off today than they were under the last years of apartheid.[1] Though peaceful, the constitutional negotiations were far from harmonious. It took over four years from the date of Nelson Mandela's release from Robben Island on February 11, 1990 until April, 1994 before even an outline of a democratic constitution was accepted by three relevant parties to the negotiations, the National Party (NP) representing the once ruling whites, the African National Congress (ANC) representing the majority of blacks and Asian South Africans, and Inkatha Freedom Party (Inkatha) representing the rural blacks of

[1] Real incomes per capita have grown by 2% peer annum for all percentiles of the income distribution from 1993 to 2007. (RSA, President's Office, *Development Indicators, 2008*, p. 23). The national rate of poverty has declined from 31 percent to 23 percent over this same period (p. 26). Rates of adult disability and infant mortality are both down (p. 38). Housing quality has improved significantly (pp. 31-34). Class sizes are smaller, school enrollment is up, and the national rate of literacy has increased (p. 49). The two adverse developments since the end of apartheid are the significant increase in the incidence of HIV and a resulting fall in life expectancy (p. 42) and the increase in the national crime rate (pp. 60-61). Crime rates have fallen since 2003, however, and today's rate of just over 5,000 crimes per 100,000 residents is comparable to the rates in most large U.S. cities.

the historic Zulu nation.

This initial agreement, known as the Interim Constitution, outlined the broad parameters of the new democracy. First, it detailed the rules for the election of a National Assembly from which would be chosen the President of the new republic; rules crucial to assure the increasingly impatient black majority they would have an equal voice in a truly democratic new South Africa. Second, it created nine provincial governments each with a separately elected legislature and premier (or governor); governments whose boundaries were explicitly negotiated to assure the white (NP) and black (Inkatha) political minorities control over public resources and policies in at least one province each. These initial negotiations established the new republic as a federal democracy, but beyond that the Interim Constitution was strikingly silent. It took another two years of full time negotiations before a final constitution was presented to the National Assembly, on October 11, 1996, for its unanimous approval.

The final constitution established three important principles for the governance of the new federal democracy. First, it accepted the geographical boundaries of the nine provinces, and thus their anticipated political control, as specified by the original Interim Constitution. Second, provinces were given responsibility for the provision of K-12 education, health services, and housing, and for the administration of transfers to the poor and elderly. Third, the national government was required to share national tax revenues with the provinces to finance assigned provincial services. Following the lead of Lijphart (1979) we argue below in Sections II and III that these constitutional provisions for provincial borders, provincial service responsibilities, and provincial financing were *necessary* for a viable federal democracy in the new South Africa.

Whether these federal institutions will prove *sufficient* is the Zuma Challenge. In the recent elections on April 23, 2009, the voters of South Africa chose the ANC as the majority party for the national parliament, allocating 65 percent of the Parliament's seats to the ANC. The ANC in turn has chosen Jacob Zuma as the country's next president. Mr. Zuma is a controversial choice. At the December, 2007 ANC

convention Mr. Zuma's supporters ousted then President Thabo Mbeki as chairman of the ANC. Mr. Zuma's core supporters in the ANC are the trade unions and the communist party, while President Mbeki's support came from the more moderate wing of the ANC, those committed to continuing President Mandela's policy agenda of a competitive markets, constrained taxation of business, economic stability, and long-run growth. In anticipation of Mr. Zuma's likely redistributive agenda, the moderate wing of the ANC left the party to form an new party known as the Congress of the People (COPE).

COPE finished third in the national election, behind the other moderate party, the Democratic Alliance. Important to our argument that federal governance can be a moderating force on redistributive fiscal policies, however, is the fact that together these two parties have won majority control over the economically important province of the Western Cape. Section IV asks whether this political control of a key province will be sufficient to check the likely strong redistributive preferences of Mr. Zuma's core supporters in the new, more radical, ANC. This is the Zuma Challenge to South Africa's original democratic bargain. Section V concludes with possible general lessons from our analysis.

II. Federal Institutions as a Check on Majority Redistribution

A. ***Background***: At the time of the initial constitutional negotiations, Robert Mugabe's Zimbabwe provided a strong reminder to the leadership of both the NP and the ANC of the risks of simple majority rule in an economy marked by wide disparities in incomes and assets. Even with fair elections, the temptation for the poor majority, or their elected representatives, to expropriate upper income assets might prove irresistible. It was clear to both the ANC and the NP that a peaceful transition would require a credible commitment to protect elite incomes.[2] To this end, the NP and the ANC compromised on an interim

[2] Waldmeir (1997), p. 157 quotes Nelson Mandela in his initial address on reconciliation as saying: "(T)he ANC is very much concerned to address the question of the concerns of whites. . . They insist on structural guarantees to ensure that . . . majority rule does not result in the domination of whites by blacks. We understand that fear. The whites are our fellow South Africans. We want them to feel safe."

constitution establishing a federal democracy with a simple majority rule National Assembly, a President elected by the Assembly, and nine provincial governments with boundaries drawn to ensure NP control of at least one, ideally two, provinces.[3] Left unspecified was the hard matter of policy assignment between the national and provincial tiers of government. That difficult task was delegated by an Interim Constitution to a panel of experts to be appointed by the new President and to be known as the Financial and Fiscal Commission (FFC). The Commission was equally balanced in its representation between the ANC and the NP, and each member of the Commission was an expert in at least one area of government policy-making: finance, administration, or accounting.

The Commission began by accepting without debate the provincial boundaries, and their likely voting outcomes, as proposed in the Interim Constitution (final Constitution, Chapter 6, Section 103). On the crucial matter of who should decide taxation and redistribution policies, the FFC gave control over all important taxes, in particular income and profits taxation and the VAT, to the central government (final Constitution, Chapter 13, Section 214), but then assigned control for the provision of redistributive services of K-12 education, health care, and housing and the payment of poor and elderly transfers to the provinces (final Constitution, Schedule 4).[4] Finally, redistributive services were to be funded by a constitutionally required sharing with the provinces of national tax revenues (final Constitution, Section 227). These recommendations were unanimously approved within the FFC and incorporated directly into the unanimously approved final constitution.

---

[3] Waldmeir (1997), Chapters 10-13 provides an valuable overview of the transition negotiations. Differences over the structure of the federal contract are summarized on pp. 193-197; 241-244. For a summary of how the number and boundaries of the new provinces were decided, see Muthien and Khosa (1998). These boundaries negotiated for the Interim Constitution were accepted directly as part of the final constitution (final Constitution, Chapter 6, Section 103). The NP had hoped to win political control of the rural Northern Cape and urban Western Cape, but they misjudged voter turnout by alienated conservative white farmers and farm workers. As a consequence the ANC won control of the Northern Cape. The NP and their allied parties were successful in the Western Cape.

[4] The final constitution does allow provinces to have their own taxation administered as a surcharge on the national income tax, but such powers must be first approved by legislation from the National Assembly (final Constitution, Chapter 13, Section 228). To date such provincial taxing powers have not been approved by the National Assembly.

The end result was to create an annual redistribution policy game in which a majority ANC controlled central government and elite controlled province(s) each sets one redistributive policy instrument of importance to the other, taxes controlled by the ANC and redistributive spending by the elite. Sections B and C specifies the details of this policy game and its likely outcomes. Under well defined conditions – low-cost elite providers of redistributive services and a sufficiently restrained and patient majority – this annual policy game when played repeatedly can check the redistributive incentives of the national poor majority.

B. ***The Annual Redistribution Game***: In the annual redistribution game the central government decides an aggregate redistributive tax per elite resident (τ) with the proceeds of this tax then allocated equally to the nine provincial governments as a redistributive grant per lower income resident (g). The central government also sets national standards for provincial spending on redistributive service inputs (q). We assume that the central government can monitor the provision of these inputs. In South Africa, the constitutionally assigned services inputs are for K-12 education, primary health care, and child and old-age income transfers. The average cost per poor resident of providing these service inputs will be s(q). After satisfying the required service standard q, provinces are free to spend the remainder (b = g - s(q)) on services of their own choosing.

The Budget Constraint: All fiscal policies are decided subject to the aggregate redistributive budget constraint:

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

$g(\tau)$ is average redistributive provincial grant per majority resident, $\tau \bullet N(\tau)$ is the central government's redistributive revenues from the imposition of redistributive tax rate per elite resident τ on the taxpaying minority $N(\tau)$, Z are payments to (> 0) or revenues from (< 0) parties outside the minority-majority fiscal agreement, and M is the size of the majority population. In the early years of the South African democracy Z (> 0) was paid as "out-of-formula" grants to the province of KwaZulu-Natal to ensure the original participation of Chief Buthelezi and Inkatha in the new democracy.

The taxpaying minority is free to leave the country, or to adopt tax avoidance strategies, as the redistributive tax rate increases; thus $N(\tau)$ may declines as τ rises. Emigration is a possibility but it has not proven to be significant quantitatively, popular stories aside. Tax avoidance is the primary means by which the elite reduces its tax payments. There is a revenue hill (aka "Laffer curve") for redistributive taxation. Revenues initially increase as τ rises, reach a maximum at τ_U, and then decline. Majority dominated unitary governments always select the maximum rate. Given the revenue potential of national redistributive taxation, we want to know if political institutions, and in particular democratic federalism, might allow an equilibrium redistributive tax rate, denoted τ_F, that is less than τ_U.

The Cost of Providing Redistributive Services: The primary service inputs used by the provinces to provide education and health care are teachers, doctors, and nurses. Income transfers to children and the elderly require cash and an honest public administration. We specify three classes of public employees: white minority providers with a_e years of training, trained majority providers with a_m years of training, and untrained majority providers with a_u years of training, where $a_e > a_m > a_u$. Better trained public employees are more productive and more honest. All public employees are assumed to be paid a common civil service wage, *S*, which is only imperfectly related to their individual productivity.[5] Therefore, more productive workers will therefore be less expensive when providing any required service input bundle, q. The cost per majority resident of providing q is specified as s(q), with $s_e(q) < s_m(q) < s_u(q)$, using highly trained minority, majority trained, and majority untrained providers, respectively. The minority elite's control over the low cost technology for providing valued redistributive services will be crucial for its ability to check redistributive taxation as a whole. The majority needs the elite and therefore has an incentive to retain their participation in the provision of redistributive public services.

[5] Having wages fully independent of employee productivity is not essential to our arguments and analysis, but an imperfect matching of wages to productivity is important. As a consequence of the decision to not discriminate by race, South Africa does have a common wage structure for positions in the civil service, without careful regard for background or training.

Redistributive Fiscal Effort: We assume the central government can successfully monitor the inputs allocated by the provinces to redistributive services, so once the standard for q is set by the central government, the provinces comply. What the central government cannot monitor, however, is how any redistributive transfers not required by q are spent. Elite controlled provinces might choose to spend their remaining redistribution grant, denoted as $b = g(\tau) - s(q)$, on services consumed by their elite residents. The share so allocated (φ) measures a *lack of* redistributive effort by the province. In the public finance literature, φ is often called the "rate of fungibility" or "provincial capture" of targeted grants.[6] The majority run central government would like minimal provincial shirking of redistributive effort with $\varphi = 0$, and the majority run provinces always comply. Elite controlled provinces, however, will shirk their redistributive responsibilities and seek to push φ as high as possible.

We assume there is some value of fiscal effort φ_L, perhaps very small, that the elite province can always allocate to elite residents without detection or penalty, but there is an upper limit φ_H as well: $0 < \varphi_L \leq \varphi \leq \varphi_H \leq 1$. The upper limit to capture defines maximum shirking and is set by the possibility that majority residents of the elite run province will leave the province and choose to relocate to a majority run province where there is no shirking ($\varphi = 0$). The upper limit is set to equalize the welfare of a typical poor resident in the elite and majority run provinces. As poor residents leave the elite run province, redistributive grants decline proportionally and the leaders of the elite province stop capture. Finally, when the rate of capture exceeds its lower bound and services to lower income residents are noticeably reduced, poor residents within the elite province impose a "protest" penalty on each elite resident of ρ Rands per elite resident. These costs come as the consequence of spontaneous marches or perhaps riots or from formally organized strikes. The costs to elite residents of such protests discourage redistributive "shirking" via high capture.

Resident Economic Welfare: The private economy together with our specification of the

[6] The current empirical literature, both for developed and less developed countries, estimate the rate of capture of central government grants by the provincial governments for their own use to range from $.30 \leq \varphi \leq 1.0$; see Inman (2009).

government's budget constraint, the costs of public services, and the fiscal choices of the central (τ and q) and provincial (φ) governments will define the economic welfare of the poor majority and the rich minority. We do so for two forms of governance, *federalism* as now specified by the South African constitution and a "default" regime we call *de facto unitary* governance. The constitution does allow the majority run central government to take over the direct provision of redistributive services at any time. If so, we assume the well-trained minority teachers, nurses, doctors and civil servants will reduce their effort, or more likely, exit the public sector for careers in the private economy.[7] We specify majority and minority resident welfare under both forms of governance.

Under *de facto unitary governance*, the average majority resident receives their private sector income (W) plus the utility value of redistributive services q – denoted $\lambda \bullet \upsilon(q)$, where $\upsilon'(q) > 0$ and $\upsilon''(q) < 0$ – plus any redistributive revenues left over after providing q:

$$\omega(U) = W + \lambda \bullet \upsilon(q) + [g(\tau_U) - s_U(q)],$$

where $[g(\tau_U) - s_U(q)]$ is the residual revenues available for the majority resident after providing q at a cost per majority resident of $s_U(q)$. With unitary governance, the majority goes to the top of the revenue hill to by selecting τ_U to maximize redistributive revenues. The high productivity, low cost minority providers do not participate in the public sector under unitary governance, or are less productive if they do so. As a consequence, the average costs of providing public services under unitary governance will be greater than costs under federalism. Welfare for the average rich minority resident under unitary governance is simply their income ($Y > W$) minus their tax payment, τ_U:

$$y(U) = Y - \tau_U.$$

Middle and upper income residents do receive services from their province but not from the redistributive

[7] This assumption follows from the theoretical work of Akerlof and Kranton (2004) and the extensive empirical work on the adverse consequences for organizational efficiency of racial and educational diversity between managers and workers; see Williams and O'Reilly (1998). In our analysis we assume $s_U(q) = m \cdot s_m(q) + (1 - m) \cdot s_u(q)$, where m is the share of public employees under unitary governance who are formally trained while (1-m) is the residual number of public employees hired with only limited training.

budget. Only the qualified majority residents receive q.

Under *federalism*, a fraction μ of the poor majority residents will reside in elite-run provinces while $(1 - \mu)$ of the majority will live in the majority-run province(s). For majority residents in a majority provinces, welfare will be:

$$\omega_m(\tau_F, \varphi) = W + \lambda \bullet \upsilon(q) + [g(\tau_F) - s_m(q)],$$

where τ_F is the equilibrium tax rate chosen by the central government under federal governance and $\tau_F \leq \tau_U$. We will be looking for a stable equilibrium to fiscal politics where $\tau_F < \tau_U$. For majority residents in the elite-controlled provinces, their welfare will be:

$$\omega_e(\tau_F, \varphi) = W + \lambda \bullet \upsilon(q) + (1 - \varphi) \bullet [g(\tau_F) - s_e(q)],$$

where $(1 - \varphi)$ is the share of the residual redistributive grant that *has not been* captured by the elite for elite services or tax relief. The majority party is assumed to be interested in the welfare of the average majority resident for the entire country equal to:

$$\omega(\tau_F, \varphi) = \mu\omega_e(\tau_F, \varphi) + (1 - \mu)\omega_m(\tau_F, \varphi),$$

or after substitution as:

$$\omega(\tau_F, \varphi) = W + g(\tau_F) + [\lambda \bullet \upsilon(q) - s_F(q)] - \mu \bullet \varphi \bullet [g(\tau_F) - s_e(q)],$$

where $W + g(\tau_F)$ is wage income plus the aggregate redistributive transfer, $[\lambda \bullet \upsilon(q) - s_F(q)]$ are the benefits from the provision of the redistributive public goods net of the average costs of providing those goods using provinces ($s_F(q) = \mu \bullet s_e(q) + (1 - \mu) \bullet s_m(q)$), and $\mu \bullet \varphi \bullet [g(\tau_F) - s_e(q)]$ is the transfer income lost because of elite capture in the elite province averaged over all of the poor majority.

Rich minority residents are assumed to live only in the elite-run province. If the rich minority adopts low capture so that shirking goes unnoticed, they receive an annual income (welfare) of:

$$y(\tau_F, \varphi_L) = Y - \tau_F + \varphi_L \bullet [g(\tau_F) - s_e(q)] \bullet [\mu \bullet M/N(\tau_F)],$$

where the elite resident pays τ_F in the federal regime and captures back φ_L of the residual redistributive grant paid to the poor residents living in their provinces ($\mu \bullet M$), shared equally among the all the rich minority

residents (N) in the elite province. If the rich minority chooses maximal shirking, however, they receive φ_H of the residual redistributive grant but bear the penalty of ρ Rand in protest costs. The protest costs imposed by the majority on the minority are pure economic waste. Together,

$$y(\tau_F, \varphi_H) = Y - \tau_F - \rho + \varphi_H \bullet [g(\tau_F) - s_e(q)] \bullet [\mu \bullet M/N(\tau_F)].$$

For both the poor majority and rich minority in the new democracy, welfare depends upon the governance regime and, for the federal regime, the choice of the redistributive tax rate by the majority controlled central government and the choice of capture by the minority controlled province. Under majority rule unitary governance, the elite cannot prevent the choice of τ_U. Under democratic federalism they can, *if* their constitutional control of redistributive spending allows for punishments of maximal redistributive taxation. The punishment strategy is for the elite province to adopt φ_H whenever the central government selects τ_U. When is such a strategy credible?

Credible Elite Punishments: For the φ_H to be a credible elite punishment, two conditions must hold. The first, which we call the *Assignment Constraint*, ensures the majority cares enough about assigned provincial services and the elite is sufficiently efficient in the provision of those services that the majority will not by-pass provincial governments and simply choose to provide all services through a de facto unitary government. The second condition – the *Border Constraint* – sets an upper and lower bound on the number of majority residents who live in the elite province. When the Assignment and Border constraints hold, the elite's choice of φ_H will be politically feasible and economically credible.

The Assignment constraint requires the majority to prefer federal over de facto unitary governance even if the elite adopts the high capture strategy – that is, $\omega(\tau_U, \varphi) > \omega(U)$. For this to be true, the level of assigned redistributive services must exceed some lower bound, $q > q^{min}(\mu)$, where q^{min} increases as μ increases. Since unitary governance prevents capture of redistributive transfers by the elite province, and the level of capture increases the more majority residents (μ) who reside in the elite province, a switch from federal to unitary governance becomes more attractive to the majority as μ increases. The only way to favor

the continued use of provincial governments then is to increase the importance of redistributive services, services where elite providers in the elite province have a significant cost advantage. Thus as μ increases, so too must q^{min}; see Figure 1.

The level of centrally required redistributive services cannot be too large, however. Assigned redistributive services cannot exceed $q^{max}(\mu)$ defined by that value of q where the economic benefits of high capture to the elite – $\varphi_H \bullet [g(\tau_F) - s_e(q)] \bullet [\mu \bullet M/N(\tau_F)]$ – just fail to compensate for the penalty ρ of using high capture. If so, the elite province cannot credibly threaten to choose φ_H. As μ increases and more majority residents are assigned to the elite province, q^{max} can increase too and still permit the elite to credibly threaten to adopt high capture; see Figure 1. For a credible elite punishment, the federal constitution must identify redistributive services where the elite has a clear cost advantage in service provision and then assign a level of those services so that $q^{max}(\mu) \geq q > q^{min}(\mu)$ is satisfied. Figure 1 shows our estimates of $q^{max}(\mu)$ and $q^{min}(\mu)$ for the South African political economy.

The Border constraint sets an upper and lower bound on the share of poor majority residents who live in elite controlled provinces. The upper bound, μ^{max}, ensures that the elite population will still be a political majority in their province, even if the central government selects τ_U – that is, $N(\tau_U) \geq \mu \bullet M$ or $N(\tau_U)/M = \mu^{max} \geq \mu$. The lower bound, $\mu^{min}(q)$ ensures that when the majority does choose τ_U, the elite will adopt the φ_H punishment. This requires $y(\tau_U, \varphi_H) > y(\tau_U, \varphi_L)$. For this to be true, there must be enough majority residents residing in the elite province, and thus a large enough redistributive transfer, so that rewards from high capture compensate for expected protest costs. Redistributive transfer available for capture – $[g(\tau_U) - s_e(q)] \bullet [\mu \bullet M/N(\tau_U)]$ – declines as q rises, however. If the central government increases its level of assigned redistributive services, then provincial borders will need to be re-drawn to allocate a larger share μ of the majority to the elite province. Thus $\mu^{min}(q)$ depends on q, where μ^{min} increases as q rises; see Figure 1. The Border constraint requires $\mu^{max} \geq \mu > \mu^{min}(q)$. Figure 1 shows our estimates of μ^{max} and $\mu^{min}(q)$ for the South African political economy.

Importantly, there is no guarantee that the Assignment and Border constraints will be jointly satisfied by a country's political economy. It may be that the minimal level of assigned redistributive services required to protect elite provinces as providers, $q > q^{min}$, so reduces the size of redistributive transfers available for capture, and thus so increases $\mu^{min}(q)$, that the Border constraint cannot then be satisfied – that is, $\mu^{min}(q) > \mu^{max}$. If so, there is no credible elite punishment, and democratic federalism will fail as a check on majority expropriation.

The Annual Redistribution Game: For those political economies where the Assignment and Border constraints can both be met, Table 1 shows the annual welfare pay-offs for the poor majority and the elite minority. Each party to the federal bargain can play either of two strategies. The poor majority controls the central government and can choose a "cooperative" strategy and set $\tau_F < \tau_U$ or a "defect" strategy and set $\tau_F = \tau_U$. Under the defect strategy the central government effectively sets all fiscal policies and simply uses the provinces as providers of redistributive services; a regime we call *de facto unitary* governance.

The rich minority controls the fiscal policies of at least one province, where they can adopt a "cooperative" low shirk, low capture (φ_L) strategy or a "defect" high shirk, high capture (φ_H) strategy. The economically efficient outcome is when both parties choose their cooperative strategies; a regime we call *democratic federalism*. The majority's choice of the low tax rate strategy increases aggregate national income by increasing the participation of the talented minority in the taxable private economy; the elite's choice of a low shirking, low capture strategy avoids wasteful protest costs by the majority. The inefficient outcome is when both parties defect from democratic federalism to de facto unitary governance by choosing high tax rates and high shirking.

Unfortunately when the annual redistribution game is played just once, Table 1 reveals the majority always prefers to defect. If the elite province cooperates and plays φ^L, the majority is unambiguously better off by choosing τ_U rather than a smaller τ_F; the high tax rate maximizes their net transfer income as $g(\tau_U)[1 - \mu \bullet \varphi_L] > g(\tau_F)[1 - \mu \bullet \varphi_L]$. Knowing that the majority will defect, the elite's best strategy is to defect to

maximal shirking and adopt high capture, a result which follows from our Border Constraint requiring $y(\tau_U, \varphi_H) > y(\tau_U, \varphi_L)$. The inefficient fiscal outcome results. Clearly a federal constitution alone, even one protecting elite provinces by satisfying our Assignment and Border constraints, is not sufficient to protect elite incomes from majority expropriation. Something more is needed.

C. Sustaining Democratic Federalism: The central feature of Table 1's redistribution game is the temptation for the majority to defect from the cooperative federal allocation to maximal redistributive taxation. When they do, then a socially inefficient outcome results. How might we discourage defection? As in all games of this form, of which the usual prisoners' dilemma is most common, the key is to play the game repeatedly and to give the cooperating elite province a means to impose a sufficiently large penalty on a majority central government when they defect.[8] One possible penalty is for the elite province to adopt the "grim trigger strategy" where it plays low capture as long as the majority central government has adopted a less than fully exploitative tax rate, but if the central government selects τ_U, then the elite province adopts high capture forever. For this redistribution game, this grim trigger strategy is the toughest penalty the elite can impose on the majority; if this penalty cannot discourage maximal taxation, then for this game, nothing will.[9]

In Inman and Rubinfeld (2008), we prove the following Proposition for the redistribution game outlined in Table 1. The Proposition assumes that both the rich minority and the poor majority play grim trigger strategies with their respective fiscal policies by defecting forever to φ_H and τ_U respectively whenever the other deviates from their cooperating strategy of φ_L and τ_F.

SUSTAINABLE DEMOCRATIC FEDERALISM: *For a political economy described by the payoffs of Table 1 and satisfying the Assignment and Border constraints, there exists a grim trigger strategy equilibrium in which democratic federalism is sustainable as a long-run equilibrium of the redistribution game, and in that equilibrium:*

---

[8] See, for example, Gibbons (1992), pp. 88-99.

[9] See Gibbons (1992), pp. 100-102.

1) The central government majority chooses a level of intergovernmental transfers (tax rate) bounded between a maximal grant (tax rate) acceptable to the elite and a minimal grant (tax rate) acceptable to the majority specified as:

$$g(\tau_U) > g^{max}(\delta) \geq g(\tau_F) > g^{min}(\delta) > 0,$$

for some discount factor ($\delta \leq 1$), and;

2) The elite province(s) adopts the fiscal strategy φ_L.

The discount factor δ measures the equivalent value today of 1 Rand received a year from now, where $\delta = 1/(1 + r)$ and where r is the familiar rate of time preference. Patient people have low rates of time preference and thus values of δ very close to 1; impatient people have high rates of time preference and values of δ much below 1. It is easy to generalize the Proposition to discount factors unique to each player as well, as $g^{max}(\delta)$ depends only upon the rich minority's discount factor while $g^{min}(\delta)$ depends only upon the poor majority's discount factor. More generally then, the Proposition says that for some pair of δ's there is a maximal level of redistributive transfers $g^{max}(\delta)$ that the rich minority will pay each year and still cooperate and a minimal level of redistributive transfers $g^{min}(\delta)$ that the poor majority will accept each year and still cooperate. At least for least those δ's, $g^{max}(\delta) > g^{min}(\delta)$ and democratic federalism can be sustained.

Unfortunately there is no guarantee that these sustaining rates of time preference match those of the rich minority and poor majority in South Africa. Inman and Rubinfeld (2008) show that as players become more impatient so that r rises and δ declines, the maximal transfer the rich will allow, $g^{max}(\delta)$, declines and minimal transfer the majority will accept, $g^{min}(\delta)$, rises. This makes sense. Impatient players want more now and if they are not compensated sufficiently over the long-run they will defect to their "grabbing" strategies – here τ_U and φ_H.

It is possible that $g^{min}(\delta) > g^{max}(\delta)$. In this case, and despite the fact that the federal constitution satisfies both Assignment and Border constraints, there is no tax rate or transfer level for which democratic federalism is sustainable as a long-run equilibrium. For democratic federalism to be a successful check majority redistribution, both the rich minority and the poor majority must be relatively patient players of

government's redistribution game.

D. Summary: South Africa turned to federal governance as a solution to one of transition politics' central challenges: How can the new poor majority credibly promise not to exploit the now vulnerable rich minority? We have outlined the conditions necessary for a federal constitution to provide this protection. From the Assignment constraint, the elite must be a low cost provider of redistributive services important to the majority and those services must be assigned to provincial governments. From the Border constraint, the elite must politically control at least one important province and, given the central government's level of assigned services, have an incentive to punish the majority by capturing intended redistributive transfers when the central government's redistributive tax rate gets too high. Finally, both the rich minority and the poor majority must be sufficiently patient that the long-run economic benefits of the cooperative, federal outcome are preferred. Under these conditions, democratic federalism does offer the promise of elite protection. Whether this promise can be realized in South Africa, particularly under the new Zuma presidency, is our next question.

III. Can South Africa's Federal Democracy Check Majority Redistribution?

To understand whether South Africa's commitment to a federal democracy can provide a credible check against majority expropriation of the wealthy new minority, we first specify the underlying fiscal redistribution game of Table 1 for the South African political economy as anticipated by the NP and ANC negotiators at the time of the new constitution, 1996. Given this specification we then ask if our Assignment and Border constraints were satisfied, allowing elite high capture to stand as a credible punishment if the majority were to defect from their federal promise not the exploit the rich minority. Finally, we check to see if this threat of punishment is sufficient to ensure a federal fiscal allocation as a long-run equilibrium to the redistribution game.

A. Political Economy: At the time of the transition, the economic elite voting age population (N_0)

is estimated as 9.6 million residents.[10] At the time of the transition, average income earned by a typical middle and upper income resident (Y) is estimated as 86,000 (real 2000) Rand/Elite Adult (≃$22,000 USD in 1996). The poor majority population (M) at the time of the transition is estimated to be 25 million residents. The average income of a typical poor majority adult (W) is estimated at 9,700 (real 2000) Rand/Majority Adult (≃$2,500 USD).

Redistributive taxes paid by upper income households is approximated by surcharge of τ Rand per upper income resident. Redistributive revenues will equal $\tau \bullet N(\tau)$. The simplest specification for elite tax avoidance is $N(\tau) = N_0 - \beta \bullet \tau$, where N_0 is the initial minority elite population of 9.6 million and β (> 0) measures the degree of tax avoidance as τ rises. We calibrate β to imply a plausible peak to the national revenue hill from elite resident taxation. Setting $\beta = .00015$ sets the revenue maximizing tax rate per elite resident for redistributive services at 32,000 Rand/elite resident, or approximately 37 percent of average middle income residents' incomes (= 37,000R/86,000R) paying taxes. Since South Africa currently taxes income at an average rate of 30 percent for non-redistributive services (defense, domestic protection, higher education, infrastructure spending), the implied maximal tax rate on earned income is estimated to be .67. This rate is consistent with most recent estimates of revenue maximizing rates for developed economies and seems plausible estimate for the formal economy of South Africa.[11] The poor majority does not pay the redistributive tax surcharge. At the time of the transition, this maximal tax rate will generate approximately 6200 Rand/majority resident. If we allow for the 8 percent growth in real incomes per capita since 1996, then maximal tax revenues can rise to 35,000 Rand/elite resident, generating 6700 Rand/majority resident for redistribution.

---

[10] Sources for all data specifying the South African political economy at the time of the transition from apartheid to democracy is described in detail in a Data Appendix, available from the authors upon request.

[11] For estimates of the maximal rate for the U.S. economy, see Gruber and Saez (2002). Our specification of the revenue hill in terms of the number of taxpaying residents with fixed incomes can be shown to be functionally equivalent to a revenue hill defined in terms of tax avoidance and a fixed number of residents.

The exogenous payment (Z) required of the central government for "other redistributive activities" is set at 600 Million (Real 2000) Rand per year. This is the amount recommended by the FFC in their first budgets to be paid annually to the province of KwaZula-Natal as part of an implicit agreement for the IFP to peacefully join the new democracy.

The elite provinces' rates of capture of unallocated redistributive grants are set equal to recent estimates of the rates of capture of central government transfers by the administrators of Ugandan local schools: $\varphi_L = .20$ and $\varphi_H = .85$; see Reinikka and Svensson (2003, 2004). High shirking, high capture occurred when Ugandan school grants were not publicized; minimal shirking, low capture occurred when residents were informed of available transfers. The estimates are consistent with the vast international literature on the ability of local officials to capture central government transfers for their own uses; see Inman (2009).

The estimates of protests costs ρ born by the elite when they adopt high capture are from two sources. A lower bound estimate is the average economic costs of urban riots in the United States during the 1960's and 1970's as provided by Collins and Margo (2007). They estimate the decline in property values of cities having experienced a moderate to severe urban riot as ranging from ½ to 1 percent of property owners' annual income. From our estimate of elite incomes, this sets a lower bound to ρ of 430 Rand per elite resident (= .005•86,000 Rand/Elite Resident). A better organized or more militant majority may be able to impose higher protest costs. As an upper bound we use .02 of elite income or ρ = 1720 Rand/Elite Resident, based upon our estimates of the costs of organized Black union boycotts on the South African economy during the last years apartheid.[12]

We assume redistributive public services use a production technology specified as $q = a\bullet(X/M)$,

[12] Our upper bound estimate comes from a regression to explain South African growth from 1950 to 2000 controlling for trade openness, the rate of gross investment, and international sanctions and yields a negative impact of Black unionization (COSATU) over the period 1985-1990 of -.021 (S.E. = .008). For these regression results, see the Data Appendix, available upon request.

where (X/M*)* is public employees (X) per majority resident (M) and *a* is employee productivity measured by years of training.[13] We use as our estimates of a_e the years of schooling of white teachers (a_e = 17 years), of a_m the years of schooling of certified majority teachers (a_m = 14 years), and of a_u the years of schooling of uncertified or untrained majority teachers (a_u = 7 years). We are therefore measuring the required level of redistributive services q as "public employee training years per majority resident." For example, the FFC's recommended level of redistributive services for the initial budgets of the new democracy were to provide 1 teacher per 38 school-aged children, 3.5 preventive health care clinic visits a year for each majority adult and child, and 4500 (real 2000) Rand for each income eligible child, disabled and elderly majority resident. Together these targets require redistributive grants sufficient to pay for .038 public employees per majority resident. The average level of training of public employees in South Africa at the time of the transition was 15 years. Thus the implied target value of redistributive service inputs is q = .57 public employee training-years per majority resident (= 15 years of training *x* .038 public employees per majority resident).[14] We estimate the corresponding value of q for the last of the apartheid years to be q = .36.[15]

---

[13] This specification of the production technology for redistributive services assumes the provision of K-12 education, primary health care, transfer income transfers are best viewed as "private goods" without significant economies of scale in population. Most recent production studies for these services confirms this assumption.

[14] 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. See FFC, *The Allocation of Financial Resources Between the National and Provincial Governments*, *FY 1996/97*, September 8, 1995, p. ii. Assuming an average level of training of 15 years per public employee, this implies a value for q = .57 public employee training-years per majority resident.

[15] In the last years of apartheid the average class size for majority children was one teacher per 41 children (Development Bank of South Africa, *South Africa's Nine Provinces: A Human Development Profile*, March, 1995, p. 94). The average number of yearly clinic visits per majority residents was 1.8 visits per year (Financial and Fiscal Commission, *The Allocation of Financial Resources Between the National and Provincial Governments*, September 8, 1995). The average income transfer per eligible recipient was 262Rand per recipient or about 30 Rand per majority resident (Development Bank of South Africa, *South Africa's Nine Provinces: A Human Development Profile*, March, 1995, p. 38). Together these service requirements imply a value of .024 public employees per majority resident. Assuming training levels again equal 15 years per average public employee, the implied value of q is .36 (= 15 *x* .024).

Finally, as a benchmark for "maximal" demand for redistributive services, if the majority were to demand services fully comparable to those now received by elite residents and transfer incomes twice that of now paid to eligible children and disabled and elderly adults, then q would require funding equivalent to 1.14 public employee training-years per majority resident.[16]

The cost per majority resident of providing q in the elite province will equal $s_e(q) = S\cdot(q/a_e)$ and in the majority province, $s_m(q) = S\cdot(q/a_m)$, where S is the uniform salary paid to public employees. We estimate S as the average teacher salary in 2001 equal to $80,000 Rand per employee. For example, to fund the FFC's initial budgets of q = .57, spending of $s_e(q) = S\cdot(q/a_e) = 80{,}000\cdot(.57/17) = 2682$R/Majority Adult would be required in the Western Cape and $s_m(q) = S\cdot(q/a_m) = 80{,}000\cdot(.57/14) = 3257$R/Majority Adult in the ANC controlled provinces. On a per capita basis including children, this implies spending for redistributive services of 1448 Rand/person in the Western Cape and 1759Rand/person in the ANC provinces.[17] These predicted spending levels for q = .57 are reassuringly close to actual spending for redistributive services in the Western Cape and majority provinces, see Table 3, Columns (6) and (9).

The majority residents' demand for redistributive public goods is specified as $\lambda\bullet\upsilon(q) = \lambda\bullet\ln(q)$, where λ measures the intensity with which the majority demands redistributive goods bounded as $6100 \geq \lambda \geq 3960$. The lower bound corresponds to the level of q = .74, chosen during the last year's of the Mbeki presidency when it is plausible to assume he began to respond to the demands of the ANC's majority member.[18] The upper bound is set under the assumption that the majority will demand that level of education

---

[16] A plausible upper bound for redistributive services would assume the majority demands the current level of services inputs now provided to the elite for K-12 schooling (1 teacher per 20 students) and primary health care (7 visits per elite adult per year, or about 3.5 visits per family member) plus twice the current level of funding for children, disabled, and elderly pensions (9000R per eligible majority resident). These services will require the equivalent .076 public employees per majority resident, or, assuming an average level of training for such employees of a = 15 years, a value of q = 1.14 public employee training years per majority resident (= 15 years x .076).

[17] The ratio of Majority Adults (25 Million) to the full population (46 Million) is .54. Thus Spending/Person will be .54 *x* Spending/Majority Adult.

[18] See *Financial Times*, September 21, 2007, p. 4, "Boost for Zuma's Leadership Campaign." To estimate λ, the demand for q will be $q = \lambda/p(\mu)$. Thus $\lambda = q\cdot\bullet\ p(\mu)$. We estimate $p(\mu) = 5355$R for this political economy. Thus

and health care services currently received by upper income households plus twice the currently proposed level of transfers for poor children, disabled, and elderly persons – that is, $q = 1.14$.[19]

The value for the elite discount factor δ_e assumes a rate of time preference for the elite of .08, the real rate of return on ten year South African bonds over the period 1996-2006: $\delta_e = .93$. An upper estimate for the poor majority's discount factor is $\delta_m = .93$ under the assumption that the poor can access the capital markets at the same rate as can the elite. More likely the poor are credit constrained and can only borrow informally, however. Recent experimental evidence for South Africa's urban poor suggests such credit constrained households would borrow – and can do so profitably – at an annual interest rate of 200%! Because of information asymmetries, however, banks have not made these loans; see Karlan and Zinman (2008). The implied discount factor for the poor majority can therefore be as low as $\delta_m = .33$. Again, we shall test for the sensitivity of our main conclusions to variation of the poor majority's discount factor between the bounds, $.93 \geq \delta_m \geq .33$. Table 2 summarizes our specification of the South African political economy.

B. Satisfying Assignment and Border Constraints: To hold in check the majority's preference for maximal redistribution, the leaders of the elite province must be able to credibly adopt high capture as a permanent punishment for such a violation of the federal compact. This requires the federal constitution's choice of provincial borders and services assignments jointly satisfy our Assignment and Border constraints. The Border constraint requires the share of the majority population assigned to the elite province satisfy the bounds $\mu^{max} \geq \mu > \mu^{min}(q)$. The Assignment constraint requires that the level of centrally chosen redistributive services satisfy the bounds $q^{max}(\mu) \geq q > q^{min}(\mu)$. The shaded area of Figure 1 shows all the values of μ and q that satisfy both constraints for the political economy specified by Table 2. Does South

---

for $q = .74$, $\lambda = 3962 = .74 \cdot 5355$.

[19] Thus the upper bound corresponds to our estimate of the maximal demand for redistributive services equal of $q = 1.14$; see footnote 26. For this value of q, $\lambda = 6105 = 1.14 \cdot 5355$.

Africa's constitution's specification of μ and q fall within the shaded area?

Provincial borders must be drawn so that the voting age population of the majority living in the elite province never exceeds the minimal level of the voting age population of upper income residents in that province, which we take to be $N(\tau_U)$. Thus the maximal share of all majority residents that can be assigned to that province will be $\mu^{max} = N(\tau_U)/M = .192$ for our specification of the revenue hill where $N(\tau_U) = 4.8$ million elite residents.[20] At the time of the adoption for the new constitution, the Western Cape was seen as the most likely elite controlled province. The constitution explicitly specified the Western Cape border to ensure NP, or more generally elite, elected candidates would control the province.[21] Recent voting histories for the Western Cape suggests that a provincial partition allocating 18.4 percent of South Africa's adult majority population to the Western Cape would be sufficient to protect elite influence over provincial politics.[22] Thus we set the $\mu = .184$. The lower bound $\mu^{min}(q)$ depends upon the level of assigned redistributive services, q. From Figure 1 for point A, as long as $\mu \geq .055$, there will be some combination of μ and q that will satisfy both the Border and Assignment constraints. Only for potentially high values of assigned redistributive services – specifically, when $q > .86$ – will there be a violation of the Border constraint with $\mu^{min}(q) > \mu = .184$; see Figure 1.

Given a value for the borders of the elite province with $\mu = .184$, our Assignment constraint requires that $q^{max}(\mu) = .86 \geq q > q^{min}(\mu) = .58$; see Figure 1. The South African Constitution allows for the assignment of most redistributive services to either central or provincial levels of government. To date, the

---

[20] This is a conservative estimate of μ^{max} since this assumes that all partial- or non-tax paying elite residents no longer vote.

[21] See Muthien and Khosa (1998), particularly pp. 43-49.

[22] 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. We therefore define $\mu = (M_e/M)$ so that $N(\tau_F)/[M_e + N(\tau_F)] = .51$. $N(\tau_F)$ is estimated to be only slightly larger than 4.8 Million. Thus $M_e = 4.6$ Million, and $\mu = (M_e/M) = 4.6M/25M = .184$. The 2009 election results confirm that this initial population partition has been sufficient to maintain elite political control over the Western Cape.

provision of K-12 education, primary health care, and the administration of children, disabled and elderly income transfers have been assigned to the provinces. A minimal level of service responsibility and funding for that minimum are decided by the central government, however. President Mandela's budgets followed the recommendations of the FFC and funded 1 teacher per 38 school-aged children, 3.5 preventive health care clinic visits a year for each majority adult and child, and 4500 (real 2000) Rand for each income eligible child, disabled and elderly majority resident requiring an value of q = .57 public employee training-years per majority resident. Mandela's redistributive budgets were approved without dissent by the Parliament. The average budget for the Mandela years misses the lower bound only slightly; his last budget funded a value of q = .59 and just satisfies the lower bound. These initial Mandela budgets were sufficient to support democratic federalism, but they clearly favored the elite. It is instructive to note that the Mandela budgets followed the FFC recommendations and Commission membership was equally divided between NP and ANC appointments.

In contrast to the budgets of President Mandela, the Mbeki redistribution budgets grew more expansive over his tenure, funding an initial level of q = .59 in his first (FY 2001/02) budget rising to q = .74 by his last (FY 2008/09) budget. Finally, Interim President Motlanthe representing the new, now more liberal ANC, has increased q to .81 in this year's budget, a one year increase in redistributive services of nine percent. Still, the Mbeki and Motlanthe redistributive budgets continue to satisfy the Assignment constraint for credible elite punishments: $q^{max}(\mu) = .86 \geq .81 > q^{min}(\mu) = .58$; see Figure 1.

So far at least, the ANC's choice of required redistributive services has left the Western Cape's "high capture" threat in tact as a credible punishment for excessive overall redistribution. Whether this threat of is sufficient to hold aggregate redistributive spending below its maximum depends upon elite and majority discount rates satisfying the requirements of our Proposition for Sustainable Federalism.

C. Satisfying the Proposition for Sustainable Federalism: For democratic federalism to be a sustainable constitutional contract the elite minority must be able to check the poor majority's preferred

option of maximum redistribution. From our Proposition for Sustainable Federalism, this is possible if our Assignment and Border constraints are met and if the majority values future income enough that the elite's threat to adopt a maximal shirking, high capture punishment forever (the "grim trigger" strategy) imposes long-run losses on the majority that exceed the short-run benefits of going to maximal taxation. A sufficiently patient majority gives the threat of high capture the clot it needs to check redistributive taxation.

For our specification of the South African political economy, the requirement for sustainable democratic federalism that $g^{max}(\delta_e) \geq g^{min}(\delta_m)$ holds *in principle*; see Table 3, cols. (1) and (2). For the Mandela, Mbeki, and now Motlanthe administrations , the maximum level of redistribution funding that the elite will offer, denoted $g^{max}(\delta_e)$, has exceeds the minimal level of redistribution the majority will accept, denoted $g^{min}(\delta_m)$. Importantly, this requirement for sustainable federalism is met even for a very impatient majority with an annual rate of time preference of 200 percent and an associated discount factor as low as $\delta_m = .33$.

To date the actual levels of redistributive transfers, reported as $g.(\tau_F)$ in Table 3, col. (4), have fallen below the required ranges for sustainable redistributive bargains, particularly so during the Mandela presidency. Perhaps to reassure the elite during the first years of the new democracy, Mandela appears to have used his considerable political capital along with the constitution's presidential powers to hold in check majority demands for a more egalitarian public sector. Under Mandela the levels of required redistributive services (q) are low and barely satisfy the lower bound for our Assignment Constraint; see Table 3, col. (1) and Figure 1. Further, the aggregate level of redistributive transfers during the Mandela years fell well short of our predicted level of redistributive spending needed to satisfy the majority, denoted as $g^{min}(\delta_m)$; compare Col. (4) to Col. (3) in Table (3) for the Mandela budgets. Budgets favoring the elite province of Western Cape are particular evident in the first years of the new democracy, perhaps to ease the "shock" of the transition; see Table 3, Cols. (5)-(7).

The aggregate redistribution budgets of Mbeki and Motlanthe are far more responsive to the

predicted redistributive requirements of the majority. Over Mbeki's tenure, the average aggregate transfer budget increased by 34 percent in real terms, rising from 2,242R/Adult in the last year of the Mandela administration to 3,003R/Adult in the last Mbeki budget for FY 2008/2009. Most of this increase went to funding education, health care, and targeted income transfers in the more expensive, majority run provinces; compare Table 3, col. (5) to col. (8). The last Mbeki budget and the recently approved budget of Motlanthe are very close to our predicted value for the majority's required $g^{min}(\delta_m)$, both for the high (3,155R/Adult; $\delta_m = .93$) and very low (3,251R/Adult; $\delta_m = .33$) discount factors.

We conclude that while President Mandela's redistributive budgets fell well short of the predicted levels of $g^{min}(\delta_m)$ required for sustainable federalism, Presidents Mbeki's last, and Motlanthe's only, budgets are nearly there, missing $g^{min}(\delta_m = .33)$ by at most 248R/Adult ($\simeq$ \$35/Adult). Redistributive fiscal politics in South Africa appear to now met our conditions for a sustainable federal democracy. President Mandela's budgets favored the elite and were not sustainable. The political pressure was to increase redistributive spending and the Mbeki and now Motlanthe budgets responded. Whether this compromise for sustainable federalism can continue under the new leadership of Jacob Zuma is the open question.

IV. The Zuma Challenge

The new economic militancy of the ANC majority leading to the election of Jacob Zuma as President poses two challenges to the new democracy of South Africa, one economic and the second political. The economic challenge is that which faces every developing economy: Can the demands for redistributive services by the poor majority be satisfied by the economic resources of the wealthy minority? It is clear that the new ANC, represented by the interim presidency of Kgalema Motlanthe, has been responsive to those demands. Real government spending for redistributive services rose by over 9 percent in the one year of Motlanthe's presidency; see Table 3. col. (1). The real aggregate redistributive budget rose by just over 2.5 percent, from 3000 Rand per adult to 3081 Rand per adult; Table 3, col. (4).

Our estimate of the peak of the revenue hill for taxable income that can used to support redistributive services is 3400 to 3600 Rand per (elite plus majority) adult, or 6200 to 6700 Rand per majority resident only. This is the maximal redistribution that could occur from the taxation of elite earned incomes. Today's (FY 2010) redistributive budget is 3081 Rand per all adults, or 5700 Rand per majority resident. South Africa is very near the top of its revenue hill for redistributive spending.

Providing education and health care services comparable to that now obtained by the elite citizens and doubling the income transfer budget to eligible residents – that is, setting q = 1.14 in Figure 1 – would require a redistribuitve budget of 3500 Rand per adult, or 6500 Rand per majority resident. Meeting these maximal service demands is economically feasible, but such funding will fully exhaust provincial budgets – that is, "free" provincial spending measured by b(q) will driven to zero. As the opening quotation implies, if a Zuma Presidency is compelled by the militancy of the new ANC majority coalition to be "fully equalizing," then either all other provincial responsibilities must go unfunded or the central government must find other revenue sources besides earned incomes – for example, high capital taxation, the nationalization of industries, or the taking of private lands. Such policies will likely place South Africa on a new, and potentially damaging, long-run economic trajectory.

The Zuma presidency also challenges the political sustainability of the initial democratic contract as embodied in the Constitution of South Africa. That contract established provinces with free elections of provincial leadership and the capacity to exercise that leadership through credible policy choices. Such choices are only available if the provinces have at least some unconstrained fiscal resources. While the constitution does allow for provincial taxation of provincial income, the national Parliament has not yet granted those powers. This leaves intergovernmental transfers from the central government, our $g(\tau)$, as the only potential source of "free" money for provincial allocations. But as Table 3, cols. (7) and (10) make clear, unconstrained provincial resources, our b(q), have been shrinking steadily as the demands for more redistributive services, our q in Table 3, col. (1), have increased. As the redistributive requirements from

the central government increase, provincial policy discretion declines.

The redistributive demands of the new ANC majority is likely to continue this trend, with two important consequences for South African democracy. First, the original constitutional contract required for sustainable democratic federalism may be undone. If the central government's required value for redistributive services, now at q = .81 under Motlanthe, rises by only 6 percent to q^{max} = .86, then democratic federalism is no longer sustainable as a feasible check on redistributive taxation; see Figure 1. The maximal redistributive demands of the new ANC implying q = 1.14 lie well outside the feasible set for sustainable democratic federalism; again, see Figure 1. If assigned redistributive services by the central government, q, become "too large" then remaining fiscal resources under the control of the leadership of the elite province, the Western Cape, becomes "too small" to pose a credible threat to the well-being of the majority ANC political party. Free of such a threat, the ANC controlled central government can then adopt maximal redistributive taxation without penalty. Now all fiscal policies will be set by the central government. While provinces may continue to exist as political subdivisions, as policy centers they will have become simply administrative arms of the central government. For all practical purposes, this is a unitary, not a federal, state.

There is some potential flexibility to the value q^{max} that defines the long-run viability of federalism, however. Lower values of protest costs (ρ) imposed by the Western Cape majority when the elite adopts high capture makes it easier for the elite to credibly threaten to punish the majority for high taxation and thus q^{max} can rise. For values of ρ at our lower bound of 430 Rand per elite resident (about $70/person), q^{max} increases from .86 to 1.00. This gives the Zuma administration somewhat more freedom to raise redistributive services without threatening the long-run political viability of the federal contract. Interestingly, the more success President Zuma has in controlling the militancy of his ANC constituents the more room he has within the structure and spirit of the original constitution to provide redistributive services.

Second, though not part of our formal analysis here, there is ample evidence from the political history

of federal democracies that national leaders are often chosen based upon their successful record as provincial leaders. In Australia, Canada, Germany, and the United States provinces have often served as proving grounds for higher office. But for provincial leaders to offer a credible case for election, they will need a record of accomplishment independent of the central government's. This requires provincial resources under provincial control. A Zuma presidency responsive to the redistributive demands of the new ANC majority setting $q > q^{max}$ is likely to lead to a *de facto* unitary state. In this setting, provincial leaders become administrators within central government policy with future presidential candidates most likely limited to those who perform well within the party hierarchy.

The emerging pressure for redistributive services from the new ANC majority raises important challenges threatens the both the economic and the democratic future of the new South Africa. While the constitutional compromise of 1996 provided the institutions for checking these redistributive pressures, our analysis has made clear that these institutions alone are not enough. The players of the constitutional game must want it to succeed. It is clear that the fiscal choices of President Mandela and Mbeki did provide for sustainable democratic federalism. Mandela through the informal power of his personal charisma and Mbeki through the formal powers of the presidency both held redistributive pressures at bay. This is the leadership task before President Zuma. The likely best response to the Zuma challenge is President Zuma himself.

V. Conclusion

South Africa's recent political history provides valuable lessons for how a new democracy might manage a peaceful transition from elite-run autocracies to majority-run democracies. Central to such transitions is to establish credible protections for elite assets and incomes in the face of majority demands for significant redistributions. The current literature suggests three such protections: 1) Continued elite control of the military (Acemoglu and Robinson, 2001); 2) An upper legislative chamber controlled by the elite with veto powers (Lijphart, 1984); and 3) the gradual extension of the franchise timed to match the

growth of a propertied middle class (Boix, 2003; Lizzeri and Persico, 2004). The majority ANC explicitly rejected each of these alternatives, but did accept the elite NP's proposal of democratic federalism with elite control of at least one province. We have generalized the logic of this fourth alternative and clarified exactly what is required for the institutions of federal governance to credibly protect elite resources from full expropriation.

While showing the potential of federal institutions, South Africa's most recent political history reveals their limitations too. The ability of democratic federalism to check a majority's redistributive preferences depends crucially upon how badly the majority wants to redistribute. The majority cannot demand too much, too quickly. We have estimated the upper bound for majority's demands for redistributive goods and found Presidents Mandela and Mbeki were able to hold those demands below the needed threshold. The new, more radical ANC, will be pushing hard against those limits. If redistributive preferences are left unchecked by President Zuma, then democratic federalism is likely to give way to de facto unitary governance and maximal taxation. As is true for all self-enforcing constitutions, it is institutions *and* preferences that determine outcomes, not institutions alone.

REFERENCES

Acemoglu, Daron and James Robinson (2001), "A Theory of Political Transitions," *American Economic Review*, Vol. 91, September, 938-63.

Akerlof, George and Rachel Kranton (2005), "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, Vol. 18, Winter, 9-32.

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

Collins, William and Robert Margo (2007), "The Economic Aftermath of the 1960's Riots in American Cities: Evidence from Property Values," *Journal of Economic History*, Vol. 67, December, 849-883.

Constitution of the Republic of South Africa (1996), Act 108 of 1996.

Gibbons, Robert (1992), *Game Theory for Applied Economists*, Princeton, N.J.: Princeton University Press.

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

Inman, Robert (2009), "The Flypaper Effect," *The New Palgrave of Economics.*

Inman, Robert and Daniel Rubinfeld (2008), "Federal Institutions and the Democratic Transition," NBER Working Paper 13733.

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

Lizzeri, Alessandro and Nicola Persico (2004), "Why Did the Elites Extend the Suffrage? Democracy and the Scope of Government, with an Application to Britain's 'Age of Reform'," *Quarterly Journal of Economics*, Vol. 119, May, 707-765.

Muthien, Yvonne and Meshack Khosa, (1998), "Demarcating the New Provinces: A Critical Reflection on the Process," in Y. Muthien and M. Khosa (eds.), *Regionalism in the New South Africa*, Brookfield, U.S.: Ashgate, 23-56.

Reinikka, Ritva. and Jakob Svensson (2003), "The Power of Information: Evidence From a Newspaper Campaign to Reduce Capture," Mimeo. World Bank, December.

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

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

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

TABLE 1: THE ANNUAL REDISTRIBUTION GAME

MAJORITY PAYOFFS per MAJORITY CITIZEN

STRATEGIES	MAJORITY STRATEGY: $\tau_F < \tau_U$	MAJORITY STRATEGY: $\tau_F = \tau_U$
ELITE STRATEGY: φ_L	$\omega(\tau_F; \varphi_L) = W + g(\tau_F) + [\lambda \cdot \upsilon(q) - s_F(q)] - \mu \cdot \phi_L \cdot [g(\tau_F) - s_e(q)]$	$\omega(\tau_U; \varphi_L) = W + g(\tau_U) + [\lambda \cdot \upsilon(q) - s_F(q)] - \mu \cdot \phi_L \cdot [g(\tau_U) - s_e(q)]$
ELITE STRATEGY: φ_H	$\omega(\tau_F; \varphi_H) = W + g(\tau_F) + [\lambda \cdot \upsilon(q) - s_F(q)] - \mu \cdot \phi_H \cdot [g(\tau_F) - s_e(q)]$	$\omega(\tau_U; \varphi_H) = W + g(\tau_U) + [\lambda \cdot \upsilon(q) - s_F(q)] - \mu \cdot \phi_H \cdot [g(\tau_U) - s_e(q)]$

ELITE PAYOFFS per ELITE CITIZEN

STRATEGIES	MAJORITY STRATEGY: $\tau_F < \tau_U$	MAJORITY STRATEGY: $\tau_F = \tau_U$
ELITE STRATEGY: φ_L	$y(\tau_F; \varphi_L) = Y - \tau_F + \varphi_L \cdot [g(\tau_F) - s_e(q)] \cdot [\mu \cdot M/N(\tau_F)]$	$y(\tau_U; \varphi_L) = Y - \tau_U + \varphi_L \cdot [g(\tau_U) - s_e(q)] \cdot [\mu \cdot M/N(\tau_U)]$
ELITE STRATEGY: φ_H	$y(\tau_F; \varphi_H) = [Y - \rho] - \tau_F + \varphi_H \cdot [g(\tau_F) - s_e(q)] \cdot [\mu \cdot M/N(\tau_F)]$	$y(\tau_U; \varphi_H) = [Y - \rho] - \tau_U + \varphi_H \cdot [g(\tau_U) - s_e(q)] \cdot [\mu \cdot M/N(\tau_U)]$

TABLE 2: POLITICAL ECONOMY OF SOUTH AFRICA[†]

DEMOGRAPHICS and INCOMES

N_0 = 9.6 Million Elite Adults.
M = 25 Million Majority Adults.
Y = 86,000 (Real 2000) Rand/Elite Adult.
W = 9,700 (Real 2000) Rand/Majority Adult.
$N(\tau) = N_0 - \beta \bullet \tau$, where $\beta = .00015$.

BUDGET CONSTRAINT: SPECIAL INTEREST PAYMENTS

Z = 600 Million (Real 2000) Rand Paid to KwaZula-Natal Province.

ELITE CAPTURE:

$\varphi^L = .20$ (Rate of Capture per Rand of Basic Grant).
$\varphi^H = .85$ (Rate of Capture per Rand of Basic Grant).

PROTEST COSTS

ρ = Protest Cost, where 1720R/Elite Adult $\geq \rho \geq$ 430R/Elite Resident.
ρ = 860R/Elite Resident (Specification for Figure 1).

REDISTRIBUTIVE SERVICE TECHNOLOGY AND COSTS

a_e = 17 (Years of Training; Elite Public Employee).
a_m = 14 (Years of Training; Trained Majority Public Employee).
a_u = 7 (Years of Training; Untrained Majority Public Employee).
S = 80,000 (Real 2000) Rand/Public Employee (Average Uniform Salary).
$s_e(q) = (S/a_e) \cdot q = (80{,}000/17) \cdot q = 4{,}706 \cdot q$ (Real 2000) Rand/Majority Adult.
$s_m(q) = (S/a_m) \cdot q = (80{,}000/14) \cdot q = 5{,}714 \cdot q$ (Real 2000) Rand/Majority Adult.
$s_u(q) = (S/a_u) \cdot q = (80{,}000/7) \cdot q = 11{,}428 \cdot q$ (Real 2000) Rand/Majority Adult.

PREFERENCES FOR REDISTRIBUTIVE GOODS

$\lambda \cdot \upsilon(q) = \lambda \cdot \ln(q)$, where $6100 \geq \lambda \geq 3960$.

DISCOUNT FACTORS

$\delta_e = .93$ (Rate of Time Preference, r = .08).
$\delta_m = .93$ (Rate of Time Preference, r = .08) to .33 (Rate of Time Preference, r = 2.00).

BORDER AND REDISTRIBUTIVE ASSIGNMENT

$\mu = M_e/M = .184$.
q = .53 to .59 (Mandela); q = .59 to .74 (Mbeki); q = .81 (Motlanthe).

[†] Source: Data Appendix available upon request.

TABLE 3: FISCAL REDISTRIBUTION IN SOUTH AFRICA:1996-2010
(Transfers per Capita; Real 2000 Rand)*

PRESIDENTIAL REGIME (δ_e; δ_m)	q [1]	$g^{max}(\delta_e)$ [2]	$g^{min}(\delta_m)$ [3]	$g_.(\tau_F)$ [4]	$g_e(\tau_F)$ [5]	$s_e(q)$ [6]	$b_e(q) = g_e(\tau_F) - s_e(q)$ [7]	$g_m(\tau_F)$ [8]	$s_m(q)$ [9]	$b_m(q) = g_m(\tau_F) - s_m(q)$ [10]
Mandela, 1996 (.93; .93)	.53	3299	3088	2189	2923	1371	1552	2119	1356	763
Mandela, 1996 (.93; .33)	.53	3299	3227	2189	2923	1371	1552	2119	1356	763
Mandela, 2001 (.93; .93)	.59	3301	3108	2242	2185	1455	730	2247	1778	469
Mandela, 2001 (.93; .33)	.59	3301	3233	2242	2185	1455	730	2247	1778	469
Mbeki, 2002 (.93; .93)	.59	3301	3108	2302	2196	1455	741	2313	1826	487
Mbeki,2002 (.93; .33)	.59	3301	3233	2302	2196	1455	741	2313	1826	487
Mbeki,2009 (.93; .93)	.74	3304	3155	3003	2671	2039	632	3040	2263	777
Mbeki,2009 (.93; .33)	.74	3304	3251	3003	2671	2039	632	3040	2263	777
Motlanthe, 2010 (.93; .93)	.81	3305	3175	3081	2775	2163	612	3115	2498	567
Motlanthe, 2010 (.93; .33)	.81	3305	3258	3081	2775	2163	612	3115	2498	567

* *SOURCES*: FY: 1995/96 to 1997/98: Financial and Fiscal Commission, *The Allocation of Financial Resources Between the National and Provincial Governments: FY 1997/98*, Tables 2, 3, 6b. FY 1998/99 to 2009/10: Minister of Finance, *Division of Revenue Bill, Various Years,* Part 4: Provincial Allocations.

NOTES TO TABLE 3: COLUMN DEFINITIONS

Column 1: q = Public Employee Training Years per Majority Adult for redistributive public services, defined to include K-12 education, primary health care services, and spending for children, disability, and elderly income transfers adjusted to "employees" after division by the average employee salary.

Column 2: $g^{max}(\delta_e)$ = Predicted maximum redistributive transfer per capita the upper income residents will pay for support of redistributive services (q) and basic provincial transfers (b) while remaining committed to democratic federalism and the cooperative strategy of low shirking, low capture (φ_L).

Column 3: $g^{min}(\delta_m)$ = Predicted levels of the minimal redistributive transfer per capita the poor majority residents will accept for support of redistributive services (q) and basic provincial transfers (b) while remaining committed to democratic federalism and the cooperative strategy of a less than maximum redistributive tax rate, $\tau_F < \tau_U$.

Column 4: $g_.(\tau_F)$ = Average redistributive transfer per capita paid to all provinces.

Column 5: $g_e(\tau_F)$ = Average redistributive transfer per capita paid to the elite province, Western Cape.

Column 6: $s_e(q)$ = Service grants per capita allocated to the elite province, the Western Cape, to specifically fund redistributive services K-12 education, primary health care services, and income transfer grants for the qualified children, disabled, and elderly. Known as the National Standards Grant, or "S"grant.

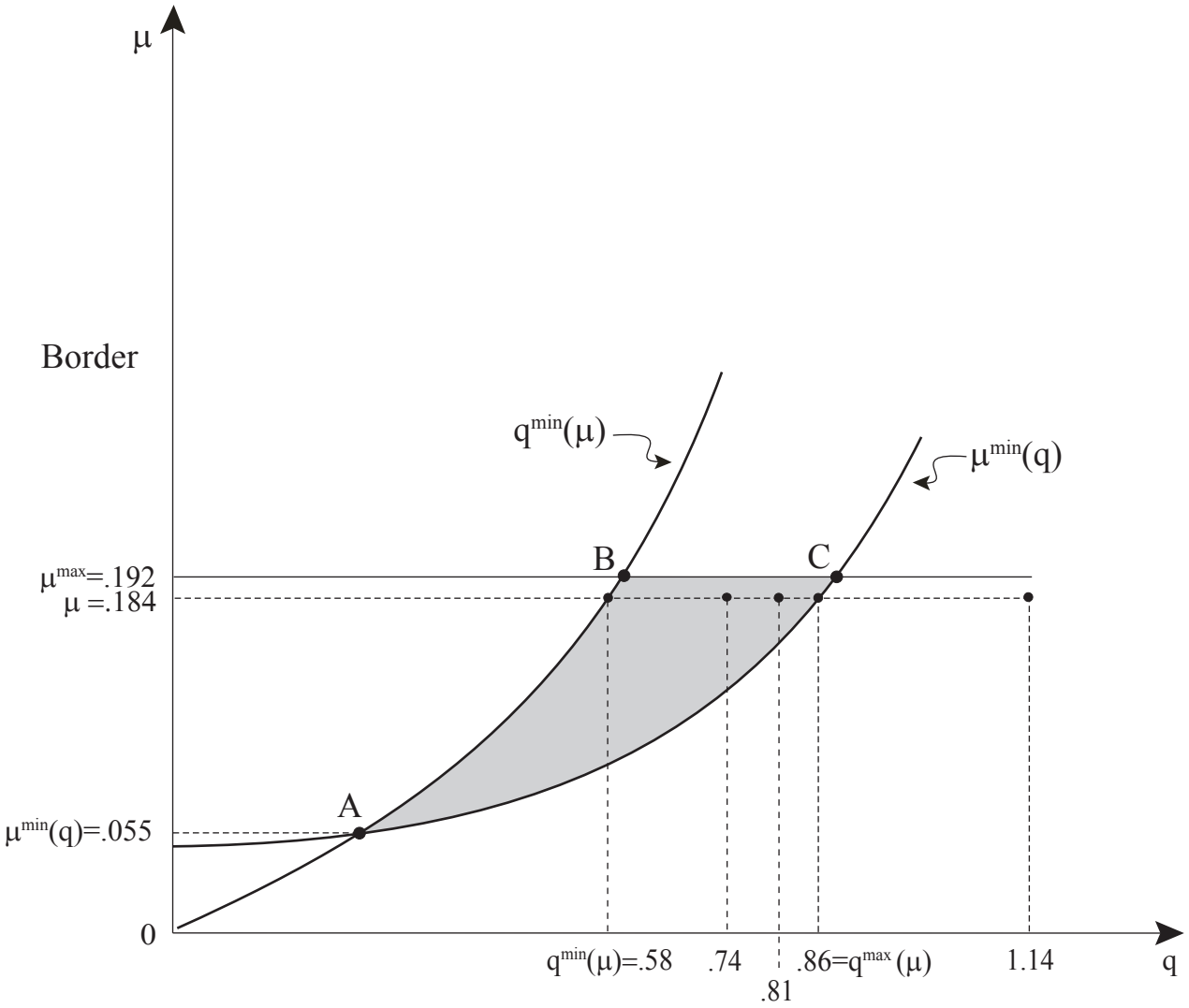
Column 7: $b_e(q) = g_e(\tau_F) - s_e(q)$ = Aggregate provincial transfer per capita paid to the Western Cape and not committed to funding of redistributive services, q. Known as the Basic Grant or "B" grant.

Column 8: $g_m(\tau_F)$ = Average redistributive transfer per capita paid to all majority controlled provinces, including KwaZulu-Natal.

Column 9: $s_m(q)$ = Service grants per capita allocated to majority provinces, including KwaZulu-Natal, to specifically fund redistributive services denoted as K-12 education, primary health care services, and income transfer grants for the qualified children, disabled, and elderly. Known as the National Standards Grant, or "S"grant.

Column 10: $b_m(q) = g_m(\tau_F) - s_m(q)$ = Aggregate provincial transfer per capita paid to majority provinces, including Kwa-Zulu-Natal, and not committed to funding of redistributive services, q. Known as the Basic Grant or "B" grant.

Figure 1: Feasible Democratic Federalism



Assignment

Coordinates for Points:

A: μ = .055; q = .25
B: μ = .192; q = .61
C: μ = .192; q = .99

APPENDIX: Defining Border and Assignments Constraints for Credible Elite Punishments

The *Border constraint* for credible elite punishments using the high capture strategy (φ_H) is specified to ensure that the wealthy province(s) have both the political ability to select that policy option – that is, the elite voters are a majority in the province – and the incentive to adopt that strategy – that is, $y(\tau_U; \varphi_H) > y(\tau_U; \varphi_L)$. The lower bound for the *Assignment constraint* for credible elite punishments is specified to ensure that the majority does not abandon the use of provinces if the elite chooses the high capture punishment – that is, $\omega(\tau_U; \varphi_H) > \omega(U)$. The specifications for $y(\tau_U; \varphi_H)$, $y(\tau_U; \varphi_L)$, and $\omega(\tau_U; \varphi_H)$ are given in Table 1; $\omega(U)$ is specified in the text. The upper bound for the Assignment constraint is set to ensure that the elite still wants to impose the high capture penalty even when they bear a penalty of ρ/Elite Resident.

We define the Border and Assignment constraints for two constitutional regimes: 1) The **q**-Regime where the central government's choice of redistributive services, q, is constitutionally required and enforced to equal an exogenously set value of $q = \mathbf{q}$; and 2) the q*-Regime where the majority-run central government is free to select its preferred level of $q = q^*$ based upon the fiscal incentives implicit in the political economy. Our analysis in Figure 1 is based upon specifications of the constraints for the q*-Regime. We specify the demand for q as $q = q^*_L(\mu, \lambda)$ when the elite province selects low capture and as $q = q^*_H(\mu, \lambda)$ for the case where the elite province selects high capture. Higher values of μ allocate more majority residents to the efficient, elite province and this increases the demand for q. Higher values of λ represent higher marginal benefits from q, also increasing the majority's demand for redistributive services.

Border Constraint:

***q*-Regime**: From the specifications in Table 1, the requirement that $y(\tau_U; \varphi_H) > y(\tau_U; \varphi_L) \Leftrightarrow (\varphi_H - \varphi_L)[g_U - s_e(\mathbf{q})][\mu \cdot M/N(\tau_U)] > \rho$, or:

$$(M_e/M) = \mu > \{\rho[N(\tau_U)/M]\}/\{(\varphi_H - \varphi_L)[g_U - s_e(\mathbf{q})]\} \equiv \mu^{min}(\mathbf{q}).$$

where μ is the fraction of majority residents who reside in the elite province. We use a strict inequality, assuming that the elite prefers to cooperate rather than defect, all else equal.

The requirement that the elite be in a political majority in their province requires that $N(\tau_U) \geq \mu \cdot M$, or:

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

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

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

***q**-Regime**: In the q*-Regime, $y(\tau_U; \varphi_H) > y(\tau_U; \varphi_L)$ requires:[1]

$$\{\varphi_H \cdot [g_U - s_e(q^*_H(\mu, \lambda))] - \varphi_L \cdot [g_U - s_e(q^*_L(\mu, \lambda))]\} \cdot [\mu \cdot M/N(\tau_U)] > \rho, \text{ or:}$$

[1] We are implicitly assuming that high capture is profitable so that: $\varphi_H \cdot [g_U - s_e(q^*_H(\mu,\lambda))] - \varphi_L \cdot [g_U - s_e(q^*_L(\mu, \lambda))] > 0$. This constraint places an absolute value upper bound on the majority's price elasticity of demand for assigned goods, generally no larger than 2.

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

For each pair of values of μ and q^*_H there is an associated value of λ and thus of q^*_L which then allows us to specify a value for $\mu^{min} = \mu^{min}(q^*)$. To ensure the elite politically controls its province the upper bound, μ^{max}, is defined as above. Together, the *q*-Border Constraint* is specified as:

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

This is the constraint shown in Figure 1.

Assignment Constraint

q*-*Regime: From the specifications in Table 1 and the text, the requirement that $\omega(\tau_U; \varphi_H) > \omega(U)$ implies:

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

where the LHS represents the additional expenditures needed to provide **q** under unitary governance and the RHS represents the savings in less capture by adopting unitary governance. This inequality holds when **q** is sufficiently large, defined by:

$$\mathbf{q} > \mathbf{q}^{min}(\mu; \varphi_H) = (\varphi_H \cdot \mu \cdot g_U)/[S \cdot \hat{a}(\mu; \varphi_H)].^2$$

It is also important that **q** not be too high, however. As **q** increases, the net return to capture declines for the elite and may eventually fall below the penalty costs of ρ/Elite Resident imposed when choosing high capture. The maximum value of **q** will be that value where the $\mu = \mu^{min}(\mathbf{q})$ just holds for the constitutionally chosen value of μ; see Figure 1. From the definition of $\mu = \mu^{min}(\mathbf{q})$:

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

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

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

q-Regime*: In this case we require $\omega(\tau_U; \mu, q^*_H(\mu, \lambda), \varphi_H) > \omega(U; q_U^*(\lambda))$ to ensure provinces survive the majority's decision to punish any defecting elite province. From Table 1's specifications of $\omega(\tau_U; \mu, q^*_H(\mu, \lambda), \varphi^H)$ and the text definition for $\omega(U; q_U^*(\lambda))$, this requirement reduces to:

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

where the LHS measures the difference between the consumer surplus earned by a typical majority resident under federalism with high capture when the price of assigned services is $p_H(\mu)$ and that surplus earned by

---

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

the majority resident under unitary governance when the price of a comparable service bundle is p_U.[3] Since $p_U > p_H(\mu)$, consumer surplus is greater under federalism with high capture. Because of elite capture, however, administrative federalism also imposes an income loss $\varphi_H \cdot \mu \cdot g_U$ on the average majority resident. The higher is the value of λ, the more important are assigned services to the majority, and thus the larger becomes the gain in consumer surplus from moving to federalism from unitary governance and the more likely it is that federalism is preferred. The value of λ where the inequality above just holds defines a minimal value for λ, denoted as $\lambda^{min} = \lambda^{min}(\mu)$. For each value of μ there is an associated value of q^*_H that defines the minimal q^*_H consistent with a credible elite punishment:

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

As for the **q**-Regime here too there is an upper bound on majority demanded q* consistent with a feasible federal allocation, now specified as an upper limit on λ. If the assigned services are too important, the majority demands (and can enforce) a high value of $q = q^*_H(\mu, \lambda)$ which reduces the amount of resources that can be captured by the elite when adopting strategy φ_H. Given μ and the cost of high capture, ρ, there is a value of λ for which high capture is no longer a credible choice for the elite. This maximal value for λ, denoted λ^{max}, is defined by $\mu^{max} = \mu^{min}(\lambda)$: $\lambda^{max} = \lambda^{max}(\mu)$. For each value of μ, there is a maximal q* consistent with a credible elite punishment specified as: $q^{*max}(\mu) = q^*_H(\mu, \lambda^{max}(\mu))$. Given a choice of provincial borders μ, $q^{*min}_H(\mu)$ and $q^{*max}_H(\mu)$ define the lower and upper bounds for q^*_H – and implicitly the bounds on service assignment, λ – for the q*-Assignment Constraint:

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

This is the constraint shown in Figure 1, where we simplify the notation in the Figure to be $q^{max}(\mu)$ for the upper bound and $q^{min}(\mu)$ for the lower bound.

[3] The prices are specified as the marginal cost of an additional unit q in each regime: $p_H(\mu) \equiv s_F'(q) - \varphi_H \cdot \mu \cdot s_e'(q)$ and $p_U \equiv s_U'(q)$.