

Chapter 10:

Creating Incentives for Fiscal Discipline in the New South Africa¹

Junaid Ahmad

In a recent compilation of lessons learned about decentralization, Roy Bahl (1999) highlighted three channels through which central interventions often undermine the discipline of a hard budget constraint for local governments. These are “deficit grants” to cover year-end deficits on local government budgets, direct central government coverage of shortfalls on specific items of expenditure, and bailouts on delinquent debt. All three channels involve direct or indirect fiscal assistance at the end of the year from central authorities. Repeated over time, such assistance can potentially create a regime of soft budgets, effectively undermining the fiscal resolve of sub-national governments and the efficiency gains expected from moving government closer to communities.

In the case of South Africa, all three of the channels-- deficit grants, coverage of specific expenditure shortfalls, and bailouts-- have been utilized by the central government to provide assistance to sub-national governments in both the apartheid and the present democratic eras. That South Africa, at first glance, conforms to the international experience of *how* hard-budget constraint for sub-national governments have been violated adds little new information to the vast policy literature on decentralization that has emerged over the last decade. Instead, the case of South Africa provides insights into *why* and under *what conditions* the constraint of hard-budgets may

¹ This paper benefited from the comments and suggestions made by Gunnar Eskeland.

not be binding on any tier of government in a decentralized system. In addition, South Africa's case suggests important lessons for managing the implementation of a new intergovernmental system and how to address the issue of potential budgetary game-playing between different tiers of governments during a period of policy transition.

Sections 1 and 2 summarize the evolution of the intergovernmental system in South Africa from the apartheid to the democratic era. Each section illuminates the key factors in the design of the intergovernmental system in South Africa that has created scope for a soft budget constraint at the local level. Section 3 provides a summary of the lessons that emerges from South Africa's experiences.

I. South Africa's Inter-Governmental System: The Apartheid Era

The political design of apartheid – separation on the basis of race – was entrenched into South Africa's intergovernmental system. What emerged was a *dual* structure of sub-national governments. The violation of the hard-budget constraint for sub-national governments was a direct result of the *political* underpinnings of this dual system.

The Inter-governmental System

For the white population², the apartheid leadership developed an inter-governmental system composed of a central government at the apex, four provincial administrations under it and within each province a tier of independent local

² Differences between white and black areas were not the only ones; apartheid also imposed differences between the Indian and Coloured population. For expositional economy, this focuses on the differences between the black and white population.

governments. The system was highly centralized. Major sources of revenue, representing about 85% of the total tax collection were assigned to central government. These included the personal income tax, sales tax (and later VAT), and the corporate income tax. The Provinces, on the other hand, acted as *delegated administrations* of the center and were dependent on central transfers for their functioning. Less than 5% of the expenditures implemented by the provinces were financed from own resources.

The local governments, known as white local authorities (WLA), were independent political entities. Each WLA had elected council members. In addition, WLAs were assigned revenue sources that are typical of municipal governments, including residential and commercial property taxes and user charges and were responsible for a limited set of municipal services such as water and electricity distribution, waste removal, and traffic control. Furthermore, each WLA was responsible for land planning and zoning regulations. Finally, WLAs could raise capital funds directly from the capital markets and even received central guarantees for their borrowings. Overall, this system of central government, provincial authorities and local governments for white communities covered 80% of the land for about 20% of the population.

In parallel, a separate intergovernmental system was established for the black population. The black community representing 75% of the population was forcibly herded into 10 regions covering 20% of the land that was basically economically defunct.³ The regions were given different levels of political and economic autonomy, but as artificial regions enjoyed neither. Their survival depended on annual fiscal transfers from central government of (white) South Africa that, at the height of the

apartheid system, represented about 7% of GDP. As none of these homelands were creditworthy, their access to capital markets required central guarantees. The South African central government established a public financial parastatal-- the Development Bank of Southern Africa-- to borrow from domestic capital markets and on-lend to the homelands. The financial parastatal in turn had the guarantee of central government.

The two parallel systems were in effect completely linked. The public finances of the “white” intergovernmental system were being used to keep the homelands operative. But, this “cross-subsidy”⁴ of the homelands through direct fiscal transfers and guarantees to secure loans failed to achieve the ultimate goal of apartheid: to keep the races physically separate. Economic needs of the white communities, in particular the white local authorities, required that blacks were permitted to temporarily migrate to the urban centers to provide labor for white businesses and households in the WLAs. But, in practice, the migration to the urban centers of white South Africa was permanent and forced the apartheid leadership to sanction the establishment of townships as black local authorities (BLAs). These townships were developed at some distance at the peri-urban areas of the WLAs. Like their regional counterparts, the homelands, the BLAs had neither political nor economic autonomy. Black households were not allowed to own property and by law formal and informal economic activities were banned from operating in BLAs. By definition, therefore, BLAs had no chance to develop a fiscal base. In addition, as a product of the apartheid system, the leadership of the BLAs had neither legitimacy nor accountability with communities.

These regions were known as homelands and independent countries.

⁴ In fact, as Swilling, et al (1991) point out, the apartheid system was a tax on the black community to keep the apartheid system alive and resulted in a redistribution from the low-income black-households to high-income white households. For an elaboration of the latter point see Wildasin (1996).

To keep these BLAs operating, a system of cross-subsidies similar to the ones given to the homelands had to be developed. Fiscal transfers were provided from the central authorities via the provincial administrations along with financial guarantees. In addition, WLAs provided fiscal transfers to adjacent BLAs, financed through payroll and turnover taxes. While central transfers were designed to provide operating funds, the horizontal flow funds from WLAs to BLAs were earmarked for capital expenditures. In practice, both flows eventually financed operating expenditures of BLAs.

Implications for Implementing A Hard Budget Constraint

The apartheid architecture of dual inter-governmental systems opened the door to deficit grants, or year-end grants, to cover revenue shortfalls for both the homelands and the BLAs. A straightforward interpretation would suggest that the *design* of the “cross-subsidy” scheme is the primary reason for this fiscal outcome. Ad-hoc and determined on an annual basis at the *end* of the fiscal year, the fiscal transfers were inherently prone to becoming open-ended funding mechanisms. In addition, by central design the homelands and BLAs had no source of own-revenues. They were, as a result, completely dependent on central funds with no recourse to raising any local resources to meet any year-end shortfalls. Central government could not expect fiscal effort by sub-national governments to contribute to the financing of local expenditures. Equally important, without access to own-tax instruments, BLA leadership had little incentive to reduce spending.

But, the real explanation for the deficit grants, as Iradj et al. (1999) point out, had little to do with the technical design of the fiscal system. Rather, the system was based

on a *political economy exposure* that made soft-budget constraints a foregone conclusion. The apartheid leadership was determined to sustain the apartheid architecture. As a result, explicit fiscal transfers and implicit financial transfers were used to sustain the dual inter-governmental system. For those charged with running the BLAs and homelands this was an open invitation to be profligate in their expenditures with the expectations that end-of-the-year deficits would be financed. The ad-hoc nature of the inter-governmental transfer system was therefore a symptom of the problem, not its cause. The political exposure of the center left it predictably at the mercy of the local authorities.

Not surprisingly, over the decade of the eighties into the nineties, central budget deficits increased gradually to reach a height of 8% of GDP in 1992. If the implicit liabilities of the center were added, the deficit may well have crossed into double-digit figures.⁵

While the apartheid structure represents an example of a system that creates an extreme and aberrant set of political exposures leading to soft-budget constraints, the lesson is telling. Central governments that create political expectations for which they can be held liable open their fiscal systems to abuse. In other words, no amount of technical design would have prevented the incentives of open-ended financing created by apartheid's political and economic objectives. On the other hand, in a more rational political system, the vulnerability to political exposures can be minimized with better design and management of intergovernmental systems. This is the lesson emerging from the *new* South Africa.

⁵ One must add with cynicism that the deficit was perhaps controlled ultimately by "investing" in "police and security" as a form of hierarchical control.

II. South Africa's Intergovernmental System: The New Era

The post-apartheid central government has moved in progressive stages to fundamentally alter and rationalize the intergovernmental structure. Three sets of reforms have characterized this process: *administrative amalgamation*, *fiscal and financial restructuring*, and *reform of the municipal delivery systems*. This ambitious policy agenda has resulted in a new intergovernmental system that has become an example of “best practice” for many developing and transition economies in terms of both the process of implementation and the outcome. Despite its success, the reform process has also created incentives favoring soft budget-constraint for sub-national governments. At the provincial government level, this has been a result of both political exposure and the design of the fiscal system that links provinces and central authorities. At the local government level, the budget softness is a result more of the access to financial markets through public financial parastatals.⁶ Yet, interestingly, the decentralization of economic and political responsibilities was undertaken during a period in which central government budget deficits fell from 8% of GDP in 1994 to less than 3% in 1999. South Africa's case thus provides an interesting story of how decentralization is consistent with fiscal discipline at the macro level in contrast to the apartheid system of central control that led to fiscal indiscipline.

Amalgamation

⁶ The new Constitution of South Africa has established three independent tiers of Government: central, provincial and local. Unlike many constitutions, such as the US and Indian, in South Africa the local tier is *not* a creation of the regional or provincial level of government.

The democratic process required an abolishment of racial jurisdictions. This objective was achieved through re-drawing of boundaries and administrative mergers. At the regional level, homelands and provincial administrations were merged to form new provinces. Similarly, at the local level, the BLAs and WLAs were unified to form non-racial local governments. These boundary demarcations and administrative mergers have resulted in a three-tiered governmental structure: central government, provincial government, and local governments.⁷ The local government tier, in turn, is a two-tiered system with each province being sub-divided into district councils and each district council overseeing several municipalities. In total there are nine provinces, 51 districts, and over 200 municipalities.⁸ The rights of these new tiers-- provinces, districts, and municipalities-- have been established into the new constitution as independent tiers of government.

Fiscal and Financial Restructuring

Complementing the amalgamation process, South Africa's new leadership also adopted a phased approach to the fiscal and financial restructuring of the intergovernmental system. This involved a fundamental reform of the (i) expenditure and tax assignment between different tiers of government; (ii) the fiscal transfer system and (iii) the rules about access to capital markets.

The expenditure assignments adopted followed to a large extent the principles of benefit-spillovers and redistribution. As a result, education, health and welfare – expenditure benefits of which are regional if not national in scope and the financing of

⁷ The Local Government White Paper provides a description of the different phases of this process.

⁸ Six of the districts councils are in fact metropolitan governments overseeing administrative wards

which would have large re-distributive implications – were assigned *jointly* to central and provincial governments. On the other hand, services like water and electricity distribution, whose benefits are *local* in nature, were left as the responsibilities of local governments.

In terms of tax assignment, fiscal efficiency objectives were achieved by placing the responsibilities for the administration of corporate and personal income tax and the value-added tax at the central government level. Similarly, the administration and management of property taxes were assigned to local authorities. Local authorities were also assigned payroll and turnover taxes for the financing of capital expenditure in the municipal infrastructure sector.⁹ The newly created provinces, like their old counterparts, however, were not given any own-source tax powers.¹⁰

In conjunction with the reform of the expenditure and tax assignment, Government focused on rationalizing the flow of fiscal transfers between different tiers of governments. The objectives were to manage the vertical imbalance that was created by expenditure and revenue assignments, to replace the ad-hoc, inefficient, and inequitable system of transfers that characterized the apartheid system, and to make the system more accountable. As a result, Government has introduced a new set of fiscal transfers based on clear economic principles and administered in a more transparent and accountable way (Box 1).

⁹ While good fiscal practice would suggest that these two taxes are not best suited for financing infrastructure, in reality they are imposing limited inefficiencies in the fiscal system. Both the rates of the turnover and payroll have been kept very low (less than 0.5%) and are determined centrally for the country as a whole. Finally, they are levied at the metropolitan/district level rather than at the municipal level. These factors limit the potential sources of inefficiencies including the cascading effects of turnover taxes, “race to the bottom”, and spatial dislocations

Box 1 South Africa's Fiscal Transfers to Local and Provincial Governments

Local Government

- A **municipal basic services grant** to enable poor residents in all local government jurisdictions to receive access to basic municipal services. The basic approach involves estimating the number of people in poverty (household income less than R800 per month in 1998 prices) and the current cost of providing basic services for each person.
- A **municipal institutions transfer** for those jurisdictions currently lacking the basic administrative capacity to raise their own revenue or lacking the basic infrastructure necessary to function as local authorities. This transfer pays for a minimum level of resources to provide and maintain basic facilities for the operation of local government (such as community centres and an office for elected officials).

A detailed description of the formula can be found in the 1997 framework document "*An Equitable Share of Nationally Raised Revenues for Local Government*". As the new system targets the poor, a number of municipalities will receive lower subsidies in the future than they previously enjoyed. A once-off transitional grant was introduced in 1998/99 to facilitate this adjustment.

Provincial Government:

The actual formula used by the Budget Council to determine each province's equitable share is based on the provinces' demographic and economic profiles. It consists of:

Education share, based on the average size of the school-age population and the number of learners actually enrolled (40%). **Health share**, based on the proportion of the population without private health insurance and weighted in favour of women, children and the elderly (18%). **Social security** component based on the estimate of people entitled to social security grants (elderly, disabled, children) (17%) **Basic share**, based on each province's share of the total population of the country (9%). **Backlog component** based on the distribution of capital needs as captured in the schools register of needs, the audit of hospital facilities and the share of rural population in each province (3%). **Economic output share**, based on the estimated distribution of gross geographic product (GGP) (8%). **Institutional grant**, divided equally among the provinces (5%).

Source: Iradj and Ajam (1999)

Government is now in the process of reforming the rules about access to capital markets by sub-national governments. The overall policy approach is to decentralize borrowing powers to *local* governments. To achieve this objective, Government has announced an end to the provision of central guarantees for local government borrowing from capital markets. In addition, Government is legislating an act to regulate the borrowing powers of local governments. The act will specify the disclosure of information related to local government liabilities and assets and the types of revenue

¹⁰ As argued later in this chapter, this policy decision is an important source of the soft-budget constraint incentive faced by provincial governments.

sources that can be pledged as collateral. In addition, it will specify the rules about “municipal bankruptcy.” This will include a bankruptcy process to be mediated by the judiciary rather than through a political or administrative system. The act will also specify the *minimum level* of delivery of essential services that will be required to be kept functioning with municipal taxes in the case of bankruptcy.¹¹ Finally, the act will regulate “off-budget” borrowing by making it illegal for municipalities to finance expenditure through municipal corporations or by manipulating the financing of statutory obligations such as pension funds.¹² All of these measures from full disclosure of assets and liabilities to bankruptcy procedures incorporated into legislation enable central authorities to decentralize borrowing powers while significantly reducing the potential of moral hazard of ultimately inheriting local liabilities. The legislation also entrenches by law the costs that will be imposed on local authorities, in particular the elected politicians for fiscal mismanagement.

While provincial governments have the constitutional right to access capital markets, the central government has reached a policy understanding with provincial authorities that provinces will not borrow from private sources. Given that provinces do not have access to own-taxes, this policy measure is indeed necessary. Otherwise, the authority to borrow from capital markets will be taken as a signal that financial transactions of provinces are backed by central authorities.

Institutional Reforms

¹¹ Without this clause, central government may face the problem of time inconsistency thus reducing the credibility of their policy measures *ex-ante*.

¹² See Ahmad (1998).

Complementing the amalgamation process and the fiscal and financial restructuring, the reform of the intergovernmental system also involved implementing legislation to facilitate certain institutional reforms at the sub-national level. These include the regulatory framework for privatizing and corporatizing public services, developing community delivery systems and establishing metropolitan governments in the context of the larger cities (see Box 2 for an example of metropolitan restructuring).

The corporatization and privatization of municipal services and the development of community delivery systems suggests that de-centralization policies is not only about regulating the political, economic and administrative relationship between different tiers of government. It also includes the de-centralization of responsibilities into the private sector and communities. The involvement of the private sector – market decentralization- and communities in the delivery and financing of services provides an additional mechanism for unbundling government, creating competition, and fostering incentives for better governance.

But, in setting up the involvement of the private sector in service delivery raises the issue of regulating quasi-monopolies especially in the infrastructure sector. While active bidding for concessions in, for example, the water and electricity distribution sectors enables the “competition *for* the market, ” once allocated the rights the concessionaire does not face “competition *in* the market.” As a result, a regulator may be needed. South Africa faces a choice in the design of the regulatory system. It can create a centralized system of regulations with a central regulator or allow decentralized regulation to emerge. The latter may include the mandatory publication of information on performance and thus allow communities to compare between service deliverers and

create a bottom's up pressure on local governments. Decentralized regulatory systems would also include the ability of the local governments to set up their regulatory system. This maybe the model adopted in the water sector while in the electricity sector a centralized regulator is the model that has been adopted.¹³

¹³ As elaborated later in this chapter, the choice and credibility of the regulatory institution is an important determinant in a decentralized system that includes market decentralization of whether a soft-budget regime exists for sub-national governments. A failure of the regulatory body either because of regulatory capture by the agents to be regulated or because of political interference by central government can leave central government with sub-national liabilities.

Box 2
Metropolitan Governance: The Case of Johannesburg

With municipalities fragmented along racial and economic lines into WLAs and BLAs (see earlier description), the challenge for the leadership in Johannesburg was to create a unified, non-racial system of metropolitan governance. The racial boundaries of the municipalities were first re-configured with the merger of adjacent black and white local authorities. Four new municipalities were thus created. The challenge was now to ensure that the four municipalities would coordinate to address problems of externalities, economies of scale and redistribution in the delivery of services. The political debate focused on two models of metropolitan coordination: the Toronto “mega city” model and the Minneapolis-St.Paul “risk-pooling” model. In first approach, the municipalities would be converted into administrative wards linked through a metropolitan government that would have the fiscal and delivery responsibilities. In the second model, a more “leaner” metropolitan government would link the four municipalities. This metro tier would inherit and pool the “business taxes” of the municipalities and redistribute the revenues through a formal grant system based on fiscal and poverty indicators. The pooling of the business taxes at the metropolitan level would ensure that municipalities could not enter into a “beggar-thy-neighbor” type of competition. The revenue sharing would also achieve a level playing field and form the basis of an “insurance pool” whereby all municipalities would receive the fiscal benefit of economic development regardless of the location of the economic activity in the metropolitan area. Johannesburg’s one city plan has adopted a third approach. It has followed the centralized, coordination of the mega city model by centralizing political authority, treasury management and spatial planning under one central node. But, the city is implementing delivery of services through “economic and administrative decentralization.” For goods for which some form of user charge can be levied – such as water and sanitation, waste services, electricity production and distribution – the city has transformed its line-function delivery system into corporate bodies with different forms of private sector involvement. For goods with more “public goods” characteristics, such as slum-upgrading, primary health, and libraries the city is delivering the services through decentralized administrative wards run by councilors and CEOs with a multi-year budget allocated by council. While unbundling the metropolitan government into corporations and community delivery allows greater accountability, it also raises the issue of how to manage the potential problem of moral hazard – between the metro government and the corporate and community bodies and between the metro and upper tier government -- implicit in this unbundling.

Source: Ahmad (1996) Structure of Urban Governance in South African Cities (Journal of International Taxation and Public Finance) ; Ahmad & Inman (1998) One City: A Proposal for Cape Town (Policy Brief)

Implications for a Hard Budget Constraint in the Provinces

To understand the sources of softness in the budget constraints sub-national governments in South Africa’s new intergovernmental system, one needs to assess the case of the provincial middle-tier *separately* from the situation of the local government tier. The provinces are considered first. Four factors have created the opportunity for

budget games between the provinces and central government: constitutional mandates, ambiguity in expenditure responsibility between different tiers, lack of own-taxes and centralized-labor mandates. The new Constitution of South Africa places the responsibility of financing and delivering health, education and welfare *jointly* in the hands of central and provincial governments. But, both the Constitution and policy legislation focused on the intergovernmental system are *ambiguous* about defining what is meant by “joint responsibilities” and, in particular, what aspect of this “jointness” is central versus provincial expenditure responsibility.

In addition, even though permitted by the Constitution, Provinces do not have any sources of own-taxes. The responsibility for financing the expenditure on health, education, and welfare lies solely in the hands of the center. In this case, central government has opted to provide grants directly to provinces. These are, however, not provided as conditional grants linked to the delivery of the social expenditures. Instead, the grants are provided as unconditional flows, allowing implicitly the provinces the flexibility to shift funding between expenditure line items.

The lack of own-taxes in conjunction with the ambiguity in the definition and separation of the responsibilities between the center and the provinces opens the opportunity for budgetary game-playing between the two tiers of government. There is scope for the provinces to pass the buck back to center on the grounds of constitutional mandates and lack of funding opportunities. In turn, the center has limited ability to hold provinces accountable for delivery. Without having provided any own-sources of revenues, the center is unable to “tax” the provinces to finance any shortfall and hence force the political responsibility for delivery back onto the provinces. As a result, the

provinces will, in turn, have limited incentive to spend efficiently knowing fully well that they cannot be taxed or punished for non-delivery.

This scope for budgetary game-playing is further aggravated by South Africa's system of centralized wage bargaining. The latter, in effect, locks in the expenditure commitments of the "unconditional" grants provided by the center. In fact, 60-80 percent of the funds in health, education, and welfare is already spoken for in terms of personnel costs even before reaching the provinces. Without having been part of central bargaining, provinces naturally do not have the political incentive to challenge the unions (see Box 3).

Box 3: Central collective bargaining structures

The present system of intergovernmental fiscal relations was super-imposed on existing centralised public sector bargaining structures. This creates enormous tensions. While provinces ostensibly have significant control over how they allocate their budgets across sectors, the reality is that personnel costs, the single largest component of provincial budgets, lie virtually outside their control. While in principle provinces have the right to hire and fire, in practice both in terms of the politics and the collective bargaining system this is difficult for provinces to implement. The system of centralised determination of conditions of service within the public sector, therefore, has imposed severe rigidities on provincial budgets.

There are problems surrounding:

- *Size of government:* Supernumeraries-- or excess workers-- (concentrated in a few provinces) are estimated at 60 000. The situation is compounded by the lack of an affordable retrenchment tool. There has been a moratorium on retrenchments since 1996. Instead government has relied on the Voluntary Service Package, which has not been an effective retrenchment tool. There are current efforts to renegotiate this.
- *Wage bargaining:* Government negotiators are mandated by a national Cabinet subcommittee. Provincial governments make extremely limited input into mandating procedures. Bargaining outcomes therefore do not necessarily reflect provincial needs and constraints. Annual increases in conditions of service appear in a separate vote on the national budget. Nevertheless agreements may be made which impose unanticipated costs on provinces. In some cases, due to the centralised mandating procedures, provinces may not even be aware of obligations arising from collective bargaining agreements. The *Budget Review* indicates that the vote would be devolved to provinces in 2000/01.
- *Work organisation:* The Personnel Administration Standards (PAS), which is used in the determination of public service remuneration and conditions of employment, effectively prescribe work organisation centrally and provide rigid job descriptions and is regarded as overly prescriptive in respect of promotions and employment. The Minister of Public Service Administration will soon issue a Code of Remuneration (CORE) which will replace the PAS. The CORE will contain sections dealing with prescripts (salary range, salary code, job

weight) and guidelines (nature of job, key competencies etc). The prescripts will be negotiated centrally. The guidelines will be decentralised, giving departments and provinces more latitude on making appointments and promotions.

Abedian et. al (1999)

Not surprisingly, provinces began to show significant deficits within one year of being established. By 1997/98, the total provincial budget deficit amounted to approximately 1% of GDP. These have been financed by the use of reserves and overdraft facilities from commercial banks. The willingness of the private banks to provide overdraft facilities to provinces in absence of own-taxes suggests that even the private sector perceived the deficits to be ultimately backed by central government. This was ultimately confirmed by central government's willingness to provide additional funds of the amount of 2.1 billion rands in 1997/98 to relieve provincial budgetary pressures.¹⁴ Interestingly, 1.5 billion of this amount was directed to provinces that had invoked Section 100 of the Constitution, in effect announcing a failure to perform provincial functions and entering into a financing arrangement with the center.¹⁵

Controlling the Provincial Budgets

Faced with the potential of growing deficits, central government implemented a series of hierarchical controls to manage the budgets of provincial governments. The

¹⁴ A significant portion of the payment was for pension and welfare grants. This specific expenditure has been at the center of court cases levied by individuals that have not received their welfare grants. The court has passed judgement that ultimately the center is responsible for this payment despite the fact that the Constitution broadly defines this expenditure as the joint responsibility of the center and the provinces. This expenditure responsibility best defines the ambiguity underpinning the concurrence list of the Constitution.

¹⁵ Abedian et. al point out that the application of Section 100 for two of the provinces –Eastern Cape and Kwa-Zulu Natal did not raise much debate in media or within the public at large. They suggest that this is an indication of how in new democracies the mechanisms of political accountability are weak. On the other hand, it may also be argued, that it is well accepted that Provinces are fiscally dependent on the center and that no one including markets were surprised at the use of Section 100.

policy measures varied. These included first a commitment to *multi-year budgeting* at the central level – the Medium Term Expenditure Framework (MTEF)-- that determined the levels of resources available for intergovernmental transfers over a three to five year period. In addition, central government implemented direct monitoring of provincial budgets. The latter has included an indirect management of expenditure patterns at the provincial level through *budgetary institutions* that bring together the central ministries and their counterparts at the provincial level. Ultimately, the effect of these controls and measures were to establish a set of rules that ensure an *ex-post*, balanced budget through top-down processes.

The case of South Africa clearly shows that where sub-national governments do not have access to own-taxes and where expenditure patterns are in effect pre-committed through mechanisms like wage bargaining councils, hierarchical controls will be needed to impose budgetary discipline. Such controls are even more needed if the design of the fiscal systems offers little incentive for capital markets to play their role of allocating credit in a manner which signals the credit-worthiness of the sub-national governments. The limited access to revenue instruments by provinces was a signal to capital markets that provincial debt would be backed by the center.

But the case of South Africa also suggests insights into why the hierarchical controls were binding and credible. With a relatively well-developed capital market¹⁶ and progressively open-macro economy, the threat of financial flows imposed a major

¹⁶ South Africa is unique in Africa, and also relative to its stage of economic development, in having a relatively well developed private financial system. In addition to a well-developed commercial and merchant banking system, South Africa has particularly strong contractual savings institutions –private insurance companies and pension funds. In fact, assets of insurance companies and pension funds correspond to 80% of GNP – higher than in United States and Canada – and South Africa has the highest level of insurance premium relative to GNP in the world (Vittas 1994). Local capital markets have long-term money that can be invested in municipalities and cities.

constraint on central government's fiscal behaviour. In 1996, a year after the first democratic elections and the appointment of the first ANC finance minister, capital flight put the exchange rate into a free-fall.¹⁷ Within weeks the exchange rate had depreciated by 50%. Faced with this currency crisis, the new finance minister developed and announced a fiscal stabilization program – Growth, Employment and Redistribution (GEAR)-- which publicly committed the Government to bringing down its budget deficit along a pre-announced path. Since then the ANC Government has been monitored by private markets on their commitment to this fiscal austerity plan. Indeed, one of Government's key successes has been the reduction of central government deficit to below 3% of GDP in the fiscal year 1999-00 from a height of 8 percent in the early nineties. In face of this market imposed and market monitored plan, hierarchical controls to manage the expenditures of provincial government were perceived as credible and binding.

In addition, the dominance of one political party, the ANC, and its influence on the political process has also been perceived as providing teeth to the control of the provinces by the center. Each political party at the *national* level announces their candidate for premier and then elections are held at the *provincial* level. The dominance of the ANC in seven of the nine provinces means that the national selection by national party members is in effect the determining choice of the premier. The latter then land up responding primarily to the central party's directives more than responding to community needs. Given that ANC's national priority has it committed to the principle of fiscal stability, the ANC's political control complements its budgetary control to ensure that the

¹⁷ Up until then the ANC had appointed a technocrat from the private sector to the position of the Minister of Finance.

hierarchical processes achieve their intended effect. But, ultimately, the objective of the fiscal controls – fiscal stability – was itself a result largely of market discipline created by the open-macro economy and the linkages between local and global capital markets.

A political choice – or problem of political exposure, to borrow the term from Abedian et. al. -- may be at the root of the financial crisis that emerged with the creation of the new provinces. The political negotiations for a new South Africa required the creation of a provincial tier. It allowed the rapid closure of the homelands. It also provided opposition parties -- the Nationalist Party and the Inkhata Freedom Party-- the political space to claim an identity for their ethnic constituencies in the new South Africa. To protect the interest of the Provinces in a multi-tiered government, the Constitution also created the Fiscal and Financial Commission (FFC) with reporting responsibilities to Parliament.¹⁸

Since the negotiation and adoption of the new constitution, however, there has been significant ambiguity within the political leadership of the ANC with regards to the status of Provinces. Not surprisingly, over time, the Ministry of Finance has chosen to downplay the role of the FFC, ignoring in particular the issue of provincial taxation powers recommended by the FFC. The political uncertainty is reflected, in turn, around the question of whether provinces are de-concentrated or de-centralized entities within the intergovernmental system. This political exposure may have left Provinces in a “neither fish nor fowl” status which provided the setting for the type of budgetary games that led to the under-financing of welfare expenditures in 1998-1999. But, unlike the apartheid era, the political exposure on provinces in the post-apartheid South Africa, was controlled by better design of intergovernmental fiscal rules.

Implications for a Hard Budget Constraint in the Local Governments

The case of local governments is quite different from that of provinces. The sheer number of municipalities— originally over 800 and currently amalgamated to about 300— suggest that central authorities will not be able to rely on hierarchical controls as the primary mechanism of oversight and management. Instead, Government has relied on imposing accountability and fiscal discipline through various *diversified* channels.

First, at the political level, communities can directly elect their representatives. Second, the electoral process has been strengthened by ensuring that local authorities have access to own-tax instruments, including commercial and residential property taxes. Third, local authorities have access to well defined and funded recurrent *and* capital grants programs.¹⁹ Fourth, legislation will enable local authorities to directly access capital markets with clear rules on information disclosure and a judiciary based bankruptcy process. Fifth, land markets for the commercial and industrial properties as well as the upper income households are working well enough to impose competition between the large cities. Sixth, a new financial management system is being introduced across all local authorities and will provide public information on the fiscal and financial state of each local authority on a quarterly basis.

Jointly these mechanisms ensure that communities (electoral processes and reactions to adjustments to tax rates), markets (through land markets and allocation of credit), and central processes (fiscal transfers systems, financial management requirements, and information disclosure) have limited the ability of local authorities to

¹⁸ For greater detail on the operations of the FFC, see Ahmad (1998)

¹⁹ In particular, the Consolidated Municipal Infrastructure Grant (CMIP) is the primary capital grants system through which central government finances the funding of municipal infrastructure through local government and community participation.

engage in unsustainable fiscal expansion. Not surprisingly, unlike the case of provinces, local authorities have not been able to make the center liable for local expenditure and capital markets have been less reluctant to provide them bridging finance to balance budgets.

Potential Problems

At the same time, however, there are potentially three windows within the fabric of the intergovernmental system that may offer a scope for local governments to pass on their liabilities to upper-tier governments. These are reflected as problems of metropolitan size, public financial intermediaries, and political exposure.

Metropolitan Size: Through the amalgamation process, national legislation has mandated the creation of metropolitan governments. Popularly known as “mega-cities” of which the top six represent about 40% of GDP, there is a potential of a ‘too-big-to-fail’ syndrome. The assumption here is that during financial crisis, metropolitan governments will be bailed out because of the potential macro-implications of their collapse. This potential problem is, however, partly being addressed by promoting the *economic decentralization* of cities. In other words, cities are being encouraged to decentralize municipal services into business corporations, preferably with private sector participation. This decentralization of economic services into the private sector would put the management of infrastructure into the bankruptcy laws of the private sector. In addition, private participation would bring in equity finance with the added benefit that private financiers would have an incentive to monitor debt. Finally, private participation has the potential of crowding private sector competition into the municipal infrastructure sector.

While the economic decentralization model would assist to bring in private sector discipline and competition in the financing and delivery of infrastructure, this approach would only be applicable for goods in a contestable market. The pure local public goods-- such as libraries and community facilities-- would fall outside this approach. In this case, the metropolitan governments, in particular Johannesburg, are looking at the model of *community decentralization*. The city is being carved out into ward zones, with ward councilors and an administrative CEO. The latter will be given a fixed sum of resources on an annual basis through a transparent budgeting system to be spent through community participation in the ward structure. Thus, hierarchical control of the ward administration through the budgetary system by the metro tier will be complemented by bottom-up control through community participation in the ward structure.

The financial sector in South Africa may also provide a mechanism to help central government by establishing a buffer between the local and upper tiers of governments. In particular, there are discussions being undertaken by the private sector to establish municipal bond insurance. In case of a default by a municipality, the bond insurers guarantee the timely payment of interest and principal in accordance with the issuer's original payment schedule. The insurance agency will work with the issuer to address financial problems and thereby minimize its own losses. As in the US market where over 40% of the municipal bonds are insured, the insurance guarantee is generally irrevocable, and the instruments are guaranteed for the life of the bond, regardless of what may happen to the insurer. To reduce their exposure, in turn, the insurance agencies will diversify their risks by re-insuring with specialized agencies. In the case of South

Africa, already several international re-insurance agencies are entering the local capital market.

Public Financial Parastatals: International evidence has often suggested that the presence of financial parastatals offer an off-budget mechanism for central authorities to finance public liabilities. In particular, during a time of financial crisis it may be politically convenient to seek off-budget mechanisms to “hide” the problem in the expectation that a resolution to the crisis is forthcoming if the problem can temporarily be financed. International experience also suggest that such public sector intermediation run the risk of displacing private capital and dampening the role of capital markets in the allocation of credit.

South Africa retained the Development Bank of Southern Africa, the financial parastatal that was used to fund the homelands and converted it to a municipal infrastructure bank. Intended to finance the smaller municipalities that would not be able to access the capital markets, the bank also started financing the country’s large cities. In the case of Johannesburg, the bank increased its lending to the city precisely when the private capital markets were pulling out of Johannesburg as a result of its economic mismanagement. Whether this intervention by the financial parastatal delayed the ultimate restructuring of the city (see Box 2) and whether this was done with the sanction of central government will be difficult to prove directly. But, it clearly showed the potential that public intermediation has in terms of providing a back-door to soft budget constraint.

In the case of South Africa, three factors can potentially limit the negative impact of public financial intermediation. First, the presence of private actors in the capital

market provides a political lobby force that will limit the encroachment of public intermediation. On the other hand, the same political lobby group will not hesitate to shed private risk onto public books and use a financial parastatal for this purpose. The latter problem suggests that a clean separation of roles for public and private financial institutions in the financing of public or quasi-public goods is difficult to achieve in practice. A preferable option may be to avoid public financial institutions (see Box 4).

Box 4: Evolution of Capital markets and Decentralization of Borrowing Powers

In countries without a domestic capital market, it may be preferable to have central fiscal transfers – e.g. from central borrowing from international sources – provide resources to sub-national governments while policy efforts focus on developing a private commercial banking system. This approach avoids the creation of public financial intermediaries, keeps the fiscal and financial sectors separate, and can offer a better starting point for a municipal finance system. As the private banking system establishes itself, long-term finance can be provided through centrally funded discount facilities. These facilities stretch the terms of commercial bank lending to municipalities but take on only the maturity risk. Commercial banks retain the credit risk for their retail lending, thus preserving the separation of risks between the public and private sectors. Columbia offers lessons for designing discount facilities. The need for discount facilities will disappear as capital markets emerge with financial instruments offering long-term finance. And as long as the fiscal system is well designed, sub-national governments will be able to secure longer maturities. Still, for fiscally weak municipalities the fiscal transfer system will remain an important vehicle for accessing funds. In more developed capital markets, private credit rating agencies and bond insurance agencies offer a market-based mechanism for monitoring and regulating sub-national borrowing. These institutions, in turn, require public sector regulation.

Ahmad (1999)

The second factor is the independence of the reserve or central bank. This independence in effect ensures that debt of local authorities will be difficult to monetize. International experience, especially the cases of Argentina, Brazil, and China, suggest that public debt of sub-national governments is often monetized where the central bank is under the political control of the government and where sub-national governments have access to sub-national public parastatals. In Argentina, for example, the provinces were using their sub-national public banks to finance the provincial deficits knowing that in

turn the central bank would use its powers to support sub-national public parastatals. In the case of South Africa, the tradition of independent reserve bank has been preserved through the country's massive political change in the last decade. The credit goes to the political authorities for having learned from international experience the value of preserving an independent central bank. But, the presence of a relatively well developed capital markets also ensured that there was a political force capable of lobbying for the independence of the central bank. Ultimately, the threat of capital flight as experienced in 1995, remains a strong constraint on the ability of political leadership to reduce the independence of the reserve bank. The latter was entrenched in the constitution.

In addition, during the constitutional talks, and this is the third factor, the ANC rejected the establishment of provincial banks but opted for direct sub-national borrowing from capital markets.

Political Exposure: The problem of political exposure discussed earlier may also be a potential problem in the case of local governments. In designing the new intergovernmental system, central authority has clarified the rules of the game in the new fiscal regime for the local governments. But the latter has inherited the backlog of the past system in two specific ways. First, there is a vast difference in the service delivery between whites and blacks – the legacy of apartheid. Second, in legislating an amalgamation process to create a new local government structures, the financial liabilities of the old system –debt owed to the private sector—has now been inherited by the new local governments. These new authorities are questioning the appropriateness of financing from a local tax base the restructuring process, the backlog, and the old liabilities for which the new political leaders feel that they are not accountable.

Central authorities have reacted to this problem of political exposure through several channels. The first response has been political. The central political leadership took the decision early on in the reform strategy that the financing of the restructuring process and the backlog of services would be done jointly by all three tiers of government. In effect, this created a political statement that if the responsibilities of the backlog could not be so easily transferred to the center, liabilities of local governments in the future would meet the same fate. This political approach of course was fully compatible with and driven to a large part by the macro stance of lowering deficits.

In addition, however, central authorities have made a portion of the implicit political exposure an explicit part of the budget. The *capital grants* for local authorities are meant precisely for subsidizing the local governments' efforts to redress the service inequities. Similarly, Government is in the process of implementing a *restructuring grant* for the large cities to support their reform process. The commitment to macro deficit targets has forced government to reduce the level of support on the restructuring grants. As a result, the grants have been targeted to the largest of the cities where the failure to quickly reform would have the largest negative externalities on the economy. But, more importantly, by placing a portion of the political exposure directly onto the budget, the government has constrained the fiscal threat of the political liability by pushing it onto the formal budget process.

III. Lessons

South Africa's experience so far suggests that a framework of a hard-budget constraint in a multi-tiered government cannot be imposed through one specific channel;

such a magic bullet is elusive. Instead a *systemic policy* approach in which hierarchical regulatory controls, fiscal tools, markets, and community accountability *all* are needed to provide the right incentives for *each* tier of government to be held accountable for their management of economic affairs.

In South Africa, the multi-year budgeting, explicit commitment to macro fiscal targets, and financial management regulations form the core of the hierarchical systems. Jointly these systems show how much money is available from the center and to which tier of government. The money is also delivered very transparently to each tier through a system of fiscal transfer that is legislated and formula based. It limits the scope for ad-hoc bargaining and manipulation of resources. In addition, local governments have access to own-tax instruments which means that, at the margin, this tier of government can be held responsible for its own expenditure decisions.

Clearly the power to tax forces local governments to take into account the impact on economic nodes and households. In South Africa, the ability of capital-owners to shift locations of firms between the large cities, or shift capital out of the country and the willingness of communities and households to pay, or withhold payment, are examples of how the taxing power of governments create a dynamic interface with business and households. This interface is lacking at the provincial level that is solely dependent on fiscal transfers.

But, the design of the fiscal transfers can also take advantage of community-control of public officials. First, the municipal capital grants are allocated through local governments. Second, these grants are used to finance projects that are designed and

implemented by communities. Not only does this process increase the sustainability of the specific projects, it enhances the interface between local authorities and communities.

In addition, the design of the hierarchical systems and the fiscal tools will shape the interface between the public and private sector. In the case of capital markets, fiscal policy design has influenced capital market responses to provinces and local governments very differently. In the case of the former, capital markets have provided funds most probably under the assumption that the center will be liable. In the case of local authorities, creditors have been more cautious, sensing that local governments are ultimately liable for their borrowings. The different response of the capital markets to the two different tiers of government was not a sign of market imperfection, but rather a market response to different fiscal policy. Similarly, how governments ultimately design the regulatory framework for private sector participation in infrastructure financing and provision, will determine whether the power of private sector competition can be used to minimize the fiscal exposure of government and whether private sector competition can help enforce hard budget constraints. A weak regulatory regime or one that is susceptible to capture or political interference may result in weak private sector participation or one that leaves government vulnerable to implicit contingencies.

Ultimately, the design of the inter-governmental system will provide policy makers with the scope to take advantage of the checks-and-balances offered by the judicious interface between hierarchical controls, fiscal instruments, markets, and electoral systems.

South Africa's experience suggests in particular that any policy assumption that in developing or transition economies with potentially imperfect markets and nascent

democratic regimes, hierarchical controls must play a more dominant role (or in a neat framework of policy sequencing, initially a more dominant role) may need to be qualified. The issue of “who guards the guardians” in a hierarchical system is an equally legitimate policy issue as is the concern for market imperfection or slow-responding electoral systems, existence of which are used as reasons for relying on hierarchical systems of control in the first place. In fact, given the imperfection of each of these channels the simultaneous use of all may enable a more synergetic system to emerge. At the least it diversifies the risk of central government of having to rely on one policy instrument.

The case of South Africa suggests that global financial markets in conjunction with local markets offer an important constraint on central authorities’ decision-making and makes hierarchical systems more credible and binding. Similarly, the separation of powers within the public sector is itself another mechanism for signaling commitment and credibility of hierarchical rules. South Africa’s independent judiciary and independent reserve bank are, for example, enabling policy-makers to support the decentralization of borrowing powers.

In this context, the particular role of local financial markets and how policy fosters their development and integration into global capital markets is emerging in South Africa as an important component of enabling a more effective and accountable decentralized system of governance. The South African case suggests therefore that the policy issue of how to foster an independent and private capital market system is therefore at the core of establishing a hard-budget constraint for a multi-tiered system of government. This paper suggests that countries may well be advised to avoid creating

public financial institutions, especially sub-regional banks. In addition, the South African case suggests that the design of fiscal decentralization will determine the effectiveness of capital market response to local government credit-worthiness. Finally, the paper suggests that to avoid unplanned liabilities for the central government, fiscal decentralization should *precede* financial decentralization.

However well a policy is designed, its Achilles' heel will remain the vulnerability to political economy exposures. The apartheid context and the financing of the dual government structure are extreme examples of this problem. But, South Africa's new intergovernmental system also shows this vulnerability. In terms of fiscal and financial strength, this system is "hour-glass" in shape: a strong center and a local tier, with a constrained provincial middle. But, it is through the "constrained middle" – the provinces-- that welfare, health, and education are financed and delivered. The weak fiscal, financial, and regulatory powers of the provinces are a result of the evolving political economy in which the central political forces have reduced the role of a middle tier in South Africa. Yet, the Constitution provides provinces more decentralized powers. This contrast between the original intent of the Constitution and their implementation in practice has created the space for the type of budgetary games that characterized central-provincial fiscal relations during 1995-1997. The center responded by imposing strong hierarchical rules and utilized the dominance of the ANC as a political party to keep the provinces in tow. Whether this top-down control is tenable in the long run remains to be seen. Currently there are talks of converting the grants into more conditional grants (greater hierarchical control) and of selectively devolving some of the functions to the district council or metropolitan level (enabling greater control by local authorities).

Ultimately, however, as long as the political decision about the role of the provinces remain uncertain, there will be a problem of political economy exposure.

An important element of imposing hard-budget constraints is setting the tone or the rules of the game during a *transition* period. Central government was quick to respond to the currency crisis with a fiscal plan and has remained committed to the macro-targets of the plan. Even during periods of macro-economic recessions when a Keynesian type of expansion may have been warranted, central authorities maintained their fiscal posture. Similarly, central government was quick to address the issue of provincial deficits with state-of-the-art fiscal control measures. In terms of local government, central government announced the stoppage of guarantees for local borrowing and has been even more forceful in regulating and even closing sub-national financial parastatals. With all of these measures, central government has slowly but surely gained credibility in the market that it will sustain its fiscal program and expect the same from sub-national tiers. This disciplined approach was essential in a transition economy to signal to sub-national governments that the “bailout regime” of the apartheid era is indeed over. South African policy makers were quite cognizant that how governments intervene in the short run can break the trends from the past and set the expectations for the medium term.

South Africa’s intergovernmental system continues to evolve and change at a fast pace. These changes are fundamental in nature with political, economic, fiscal and financial decentralization happening simultaneously. Yet, in spite of these rapid changes and during the process of decentralization, the country has significantly improved its macro fiscal and monetary conditions. While the reasons for this success are manifold

one key determinant has been the central government's ability to impose to a large degree a hard-budget constraint on sub-national governments. It was achieved by the judicious use of hierarchical controls, markets, and electoral voice with each channel providing a check-and-balance on the other. Ultimately, this synergetic use of the different channels is perhaps the lesson that South Africa offers other countries to avoid the so-called dangers of decentralization.

References

- Abedian I. and Ajam T. 1999. "Multi-Tiered Governments and Hard-Budget Constraint: The Case of South Africa." Draft paper, World Bank Country Office.
- Ahmad, Junaid. 1999 "Decentralizing Borrowing Powers" *PREM Notes*, Number 15, January.
- Ahmad, Junaid. 1996. "Structure of Urban Governance in South African Cities." *International Taxation and Public Finance* 3(2):193-213.
- Bahl, Roy "10 Lessons on decentralization." *Georgia State University Working Paper*.
- Dillinger, William. 1997. "Brazil's State Debt Crisis: Lessons Learned." World Bank, Latin America and the Caribbean Region, Poverty Reduction and Economic Management Unit, Washington, D.C.
- Lall, Rajiv, and Bert Hofman. 1995. "Decentralization and Government Deficits in China." In Jayanta Roy, ed., *Macroeconomic Management and Fiscal Decentralization*. Washington, D.C.:World Bank.
- Litvack, Jennie, Junaid Ahmad, and Richard Bird. 1998 "Rethinking Decentralization in Developing Countries." Sector Studies Series. World Bank, Poverty Reduction and Economic Management Network, Washington, D.C.
- RSA 1999, Local Government White Paper, Pretoria, South Africa.
- Swilling, Mark, Richard Humpries, and Khehla Subhane 1991. *Apartheid City In Transition*. Cape Town: Oxford University Press.
- Wildasin, David 1993 Local Finance of Urban Infrastructure In South Africa." AFIEI Working Paper, World Bank, Washington DC.