WP/06/204

IMF Working Paper

A Principal-Agent Theory Approach to Public Expenditure Management Systems in Developing Countries

Luc Leruth and Elisabeth Paul

INTERNATIONAL MONETARY FUND

 WP/06/204

IMF Working Paper

Fiscal Affairs Department and Offices in Europe

A Principal-Agent Theory Approach to Public Expenditure Management Systems in Developing Countries

Prepared by Luc Leruth and Elisabeth Paul[1]

Authorized for distribution by Teresa Ter-Minassian and Saleh M. Nsouli

September 2006

Abstract

A well-functioning public expenditure management (PEM) system is considered a critical pillar of government efficiency, on par with a low-distortion tax system and efficient tax administration. The paper discusses PEM systems in developing countries using an analytical framework based on principal-agent theory. This simple model can be applied to various PEM systems, and allows for comparisons between institutional settings. To illustrate this, we analyze the benefits derived from the use by the Ministry of Finance (MoF) of two control instruments; ex post audits and ex ante controls, and assess their value in terms of their ability to deter cheating. We derive a set of possible "control regimes" which can be used by the MoF. Although we illustrate the use of the model using developing countries, it is also relevant to developed economies.

JEL Classification Numbers: D73, D86, H11, H50, O12

Keywords: public expenditure management, principal-agent theory, developing countries

Author(s) E-Mail Address: lleruth@imf.org and e.paul@ulg.ac.be

---

[1] Luc Leruth was Section Chief of the Fiscal Transparency Unit of the IMF when this paper was started; Elisabeth Paul is FNRS (*Fonds National de la Recherche Scientifique)* Research Fellow at the University of Liege. The authors would like to thank their colleagues at FAD and the World Bank for very useful discussions, as well as D. Bouley, B. Chevauchez, A. Corhay, J. Diamond, I. Jelovac, A. Jousten, B. Jurion, J. Lawarée, P. Mauro, S. Nsouli, T. Prakash, C. Roehler, D. Ross, V. Tanzi, and H. Tulkens for helpful comments on earlier versions of this paper. The authors would also like to thank Ms. Dyer for her secretarial assistance. The usual disclaimer applies. Elisabeth Paul is grateful to the FNRS for financing her visit to the IMF in March-May 2005.

Contents

I. INTRODUCTION

A well-functioning public expenditure management (PEM) system is considered to be a critical pillar of government efficiency by most practitioners, who place it on par with a low-distortion tax system and an efficient tax administration. It is therefore unfortunate that there is so little economic research on the design of PEM systems, especially on the theoretical side.[2] On the empirical side, papers have generally focused on the efficiency of public expenditure in key sectors (health and education), and only a few attempts have been made to quantify the welfare losses associated with a weak PEM system. They all point to rather high economic costs. For example, a public spending tracking survey in Uganda concludes that only 13 percent of nonsalary expenditures earmarked for primary schools reached the intended beneficiaries during 1991–95. The bulk of the allocated spending was either used by public officials for purposes unrelated to education or captured for private gain (Ablo and Reinikka, 1998; Reinikka and Svensson, 2004). In Ghana, a survey concluded that 20 percent of nonwage public health expenditure and 50 percent of nonwage education expenditure reached the frontline facilities (Ye and Canagarajah, 2002). In Cameroon, one observes a "loss" of about 25 percent of the value of received goods and services with respect to the value of committed funds (République du Cameroun, 2004).

The importance of a good PEM system has come to the forefront of the debate in the context of the debt initiative for Heavily Indebted Poor Countries (HIPCs), which provides substantial debt relief from the international community while requiring eligible countries to pursue good economic policies and to make their budget more "pro-poor," using the HIPC relief for spending on priority areas of a country's poverty reduction strategy (PRS). The difficulty in tracking public expenditure has become clear during the systematic assessment of the capacity of some 25 HIPCs by international financial institutions.[3] Without getting into the details of the methodology used to assess PEM systems, it is worth mentioning that it was based on 15 (extended to 16 in 2004) benchmarks relating to the three main components of budget management.[4] The studies indicate that, while progress has been made since the initiative was launched, a majority of HIPCs still require substantial upgrading of their PEM systems to be capable of reliably tracking public spending. In particular, internal control and the production of final audited accounts are the areas in most need of strengthening.

---

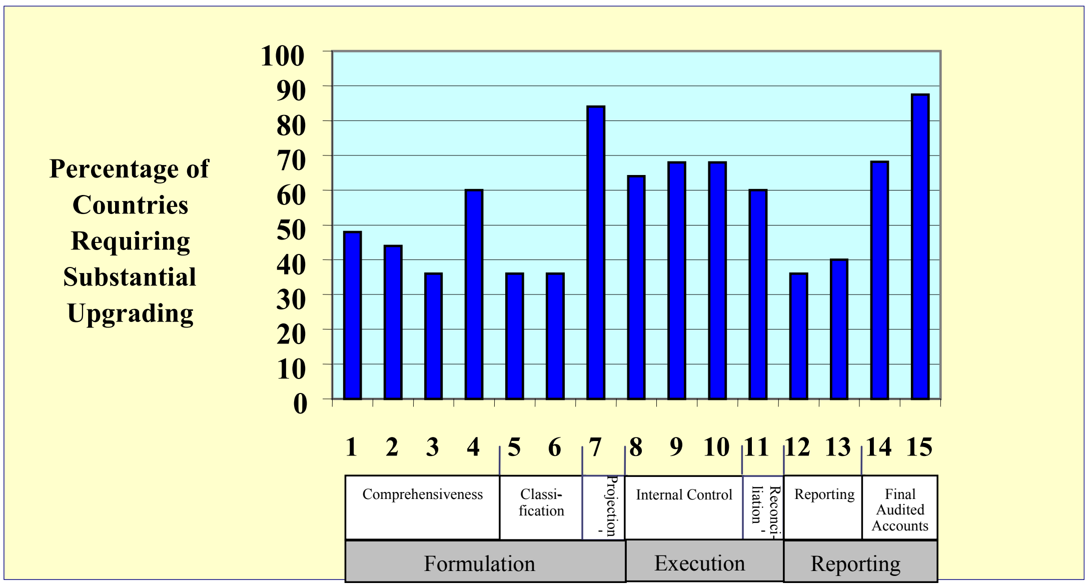
[2] However, there is a growing literature on performance, program, and output budgeting, which basically aims to improve the information available on the effectiveness of public expenditure, and hence helps improve performance through enhanced accountability.

[3] This assessment was initiated in 2002 and recently updated (see IDA/IMF, 2002, 2005).

[4] More recently, several bilateral donors and multilateral institutions have set up the *Public Expenditure and Financial Accountability* (PEFA) Program. It aims to build a strategic and collaborative approach to assessing and reforming partner countries' public expenditure, procurement, and financial accountability systems, and identifies a set of performance indicators and benchmarks, in order to help address developmental and fiduciary objectives. It has developed a Public Financial Management (PFM) performance measurement system, and assesses PFM systems against six critical objectives: budget realism; comprehensive, policy-based budget; fiscal management; information; control; accountability and transparency. The sixteen criteria are presented in Appendix I.

The problem is that the list of recommended key reforms for "getting the basics right" (Schick, 1997) is quite large and, although internal and external controls are identified as priorities, the list of priorities also covers most other areas (see Diamond, 2006). As HIPCs are, by definition, severely constrained in terms of both financial and human resources, it therefore appears critical to address the key areas of weaknesses in the most effective way possible.

Figure 1. Relative Needs for Upgrading Budgetary Systems



Source: IDA/IMF (2002).

In this paper, we use the well researched principal-agent theoretical framework to clarify the issues arising in PEM systems and help prioritize the PEM reform agenda. It has already been argued that a chain of principal-agent relationships characterizes PEM systems, which in turn raises the potential for agency problems. In the words of Tanzi (2000, p. 445), "Between their creation and their final implementation, fiscal decisions go through many stages at which mistakes, indifference, passive resistance, implicit opposition, and various forms of principal-agent problems may distort the final outcome." The author cites as examples problems occurring in formulating policies (because the behavior prescribed by the ideal role of the state may not be in the interest of the individuals who constitute the government), and problems arising between the government (as principal) and top bureaucrats (as agents), as well as between top bureaucrats (as principals) and employees (as agents).

We interpret corruption and bad governance as stemming from asymmetric information and interest divergence between those who perform tasks (the agents) and those on whose behalf tasks are performed (the principals). A rent can thus be captured by the agents at the expense of their principal.[5] The reason is that a low level of output can be due either to a low exogenous "state of nature," or to some misbehavior (such as a low effort level or corruption) by the agent. In our model, the ministry of finance (MoF) acts as the principal, providing public funds to line ministries (the LMs which are, for example, the ministry of education, the health ministry, or some other public body) to implement a set of actions. The relationship between the MoF and the LMs is an agency problem subject to asymmetric information both on some external parameters and on the actions performed. In case of low output, the MoF is not in a position to distinguish the cause, unless it uses some form of audit. The principal's problem is to design the contract that most efficiently forces the agent to meet the requirements. The contract must therefore specify a level of output (depending on the state of nature) associated with a certain level of transfer, as well as some control and sanction parameters.

The MoF has a number of instruments at its disposal to limit rent-seeking behaviors. These include, in particular, internal controls within the LM. In the so-called Francophone system of PEM (prevailing in most of French-speaking Africa), the MoF usually places some of its employees in the LMs, and their duty is to check that the operations performed by the LMs comply with the contract. In the Anglophone system (prevailing in most Anglophone African countries), the approach is different: the LM, being accountable for its performance ex post (and this is verified by the court of audit), tries to prevent non-compliance by having some of its own employees check the operations of others. In a sense, the head of the LM becomes the principal and its employees are the agents. In most cases (Anglophone, Francophone, or other systems), outcomes are verified ex post by a court of audit, in charge of analyzing the performance of the LM and reporting its findings to the relevant authorities (usually parliament).[6] If corruption is detected, the official concerned will be punished by disciplinary action or through the judicial system, sometimes entailing a hefty penalty (such as the "mise en débêt" in France).[7] These examples suggest that the literature on incentives and contract design provides the background for a potentially very useful model to analyze PEM issues and guide reforms.[8] However, most traditional models do not exactly fit the realities of PEM systems, and adjustments are needed to take into account their specific features and constraints. These will be identified and elaborated upon as the model is developed.[9]

The paper is structured as follows. We interpret a PEM system in light of the assumptions commonly found in the principal-agent literature in Section II. We then present the basic

---

[5] This interpretation is consistent with that of political and social sciences, which refer to the broader notion of "rent capture" rather than corruption.

[6] In some countries, such as Belgium and Lebanon, the court of audit may also perform some ex ante control. In this paper, we restrict our attention to the functions of ex ante versus ex post controls, irrespective of the institution performing them.

[7] The "mise en débêt" is a tool to make public accountants personally responsible for financial wrongdoings discovered in their management of public funds by the court of audit or similar body.

[8] Note that assuming a strict agency relationship between the MoF and the LM is a simplification of realities. A powerful LM could often play a very important role in budget negotiations, simply because it has more knowledge about requirements in its own area than the MoF. This could be particularly true at the time of budget preparation, where the LM could consider itself as an equal partner to the MoF.

[9] Potter and Diamond (1999) and Bouley, Fournel, and Leruth (2002) discuss several of these constraints.

features of the model in Section III. Section IV discusses ex post audits and their value in developing countries, while Section V introduces ex ante controls. We conclude the paper in Section VI.

II. Interpretation of PEM under the Principal-Agent Theory

A. The Contract

We base our analysis on standard principal-agent models involving supervision (Kofman and Lawarrée, 1993, 1996; Khalil and Lawarrée, 2003). We essentially focus on the control of LMs or assimilated bodies by the MoF, which is supposed to represent the public interest. LMs can be seen as agents of the MoF (the principal) because they are required to produce a certain level of public output—including the quality of this output—in exchange for their budget appropriation. The pair "expenditure program – budget appropriation" can be interpreted as the two components of the contract between the MoF and the LMs.[10] The objective of the MoF is to induce the LMs into implementing their expenditure programs, while the LMs pursue their own objectives. That relationship entails both hidden actions (e.g., the productive "effort" of the civil servants, possible perquisite consumption, or corruption) and hidden information (e.g., the exogenous productivity of that particular sector of the economy), with the agents having the informational advantage over the principal. Hidden information could also refer to poor program design, which would lead to inefficiency and would be difficult to dissociate from the inefficiency originating from a weak PEM system.[11]

As already indicated, a number of government operations can be assimilated to principal-agent relationships. For example, one could consider that the minister (who is the head of the ministry, but also a political appointee) heading the LM is a principal whose objective is to make sure that his agents (the civil servants) implement what he has promised to do. One could also consider that the parliament is the principal, whose objective is to make sure that the government (the executive) implements the government's program. Yet another example would be to model the central government as the principal, while the subnational governments are the agents. A paper by Ahmad, Tandberg and Zhang (2002) looks at this issue, using a principal-agent framework to analyze incentive structures that best compel local governments to truthfully reveal their ability to implement national programs. The paper focuses on the optimal contract between both levels of governments. It also insists on the need to have a multiperiod game in order to make punishments credible.

An important element of any principal-agent model is to specify an observable that will be the main element of the contract. When the agent is the LM, measuring performance should ideally be based on a mix of indicators including output, outcome, and impact. Such information is usually difficult to obtain, and although simply measuring inputs is clearly not

[10] This is in line with the approach adopted for instance in the Australian budgeting system, where so-called Service and Resource Allocation Agreements are prepared and implemented (NSW Treasury, 2000).

[11] For example, if the health system in a country does not perform well, say in terms of vaccination ratios, it can be because the health ministry focuses on other things (and could do those efficiently). It could also be because the money appropriated for the purchase of vaccines gets "lost" in the system. A weak PEM system would generally refer to the latter. See for example Gupta and Verhoeven (2000).

satisfactory, it is often the only variable for which adequate data are available. Furthermore, the level of resources in many developing countries, including the HIPCs, is such that broadening the statistics coverage can lead to the undesirable consequence of a serious degradation in the quality of data (i.e., information on inputs). Beyond the availability of data, the complexity of performing meaningful measurements, and potential biases linked to the use of performance indicators, it may also not be fully realistic to assume that the MoF has the capability to judge outcomes. Hence, we will hereafter use the term "output" in a general sense, i.e., to mean one variable that the MoF is in a position to measure, and this could include outcomes.

It also follows that we formally model a program budgeting process, although the results also apply to countries that do not explicitly use that approach.[12] In our model, LMs must make proposals on their priorities, on objectives to be reached, and on corresponding (quantifiable) targets.[13] The LM budget proposal is then negotiated with the MoF. However, we do not model the negotiation process. The contract comprises both:

- the required "output" to be produced by the LM (in terms of provision of public goods and services), and thus, implicitly, the "effort" required from the LM; and

- the LM "transfer," i.e., its budget appropriation.

A menu of possibilities can be included in the contract to take into account the general economic conditions or make relevant assumptions. For instance, it could be specified that, under a baseline scenario with realistic growth prospects, the LMs are required to operate with their existing capacities; but, under a more optimistic assumption (the country receives more debt relief, or the economy experiences higher growth), the LMs could make additional investments. In fact, this is increasingly happening in the context of PRS Papers (PRSPs), which often present a baseline scenario, plus a higher aid scenario based on the resource availability necessary to reach the MDGs.[14] We assume that, everything else being equal, the LM prefers being granted a large budget appropriation but dislikes the effort associated with the performance requirements.

[12] To quote a very practical definition from the New South Wales Treasury: "Output budgeting involves the Executive Government explicitly 'purchasing' outputs from program and service delivery agencies (the 'providers') in order to achieve desired Government outcomes [...]. With performance budgeting, the Executive Government funds (or 'purchases') an agency's program and delivery plan (a set of program and delivery strategies) which the agency has developed in order to achieved desired Government outcomes" (NSW Treasury, 2000, p. 13).

[13] Ideally, these objectives are defined in the context of a medium-term (three-year) framework and based on a comprehensive macroeconomic model. A multiyear budget framework has the potential to improve incentives, for instance by allowing the introduction of intertemporal competition across agents (Ahmad and Martinez, 2004). Although we will not address this issue in the context of the paper, it is important to note that the lack of a proper framework for medium-term budgeting has also been identified by the IMF and the World Bank as an area that requires substantial strengthening.

[14] The PRSP approach was launched in 1999 in the context of the HIPC initiative. A PRSP describes a country's macroeconomic, structural, and social policies and programs to promote growth and reduce poverty, as well as associated external financing needs. Its preparation and implementation now often condition the release of aid funds and debt relief.

B. Agency Problems

The agency problem arises from the diverging interests of the MoF and the LM, and the latter's informational advantage, both on its own actions and on the current state of nature. As standard in the principal-agent literature, the agent's effort is a necessary component of the production function—but entails some disutility. The agent may take unfair advantage of its superior information: if external conditions are favorable, the LM could exert little effort and produce a low output, while claiming that this low output is due to unfavorable external conditions. The MoF is not in a position to disentangle the two factors unless it uses some form of audit or supervision. There is thus a risk that the LM captures some rent at the expense of the MoF.[15] In the principal-agent literature, this cheating rent generally stems from lowering the level of effort vis-à-vis the compensation received. Rents, and possible reductions in public output, compared to what is economically efficient, constitute the agency costs.

In this paper, we broaden the interpretation of corruption (generally referred to as the abuse of public office for private benefits) to include misgovernance stemming from the abuse of some information asymmetry. We consider the LM's effort in terms of a combination of factors that can be good and bad, including, on the one (good) hand, an efficient and equitable allocation of resources, fiscal transparency measures, and quality of services provided; and, on the other (bad) hand, corruption, consumption of perquisites, mismanagement, and nepotism in the choice of staff or suppliers. This allows us to interpret the cheating rent, not only in terms of reduced disutility from "productive" effort, but also as corruption or misgovernance. For example, if the state of nature is high (say, favorable weather conditions), the LM could allocate some resources to unproductive areas or divert monies, if it thinks that the MOF could be led to believe that the state of nature was low. In such cases, rent capture takes place and is possible because of the information asymmetry between the principal and its agent. This interpretation enables us to link our approach with the empirical literature on corruption. Indeed, the latter identifies various factors contributing to corruption, including the overall level of potential benefits from corrupt behaviour, the cost of bribery (including penalties and sanctions), and the bargaining power and extent of discretionary powers of the various actors (Chand and Moene, 1999). Moreover, while cheating (exerting a lower effort) is probably costless for the agent, we argue in this paper that cheating, in the sense of being corrupt, may entail some costs to be concealed. This enables us to make the link with the literature on collusion in organizations (e.g., Tirole, 1986), and in Section V, we interpret ex ante controls by the MoF as increasing the cost of cheating for the agent.

As already stated, the MoF has a number of instruments and strategies at its disposal to limit agency problems. First, it can use incentive schemes, designed solely on observable information, and promise to grant the LM a transfer equivalent to the sum of a suitable compensation for the LM's effort and an informational rent (which depends on incentive compatibility constraints) in case of high productivity.[16] If such a contract exists, it prevents the LM from exerting little effort—but at the expense (for the MoF) of a loss equivalent to the

[15] Hereafter we use the term "*informational* rent" when referring to the supplementary premium that the LM is deliberately granted as an incentive to exert high effort. We use the term "*cheating* rent" to refer to the amount illegally diverted, notably through corruption.

[16] In this paper, we use indifferently performance payment (contingent on observable/verifiable results) and informational rent, although the latter is, in principle, more general (the difference between the expected utility of an agent with private information and his reservation utility).

informational rent, in addition to a distortion created by requiring a lower level of effort in some occurrences of the productivity factor. Although commonly applied to models of the corporate world (e.g., granting board members bonuses or shares), this strategy is not always directly applicable in the public sector.[17] Alternatively, the MoF can supervise the LM using a number of instruments and can threaten it with appropriate sanctions if cheating is detected. The design of the appropriate control system must take a number of factors into account, for instance the choice between ex ante and ex post (or internal and external) controls, the type of variables to be monitored (input versus result indicators), and the choice between systematic or random audits. In our model, there are two unobservable variables (effort and state of nature). Supervision could thus turn either to the exogenous productivity factor, from the observation of which the agent's behavior could be inferred (this relates for instance to public sector reforms aimed at improving the economic statistical data collection, or to audits targeted at assessing the program design), or directly to the agent's effort. In this paper, we assume that the MoF will audit the LM's effort.[18]

C. Timing

The MoF and the LMs act in a continued relationship framework, because a new budget is prepared every year. However, we consider that the external productivity occurrences are independent, so that the MoF cannot infer the state of nature—and thus the LM's effort—from what happened in previous years. This allows us to simplify the problem and limit ourselves to the study of the optimal static contract. Nevertheless, the actual repetition of the game allows the MoF to make credible commitments.

We assume that the timing of the PEM process is as follows:

(1) After negotiations (not modeled here—but note that the contract guarantees the LM its reservation utility), the LM accepts the contract proposed by the MoF; the contract is a menu of {output and implicit effort[19]/budgetary allocation} pairs, dependent on the external productivity occurrence.
(2) Nature chooses the level of the external productivity parameter, and only the LM observes it.
(3) The LM chooses its level of effort and starts producing the output; in the meantime, the MoF can exert ex ante controls.
(4) The transfer is paid through warrants (if the LM plans to reach a high level of output, it can justify it and call for additional credits).
(5) The MoF collects the output, which is publicly observable at the end of the budgetary year.

---

[17] Incentive schemes may be used in public companies (see, for instance, the regulation theory following Laffont and Tirole, 1993) and also, to some extent, in customs administrations (on the theoretical side, see, for instance, Besley and McLaren, 1993).

[18] A model close to ours, which combines adverse selection and moral hazard, predicts that monitoring the agent's action is strictly preferable to auditing private information (Kessler, 2000).

[19] As we shall see, the effort level required from the LM should be optimal with respect to the contracted output. It is thus implicitly defined. This problem is a case of "false moral hazard" where the observation of the variable does not allow to perfectly disentangle the agent's "type" and effort level (Laffont and Martimort, 2002).

(6) The MoF can order an audit if it suspects that cheating took place (or simply because it has committed to make these audits for long-term credibility purposes).
(7) The LM (or its program manager) is imposed a penalty if caught cheating.

III. THE BASIC MODEL

A. Main Assumptions

The model developed here is close to the literature on supervision as in Kofman and Lawarrée (1993) and Khalil and Lawarrée (2003). However, it differs in some assumptions so as to better reflect the features of PEM systems. The added value of this paper lies in the practical applications to PEM of the analysis, but also in those differences.

We model the agency relationship between the MoF and one LM, both assumed risk-neutral.[20] The LM produces a level of output x, which depends on two variables: a random exogenous productivity factor, θ; and the LM actions or effort, e, such that $x=\alpha(\theta,e)$; with $\alpha_e>0$; $\alpha_{ee}<0$; $\alpha_\theta>0$. The realized output is public knowledge, but e and θ are the LM's private information. The external productivity can be either high or low: θ_i with $i\in\{H,L\}$ and $\Delta\theta=\theta_H-\theta_L>0$. We also assume $\alpha(\theta_H,e)>\alpha(\theta_L,e)$; and $\alpha_e(\theta_H,e)>\alpha_e(\theta_L,e)>0$. It is common knowledge that the MoF assigns ex ante probability q to the event that $i=H$ (and probability $(1-q)$ to the event that $i=L$). When state i occurs, the LM exerts a certain effort level e_i, thus producing an output $x_i=\alpha(\theta_i,e_i)$.

The monetary equivalent of the LM's disutility from effort is represented by an increasing and strictly convex function $\psi(e)$, with $\psi_e>0$ and $\psi_{ee}>0$. To obtain strictly positive but bounded optimal efforts, we also assume that $\alpha(\theta,0)=0$; $\psi(0)=0$; $\lim_{e\to0}\psi_e(e)=0$; $\lim_{e\to0}\alpha_e(\theta,e)=\infty$; $\lim_{e\to\infty}\psi_e(e)=\infty$; and $\lim_{e\to\infty}\alpha_e(\theta,e)=0$. The LM's utility is given by $u=t-\psi(e)$, where t is the transfer (appropriation) it receives, and its reservation utility is normalized at zero. We also assume that $\Delta\theta$ is large enough so that the MoF is always better off in case of high output: $x_H-t_H>x_L-t_L$.

B. Perfect Information Benchmark

The MoF's problem is to choose the levels of effort required from, and transfers to be made to, the LM for each occurrence of the random factor, so as to maximize the expected output:

[20] In practice, the MoF would try to maximize the joint output of several LMs. By restricting the model to one LM, we assume the MoF treats all LMs equally. The case of several LMs is indirectly handled when the probability of audit is below one (notably in the mixed strategy equilibrium), which may be interpreted as follows: the MoF can possibly audit only a certain number of LMs, and each LM chooses its cheating level considering that probability of being audited.

$$\underset{e_H,e_L,t_H,t_L}{Max}\ E(X) = q\left[\alpha(\theta_H, e_H) - t_H\right] + (1-q)\left[\alpha(\theta_L, e_L) - t_L\right] \qquad (P)$$

Subject to the LM's individual rationality *(IR)* or participation constraints under each occurrence:

$$t_H - \psi(e_H) \geq 0 \qquad IR(H)$$

$$t_L - \psi(e_L) \geq 0 \qquad IR(L)$$

Under perfect information, the MoF equates the LM's marginal cost of effort with the marginal value of its product: $\alpha_e\left(\theta_i, e_i^*\right) = \psi_e\left(e_i^*\right)$, with $i \in \{H, L\}$.[21] The transfers are such that both participation constraints are binding: $t_i^* = \psi\left(e_i^*\right)$. The MoF can therefore enforce first-best, efficient efforts, and the LM gets no rent. Note too that, according to our assumptions, $e_H^* > e_L^*$.[22]

C. Second-Best Solutions

Under imperfect information, effort is not observable. It cannot be directly enforced, but must be indirectly induced. Therefore, the MoF must provide the right incentives so that the LM produces the highest possible level of effort. The reason is that when one productivity level takes place, the LM could cheat by adjusting its effort so as to produce the output corresponding to the other productivity level. We define the effort level $\tilde{e}_L$ such that $\alpha\left(\theta_H, \tilde{e}_L\right) = \alpha(\theta_L, e_L)$.[23] This means that, if $i = H$, the LM could exert a low effort $\tilde{e}_L$, thus produce x_L, while claiming to receive t_L. The "cheating rent" that the LM can get in that case is thus equal to $\left[t_L - \psi\left(\tilde{e}_L\right)\right] - \left[t_H - \psi(e_H)\right] > 0$. To ensure that it is never optimal for the MoF to shut down the contract in case of low productivity, we also assume $(1-q)\alpha(\theta_L, e_L) > \psi(e_L) - q\psi\left(\tilde{e}_L\right)$.

Traditional principal-agent models rely on incentive-compatible schemes to prevent such cheating. The well-known results from this literature apply and can be summarized as follows:

At equilibrium, the LM does not cheat and gets no rent when $i = L$. In order to induce it into exerting the right effort level when $i = H$, the LM must receive an informational rent equal to

---

[21] Superscript "*" stands for first-best values.

[22] This could be interpreted as an absorptive capacity constraint: when the state of nature is high, the LM must work harder to absorb a larger appropriation.

[23] To be complete, we should also define $\tilde{e}_H$ such that $\alpha\left(\theta_L, \tilde{e}_H\right) = \alpha(\theta_H, e_H)$, i.e., the extra effort level the LM would need to exert so as to produce x_H when $i = L$. However, this is never profitable for the LM, and is little relevant to our analysis.

$\psi\left(e_L^{SB}\right)-\psi\left(\tilde{e}_L^{SB}\right)$ (where superscript "SB" stands for "second best") in addition to its first-best transfer, where $\tilde{e}_L^{SB}$ is such that $\alpha\left(\theta_H,\tilde{e}_L^{SB}\right)=\alpha\left(\theta_L,e_L^{SB}\right)$. While production is efficient when $i=H$, the LM underproduces (compared to the full information benchmark) in case of low productivity: $e_L^{SB}<e_L^*$. Requiring a lower level of effort when $i=L$ enables the principal to decrease the informational rent granted to the LM when $i=H$. This reflects the "rent extraction/economic efficiency" trade-off, which characterizes adverse selection problems.

If we define the agency cost incurred by the MoF as the difference in expected output due to the information asymmetry, this is equal to:

$$AC^{SB}=E\left(X^*\right)-E\left(X^{SB}\right)=q\left[\psi\left(e_L^{SB}\right)-\psi\left(\tilde{e}_L^{SB}\right)\right]+(1-q)\left\{\left[\alpha\left(\theta_L,e_L^*\right)-\alpha\left(\theta_L,e_L^{SB}\right)\right]-\left[\psi\left(e_L^*\right)-\psi\left(e_L^{SB}\right)\right]\right\}$$

The two main components of this expression are thus the rent granted to the LM when $i=H$, and the productive distortion occurring when $i=L$ (minus the difference in compensation due to distortion).

D. Introducing Supervision

In order to avoid forgoing the informational rent, the principal can hire a supervisor and reduce the information asymmetry. Usually, this is combined with the threat of penalty if cheating is detected. In introducing supervision, we will, in some respects, diverge from the existing literature on principal-agent, so as to better reflect PEM concerns.

In the context of PEM, supervision may take various forms. One can distinguish internal controls (e.g., MoF or LM agents responsible for ensuring that expenditures and procurement are performed according to the rules), and external controls (e.g., a court of audit reporting to parliament). Controls may take place ex ante, (e.g., comptrollers issuing visas to allow expenditure, or automatic safeguards preventing LMs from exceeding budget appropriations),[24] or ex post (e.g., auditors checking the reliability of fiscal data or the performance of public spending). Different types of controls can be combined. For instance, going back to the example of PEM systems in Africa, it is worth noting that the so-called Francophone system rests on the principle of separation between the person who initiates spending (the *ordonnateur*) and the person who pays it (the accountant or *comptable*). The system relies on centralized ex ante controls from the MoF, which take place at various stages of the expenditure process and mainly focus on the conformity of spending with regard to procedures and budget appropriation. Anglophone countries, on the other hand, have inherited a decentralized management system, where the LMs' accounting officers are responsible for budget execution. Ex ante expenditure control is mainly exercised by the issue of periodic warrants by the MoF (cash management). The Anglophone system relies on independent, ex post controls by an auditor-general. In practice, however, notwithstanding those conceptual differences and institutional arrangements, both systems have proven to perform poorly (Bouley, Fournel, and Leruth, 2002; Moussa, 2004; and Lienert, 2003).

---

[24] For example, if a government is not allowed to use an overdraft facility with the Central Bank, the rule can be interpreted as a safeguard.

In the next two sections, we introduce two types of supervision. We first study the case of ex post audits (Section IV), and explain how a standard principal-agent model with audit may be of interest for the design of PEM systems. We then move to ex ante controls (Section V). In doing so, we assume that ex ante controls increase the cost of cheating for the LM, as is done in the literature on collusion. Finally, at the end of Section V, we propose an integrated framework which combines both types of controls.

IV. Ex Post Audits

In this section, we introduce an ex post auditor, costing z whether or not it identifies cheating.[25] It could be an external or an internal auditor, in which case the audit cost may be interpreted as the opportunity cost of using MoF resources for controlling the LM, instead of doing other tasks. The auditor observes an imperfect signal on the LM's effort (for instance, this may be done through a review of accounts to check if there has been corruption). We assume that the signal can take two values: "has complied" or "has cheated." The latter occurs only when the LM has indeed cheated, while the signal can report compliance by mistake. The monitoring function is such that σ denotes the probability of detecting actual cheating, in which case the LM is imposed an exogenous penalty P. With the introduction of supervision, the contract specifies not only the transfers and expected outputs (and thus, implicitly, the expected effort levels), but also the probability of audit. We assume that the MoF commits to audit with probability γ after x_L has been observed. When productivity is high, the LM may cheat with probability m. Given that output is low, the probability that the LM has cheated can be written as $\phi = qm/\left[(1-q)+qm\right]$. We also assume that the auditor is honest and does not collude with the LM.[26] An appealing interpretation is that developing countries are often subject to donors' external auditors, which are supposed to be honest (or not in a position to negotiate with the LM).

Two of our assumptions deserve further discussion. First, while principal-agent models typically assume that the level of penalty is a decision parameter, we assume that it is fixed (for instance, legally bounded and difficult to change).[27] We feel that, although a comparative static analysis can be done on the level of the penalty, in reality, social factors often limit the extent to which a penalty can be imposed in developing countries, where informal rules are predominant and society tolerates a certain amount of rent capture by public servants (see

[25] Assuming the auditor is paid only when reporting cheating is not relevant in a PEM system, although there are many instances where a bonus is given when cheating is detected (for example, customs employees detecting a fraud have a right to a portion of the tax recovered in many countries).

[26] Note that this assumption is also made by Kofman and Lawarrée (1993) in their regime with one costly truthful auditor, as well as in Khalil (1997), for example.

[27] The maximum deterrence principle (Baron and Besanko, 1984), which implies that the principal would choose the upper bound penalty level, does not always apply, especially in models where the agent may be punished by mistake and thus must be compensated ex ante for that risk (e.g., in some regimes of Kofman and Lawarrée, 1993). While this is irrelevant in our framework, we could have a situation (dubious and unlikely) where the principal's benefit from the penalty is higher than the audit cost, so that it is preferable for the MoF to allow cheating than to deter it. Finally, the traditional result stating that with infinite penalty, cheating is deterred and the first-best solution is reached, could apply but, as we argue later, it is not of practical relevance for our purpose.

Dabla-Norris and Paul, 2006). It is not surprising therefore that most sub-Saharan African countries are characterized by a patrimonial state, lack of accountability, ineffective control procedures, and endemic corruption (e.g., Bayart, 1993; Dia, 1994; World Bank, 1997; Mbaku, 1998). This is due to various causes, including that "[formal institutions'] ineffectiveness is compounded by the absence of the rule-of-law as a third-party enforcement mechanism, engendering a climate in which enforcing formal contracts and procedures becomes costly, often prohibitively so, and sanctions for deviant behavior are devoid of credibility" (Dia, 1996, p. 1).

Second, a crucial assumption used in the principal-agent literature is whether or not the principal can credibly commit to a given probability of audit at the time of offering the contract. We use this assumption in subsections A and B below. We relax it in subsection C, however, in order to capture the situation in institutional settings where the decision to audit is only taken ex post by the MoF. Hence, we discuss three possible regimes that the MoF can implement:[28] (i) the "cheating-proof" regime, which corresponds to the optimal, incentive-compatible contract when commitment is credible; (ii) the "cheating-inducing" regime, which is not optimal but, as we argue, matches the situation in some developing countries; and (iii) a "no commitment" case, in which there is no formal commitment to audit at the time of offering the contract, which results in a mixed strategy equilibrium. In subsection D, we discuss some applications of the model in terms of PEM and compare these regimes on the basis of their relative costs and benefits.

A. The Cheating-Proof Regime

Traditional principal-agent models with credible commitment use the revelation principle to determine the optimal contract. The revelation principle asserts that, to find the optimal payoff of a problem with asymmetric information, one can, without loss of generality, restrict it to the incentive-compatible, individually rational scheme where every agent truthfully reveals his private information. To put it simply, this means that the principal can do no better than offer an incentive-compatible contract, which therefore *deters* cheating. Under these circumstances, the penalty is never imposed at equilibrium, but its existence out of the equilibrium path acts as a deterring threat and prevents cheating.[29]

This approach is only applicable to settings in which the principal is able to credibly commit to any outcome of the contract. In a PEM system, the existence of a court of audit may be viewed as a commitment tool, enabling the MoF to make the credible commitment, at the time of offering the contract, that it will audit the LM at the end of the fiscal year with a given probability, which can be either probability one (systematic audit), or below one (random audit).[30] Under this regime, there are conditions where the audit threat is such that it prevents the LMs from cheating.

---

[28] We use a terminology similar to that of Eskeland and Thiele (1999) who refer to collusion-proof and collusion-inducing regimes, as we refer to cheating-proof and cheating-inducing regimes.

[29] This may seem a little remote from reality. However, the revelation principle characterizes the optimal payoffs and can be seen as the "truth-telling map" of a complex mechanism where cheating and punishments occur (Kofman and Lawarrée, 1993, p. 648).

[30] If there are several LMs, the latter situation could correspond to the case where the MoF announces it will audit a certain number of LMs—so that each LM knows, ex ante, with which probability it will be audited at the end of the year.

Formally, the MoF's problem is to choose the levels of transfers, required efforts, and audit probability so as to maximize the expected output:[31]

$$\underset{e_H, e_L, t_H, t_L, \gamma}{Max} E(X) = q\left[\alpha(\theta_H, e_H) - t_H\right] + (1-q)\left[\alpha(\theta_L, e_L) - t_L - \gamma z\right]$$

Subject to *IR(L)*, *IR(H),* and the following incentive compatibility *(IC)* constraint:[32]

$$t_H - \psi(e_H) \geq t_L - \psi(\tilde{e}_L) - \gamma\sigma P \qquad IC(H)$$

Note that the term $\gamma\sigma P$ relaxes the *IC(H)* constraint, compared to the second-best case. Hence, provided its cost is not too high, we obtain the intuitive result that audit benefits the principal.

As in Kofman and Lawarrée (1993), we find that the optimal contract exhibits qualitatively different types according to the value of the parameters (although the specific thresholds differ as we have slightly different assumptions; see Appendix II for proof and complete results). These types are characterized by different rents, productive distortions, and audit probabilities. The shift from one type to the other rests on a comparison between the expected benefits from audit (the penalty and reduction of rents and distortions) and its cost, with both the former and the latter depending on exogenous (country-specific) parameters.

- If $q\sigma P < (1-q)z$, audit is too costly relative to its benefits: it is therefore not optimal, and the principal can do no better than offer a second-best contract (by the revelation principle; see, e.g., Baron and Myerson, 1982).
- When the benefits from audit increase, there is a point where they exactly offset its cost: $q\sigma P = (1-q)z$. The MoF is then indifferent between auditing and not auditing.
- When the benefits from audit exceed its cost, so that $q\sigma P > (1-q)z$, it is profitable to audit.

Some general results apply to all three cases. When the state of nature is $i = H$, production is always efficient, but the LM is granted an informational rent. When $i = L$, there is no rent, but the LM underproduces (compared to the first-best). *IC(H)* is binding (i.e., the MoF should not increase incentives above the point when it can deter cheating), and the issue is precisely to fulfil the *IC(H)* constraint at the lowest cost. Its shadow cost (the Lagrange multiplier) is

---

[31] The probability of cheating and the expected penalty do not enter the principal's objective function as, under this regime, cheating is deterred. However, the expected penalty appears in the *IC* constraint.

[32] To be complete, one should also introduce an *IC(L)* constraint, aimed at preventing the LM from producing the high output when productivity is low. This would take the form $t_L - \psi(e_L) \geq t_H - \psi(\tilde{e}_H)$. But as is common in the literature, that constraint is redundant with the others and is therefore not relevant. Note also that, as we use the revelation principle and deter cheating, the optimal solution exhibits no cheating. The penalty is not collected in equilibrium and hence it does not enter the MoF's objective function (although it is present in the constraints).

thus a crucial variable in determining the optimal institutional setting (we develop that argument in Section V).[33]

When audit is efficient (last bullet), the optimal probability of audit, rents, and productive distortions vary and three sub-types can be identified:[34]

- *"Rent extraction" (RE) scheme:* the MoF audits with probability 1, the production distortion when $i = L$ corresponds to the second-best solution, and the LM gets a rent when $i = H$ (the rent is equal to that of the second-best solution reduced by the expected penalty);
- *"Effort adjustment" (EA) scheme:* the MoF still audits with probability 1, but the LM no longer gets a rent, and the productive distortion is lower compared to the second-best solution (i.e., the effort level lies between the second-best and the first-best solutions);
- *"Random audit" (RA):* the MoF decreases the probability of audit (i.e., $0 < \gamma < 1$) and still deters cheating, while the productive distortion is smaller than at the second-best.[35]

As the general specifications of the functions do not allow us, for instance, to compute the exact boundaries between the regimes identified above, we illustrate these results using a numerical example such that $x = \theta^2 e^{1/2}$ and $\psi(e) = e^3/3$. We can show that the EA and RA productive distortions when $i = L$ are smaller than at the second-best ($e_L^{EA} > e_L^{SB}$ and $e_L^{RA} > e_L^{SB}$), and characterize the boundaries between the regimes (see Appendix II). In particular, if we define

$$A \equiv \frac{\left[(1-q)(\theta_L)^2\right]^{6/5}\left[1-(\theta_L/\theta_H)^{12}\right]}{3\left\{2\left[1-q(\theta_L/\theta_H)^{12}\right]\right\}^{6/5}} \text{ and } B \equiv \frac{(\theta_L)^{12/5}\left[1-(\theta_L/\theta_H)^{12}\right]}{3\left\{2\left[1+\frac{z}{\sigma P}\left[1-(\theta_L/\theta_H)^{12}\right]\right]\right\}^{6/5}}$$

we show that the RE scheme takes place when $\sigma P \le A$, the EA scheme takes place when $A < \sigma P \le B$, and the RA scheme takes place when $\sigma P > B$.

In other words, if it is efficient, audit reduces the agency costs (distortions and rents) associated with the second-best contract—while still deterring cheating. When the expected penalty is relatively low ($\sigma P \le A$ in our numerical example), the MoF must use concomitantly other incentives to prevent cheating. For example, our model suggests that the LM could get a rent in case $i = H$, which would then lead to distorting production when

[33] Note that, from the specification of that Lagrange multiplier (see Appendix II), one observes that the higher the cost of audit and/or the probability of low productivity, and the lower the penalty, the harder it is to deter cheating.

[34] Our results are consistent with the analysis of Kofman and Lawarrée (1993) with a truthful auditor. There are some slight differences however: (i) we assume the principal cannot earn a penalty out of a complying agent; (ii) we assume a fixed penalty; (iii) our specification of the production and disutility of effort functions differ, which hampers tractability. The labels also follow Kofman and Lawarrée (1993).

[35] The random audit case may be interpreted in a context with several LMs, where γ represents the probability, for each LM, to be audited, in the framework of the general auditing policy of the MoF.

$i = L$. Nevertheless, audit enables the MoF to reduce the rent granted to the LM compared to the second-best level (hence the label "rent extraction"). When the expected penalty rises ($A < \sigma P \leq B$ in our example), the threat increases for the LM so that the MoF can reduce the degree of mobilization of other incentives. In our model, the LM can no longer get a rent, and so the productive distortion can be reduced. Audit thus increases the effort requested from the LM in case of low productivity (hence the label "effort adjustment"). Also, when the expected penalty is large ($\sigma P > B$ in our example), the MoF can reduce the probability (thus the cost) of auditing (while still deterring cheating). Finally, there remains a production distortion when $i = L$, reflecting the trade-off between efficiency and the cost of audit. Box 1 below summarizes our findings (using the numerical example).

As already mentioned, these results show that, when the expected penalty is relatively small, the MoF must concomitantly use other incentive tools, such as informational rents, in addition to audits to be able to prevent cheating. As the penalty threat increases, other incentives become less necessary. When the penalty is very high, even if the MoF reduces the probability of audit, the LM will not find it profitable to try to cheat. In developed countries, like France for instance, this principle is at the root of the "sampling" of expenditures and agencies to be audited.

Finally, note that this model may also be adapted to take into account other constraints faced by the MoF. For example, we could think of a situation where the MoF is obliged to comply with some minimum requirements in terms of output (for example, the "Education for All Initiative," or the provision of some basic health care package), so that it cannot tolerate that the LM produces below x_L^*. The MoF would still have to fulfill *IC(H)* at the lowest cost (considering the additional constraint), which could be done through a trade-off between granting the LM a higher rent when $i = H$, or raising the audit probability (if it is possible).

Box 1. Features of the Optimal, Cheating-Proof Contract

- If $q\sigma P < (1-q)z$, audit is not efficient ($\gamma = 0$) and the second-best contract is chosen.
- If $q\sigma P = (1-q)z$, the MoF is indifferent between auditing and not auditing.
- If $q\sigma P > (1-q)z$, audit is efficient ($\gamma > 0$) and the optimal contract may exhibit different types:
  - When $\sigma P \leq A$, the RE scheme is chosen: $\gamma = 1$, $e_L = e_L^{SB}$ and the LM gets a rent when $i = H$.
  - When $A < \sigma P \leq B$, the EA scheme is chosen: $\gamma = 1$, $e_L^{EA} > e_L^{SB}$ and the LM gets no rent.
  - When $\sigma P > B$, the RA scheme is chosen: $0 < \gamma < 1$, $e_L^{RA} > e_L^{EA} > e_L^{SB}$ and the LM gets no rent.

B. The Cheating-Inducing Regime

When audit is not optimal, we have seen that the MoF should offer the second-best contract, which is characterized by an informational rent paid to the LM when the state of nature is high. However, in reality, there may be circumstances preventing the MoF from offering that contract. For instance, the MoF may not be in a position to grant an informational rent when the budgeting system is input-based (line-item), or if it is confronted to a tight cash constraint. Besides, one fundamental difference between private sector operations (which have typically been used to illustrate the principal-agent theory) and public sector operations is often that output is easier to quantify when it is sold on the market. Public sector output often is not easy to quantify and must then be estimated at high cost and with uncertainty (how many children have actually learned to read and write, how many have been vaccinated).

There may also be political pressure or legal constraints forcing the MoF to use ex post controls by the court of audit, even if it is not efficient. For instance, nearly all sub-Saharan African countries possess a court of audit. Yet, findings from Country Financial Accountability Assessments (CFAAs) show that these institutions often suffer from important weaknesses, ranging from lack of independence to poor capacity (World Bank, 2004b), which impairs their ability to deter cheating in a significant manner. The solution recommended by the principal-agent theory is to grant LMs premiums so as to induce them not to cheat. But this solution is hardly (if at all) observed in reality. Rather, one often observes the coexistence of a weak supreme audit institution and high levels of cheating (including corruption), with few positive incentives such as performance premiums.

To look into cases where there exists a court of audit, which is not effective in deterring cheating, we now introduce a "cheating-inducing" regime: the MoF uses audit, but is not able to deter cheating. We are aware that this regime is not optimal for the principal with respect to the constraints usually considered in the literature, as it could get the same output without incurring the cost of audit. However, we believe that it adequately reflects the situation of some developing countries (see subsection D on that issue). Therefore, we consider the case in which the MoF audits, but the penalty and audit probability are such that $t_H - \psi\left(e_H\right) < t_L - \psi\left(\tilde{e}_L\right) - \gamma\sigma P$. In that case, it is always in the interest of the LM to cheat when feasible, and the actual output is always low. As audit takes place, penalties will be imposed at equilibrium, but they will not be a threat sufficient to deter cheating. To come back to the model, the cheating-inducing regime takes place when the MoF commits to audit with a probability $\gamma^{CI} < \left[\psi\left(e_L^*\right) - \psi\left(\tilde{e}_L^*\right)\right] / \sigma P$ (the superscript "CI" stands for "cheating-inducing"). The production is low, but nevertheless corresponds to the first-best solution and there is no rent.[36] The agency cost incurred is given by $AC^{CI} = q\left\{x_H^* - x_L^* - \left[\psi\left(e_H^*\right) - \psi\left(e_L^*\right)\right]\right\} - \gamma^{CI}\left(q\sigma P - z\right)$.

Characterizing the cheating-inducing regime is interesting in that it helps understand the reasons why the MoF is unable to deter cheating. This happens in the following cases:

[36] The latter results are explained by the fact that if the MoF is not able to deter cheating, it is not worth incurring further distortions and agency costs.

- The audit probability γ^{CI} is deliberately chosen at a level that is too low.

- The parameters preclude sufficient audit—in particular, if the expected penalty is low compared to the rent, the theoretical γ which would help deter cheating may turn out to be higher than unity, which is irrelevant; this case may also occur when the effectiveness of audit (σ) is too low.

- The MoF does not make use of concomitant incentives (e.g. rents and distortions).

Finally, note that in this case, the agency cost mostly consists of a loss of production when $i = H$. Besides, if $z > q\sigma P$, audit also increases the agency cost. The weight of this component could increase when the MoF supervises several LMs. For instance, if some ministries are less critical than others, and could reasonably function with low production levels, agency costs could be reduced by offering cheating-inducing contracts and save on auditing costs. The money saved could be used to provide the incentives to the priority LMs and offer them a cheating-proof contract, thus ensuring high production (when $i = H$) in these sectors.

C. The Case of "No Commitment" [37]

So far, we have considered that the MoF could credibly commit, at the time of offering the contract, to audit the LM with a given probability once output is observed. However, commitment to audit suffers from a time inconsistency problem: the contract determined by the revelation principle is optimal ex ante, but entails inefficiencies ex post. Indeed, as the contract deters cheating, the audit cost must be incurred without any compensation in terms of collected penalty. Moreover, as the contracted effort is not efficient when $i = L$, one obtains a Pareto improvement by renegotiating the contract once information has been revealed. This time inconsistency reduces the credibility of the commitment to audit, all the more if the MoF is facing a tight budget constraint.[38] Other reasons may also contribute to preclude the commitment to audit, for instance, the difficulty of verifying whether the principal did indeed abide by its committed random audit strategy (Mookherjee and Png, 1989), or, in practice, the absence of adequate legal institutions. In particular, we have so far assumed that the existence of a court of audit in the country consisted in a commitment control. Yet, the inefficiency of the court of audit and related institutions may reduce the credibility of the commitment. Nevertheless, in such situations where the MoF cannot credibly commit to auditing, it can call upon some external auditors for specific tasks.

---

[37] Mixed-strategy equilibriums have generally been used in the literature on collusion. The latter suggests that, instead of trying to systematically deter (or induce) collusion, it may be efficient to allow it to some extent (e.g., Kofman and Lawarrée, 1996; Khalil, 1997; Khalil and Lawarrée, 2003). This may be the case, for example, if there is a positive probability that the agent and the supervisor are honest, so that it may be too costly to provide incentives to systematically deter collusion. We hereafter apply a similar approach to cheating (corruption). We do not model it, but in a context with several LMs, the mixed-strategy equilibrium could also yield the optimal contract when some LMs are honest while others are corrupt, because preventing cheating as if all LMs were corrupt would be too costly.

[38] This argument holds in one-period games. However, in the context of repeated relationships, it is probably in the MoF's interest not to backtrack on its promise to audit with the announced probability, in order to preserve its reputation.

In this subsection, we drop the assumption that the MoF commits to auditing at the time of offering the contract. However, once output is observed, the MoF can decide to audit if it proves to be efficient ex post, i.e., if the MoF expects to get "value-for-money" out of the audit. Indeed, when the output is produced, it is too late to deter cheating—but the MoF can still earn a penalty if cheating is detected. The MoF will be willing to audit only if the expected penalty is at least equal to the audit cost. Moreover, the mere expectation that the MoF may audit will reduce the LM's incentive to cheat.

Formally, with no commitment, the revelation principle cannot be used and audit must be ex post optimal to justify its cost. Hence, cheating may occur in equilibrium (i.e., the probability of cheating is positive), and the MoF can expect to collect a penalty. The MoF's problem is to choose the levels of transfers, effort, and audit probability so as to maximize its objective function:

$$\underset{e_H, e_L, t_H, t_L, \gamma}{Max} E(X) = q(1-m)\left[\alpha(\theta_H, e_H) - t_H\right] + \left[(1-q)+qm\right]\left[\alpha(\theta_L, e_L) - t_L + \gamma(\phi\sigma P - z)\right]$$

Subject to *IR(L)* and the following constraints:

$$(1-m)\left[t_H - \psi(e_H)\right] + m\left[t_L - \psi(\tilde{e}_L) - \gamma\sigma P\right] \geq 0 \qquad IR(H)$$

$$m \in \arg\max_{m'} (1-m')\left[t_H - \psi(e_H)\right] + m'\left[t_L - \psi(\tilde{e}_L) - \gamma\sigma P\right] \qquad IC(H)$$

$$\gamma \in \arg\max_{\gamma'} \left[\gamma'(\phi\sigma P - z)\right] \qquad IC(A)$$

Note that, compared to the previous cases, the MoF's objective function and the constraints now encompass the probability of cheating by the LM (subsections A and B correspond to corner solutions of this general problem). The two *IC* constraints consist of indifference conditions and determine the game, which is played simultaneously by the MoF and the LM. Indeed, the LM is indifferent between cheating and being honest when $t_H - \psi(e_H) = t_L - \psi(\tilde{e}_L) - \gamma\sigma P$. The MoF observes x_L, and is indifferent between auditing and not auditing when $\phi\sigma P = z$.

This problem yields a mixed-strategy equilibrium. Under this regime, production is efficient and the LM gets no rent.[39] The equilibrium is obtained when the MoF audits with probability $\gamma = \left[\psi(e_L^*) - \psi(\tilde{e}_L^*)\right]/\sigma P$, and the LM cheats with probability $m = \left[z(1-q)/(\sigma P - z)q\right]$. The mixed-strategy equilibrium is associated to an agency cost given by $AC^{MS} = \frac{z(1-q)}{\sigma P - z}\left\{x_H^* - x_L^* - \left[\psi(e_H^*) - \psi(e_L^*)\right]\right\}$ (where superscript "MS" stands for "mixed strategy"). Note that penalties are collected at equilibrium, but they do not explicitly enter into the agency cost, because they are exactly offset by the cost of audit. Moreover, as it entails no

[39] The efficiency result is similar to that of Khalil and Lawarrée (2003) and rests on the assumption that the penalty is independent from the transfer. The audit probability is increased until the rent is reduced to zero, and there is no reason to distort efforts. With transfer-dependent penalties, Khalil (1997) finds the agent overproduces when the productivity is low.

production distortion, the mixed-strategy equilibrium tends to the first best when the penalty is very large, or audit is free.[40]

To our knowledge, the literature has not attempted to directly compare regimes with and without commitment, because commitment to audit is usually considered desirable as it reduces the ex ante cost of inducing truthful reporting (Baron and Besanko, 1984). The limited commitment due to the possibility of renegotiating a contract is typically handled by using the renegotiation-proofness principle. The latter, which is somewhat similar to the revelation principle, says that one can, without loss of generality, restrict the set of possible contracts to the class of contracts that are not renegotiated. Renegotiation-proof constraints are thus added to the set of *IRs* and *ICs* (e.g., Bolton, 1990; Dewatripont and Maskin, 1990; but see also Aghion, Dewatripont, and Rey, 1994). In practice, commitment may entail some costs, including those linked to the creation of an adequate institutional setting such as the creation of a court of audit. Our framework allows comparing the agency costs associated with each regime. For instance, one observes that the agency cost of our mixed-strategy equilibrium consists mainly of the loss of production (when $i = H$ and the agent cheats). However, when external audits are cheap, and/or the expected penalty is large, the cheating probability decreases and the mixed strategy becomes a better option. This could reduce the value of establishing a court of audit if it does not yet exist in the country.

D. Applying the Theoretical Framework to PEM Systems

The sections above explain the different control regimes the MoF can implement, according (notably) to the audit probability it chooses. The optimality of these regimes depends on the value of exogenous, country-specific parameters, such as the level of penalty, the quality of the supervision technique, the cost of audit, and the probability of high productivity. This suggests that the need to base the choice of the control design on a good analysis cannot be overemphasized and limits the applicability of "one-size-fits-all" solutions. In this respect, our model provides an analytical framework that can guide PEM reforms, as it allows for comparing different institutional designs, while taking into account the constraints faced by governments.

Generally speaking, the MoF should choose the audit regime associated with the lowest agency cost. Yet, the regimes vary with the institutional setting. The cheating-proof regime corresponds to a situation where the MoF can credibly commit ex ante to auditing with a given probability, for instance if a court of audit exists. This regime can be compared directly with the second-best contract, and the choice will depend on the ratio $q\sigma P/(1-q)z$. The mixed strategy relies on a different assumption: there is no commitment, but the MoF can choose ex post whether to resort to external auditors for specific tasks. To compare these two frameworks, one should add the fixed cost of setting up and running the institutions necessary

[40] Our results are quite intuitive. Both the probability of audit and the probability of cheating are decreasing functions of the expected penalty. The more cheating rent the LM can capture, the more the MoF audits. The more expensive the audit is, the more the LM cheats. The cheating probability is also influenced by the relative probabilities of the productivity occurrences: to keep the MoF indifferent between auditing and not auditing, cheating is increasing with the probability of low productivity. The agency cost decreases with the expected penalty and increases with the probability of low productivity. It is also higher when the difference of production between $i = H$ and $i = L$ is higher.

to allow commitment (e.g., the cost of creating a court of audit) to the agency cost of the cheating-proof regime.

In theory all the regimes discussed above could be implemented. In practice, however, additional constraints may restrain the choices available to the MoF, especially in developing countries. On the one hand, several factors may contribute to reducing the ratio $q\sigma P/(1-q)z$, which conditions the value of audit in those countries. The penalty faced by LMs when cheating is detected may be very small or rarely enforced (see, e.g., Dia, 1996; Lienert, 2003; Moussa, 2004), all the more if discounted at a high rate (because of the time required for implementation) and if the supervision technology is not performing well (low σ, for instance due to poor fiscal data; see, e.g., Bouley, Fournel, and Leruth, 2002); the probability of high output may be low, compared to that expected in industrial countries; and the opportunity cost of audit may be high, considering the scarcity of competent human resources. Therefore, while the MoF can probably deter cheating through ex post audits in industrial countries, this may not be so easy in developing countries.

Practical constraints may also restrict the MoF's actions:

- A tight cash constraint and/or the framework of line item budgeting may limit the availability of informational rents.[41]

- The government may have committed to provide a minimum package of services, which prevents production distortions below a certain level (the level obtained under the optimal contract).

- The MoF may be legally or politically obliged to resort to the court of audit, even if it is not working well.

Finally, if audit is not efficient (due to a low ratio $q\sigma P/(1-q)z$) and if the MoF cannot enforce the second-best contract (for practical reasons), it can only implement the cheating-inducing regime: the most unfavorable for the MoF, and a regime that is never optimal.

Note that our analytical framework may be extended to evaluate various reforms. One can think of reforms to increase the quality of audit, or to use signalling to help determine productivity (e.g., collecting economic information on the sector, and/or auditing the quality of the approved program design).

V. Ex ante Controls[42]

We have so far considered that, notwithstanding the risk of being caught ex post, cheating is costless for the agent. This assumption is common in the literature because the cheating rent generally consists of a reduction in the effort made by the agent, which remains his private

---

[41] Informational rents could be envisaged in a system of performance budgeting, as they consist of rewarding the LM for good performance (above the compensation of its effort).

[42] This section deals with an issue not often discussed in principal-agent papers where the focus tends to be on controls run ex post.

information. But the LM's effort may also comprise some negative actions (such as corruption), and this leads us to assume that cheating entails some costs to be concealed. This allows us to make the link with the economic literature on collusion in organizations.[43] The literature distinguishes two types of collusion costs, according to whether they are exogenous (e.g., negotiation costs, "physical" strategies to divert monies from their intended purposes), or endogenous (e.g., costs stemming from the risk of future detection, see Faure-Grimaud, Laffont, and Martimort, 1999; Khalil and Lawarrée, 2003). Ways in which the principal can avoid corruption include (i) create incentive payments; (ii) decrease the stake of collusion; and (iii) increase the transaction cost of collusion (Laffont and Rochet, 1997). In this section, we introduce an exogenous cost of cheating and explain how it affects the constraints of the MoF's problem. In a second step, we interpret ex ante controls, undertaken by the MoF before the commitment and/or the payment of LM's expenditures, as increasing the cost of cheating. We then discuss the relative value of ex post and ex ante controls.

A. Exogenous Cost of Cheating

We assume that the LM incurs a certain cost $\eta \geq 0$ when it cheats. That cost decreases the expected benefits from cheating. If we first consider a model without ex post audit, the MoF's problem can be written as:

$$\underset{e_H, e_L, t_H, t_L}{Max} \; E(X) = q(1-m)\left[\alpha(\theta_H, e_H) - t_H\right] + (1-q+qm)\left[\alpha(\theta_L, e_L) - t_L\right]$$

Subject to *IR(L)* and:

$$(1-m)\left[t_H - \psi(e_H)\right] + m\left[t_L - \psi(\tilde{e}_L) - \eta\right] \geq 0 \qquad IR(H)$$

$$m \in \arg\max_{m'} (1-m')\left[t_H - \psi(e_H)\right] + m'\left[t_L - \psi(\tilde{e}_L) - \eta\right] \qquad IC(H)$$

We observe that *IC(H)* is relaxed by the cost of cheating. If $t_H - \psi(e_H) < t_L - \psi(\tilde{e}_L) - \eta$, the LM will always cheat ($m = 1$), unless appropriate incentives are provided. For instance, the second-best contract in this case would also entail a rent when $i = H$, and a productive distortion when $i = L$, but these would be reduced by the cheating cost. If the cost of cheating is high enough and $t_H - \psi(e_H) > t_L - \psi(\tilde{e}_L) - \eta$, the LM will not cheat ($m = 0$) and the MoF will reach the first-best solution (no rent and efficient production, as *IC(H)* is not binding). It is thus in the MoF's interest to increase the cost of cheating. This is discussed further in the next subsection.

[43] Following Tirole (1986), that branch of the literature studies the potential for side-contracting between a privately informed, cheating agent and the supervisor hired by the principal to control him or her.

B. Ex Ante Controls as Increasing the Cost of Cheating

Most traditional principal-agent models consider monitoring at the ex post stage.[44] Deterring cheating (or collusion) ex ante is done by granting rewards and/or through a threat of future punishment. For their part, PEM systems also include a series of controls designed to prevent agents from cheating ex ante. For example, automatic tools, such as computer-based systems checking the budget appropriations before allowing spending, are designed for that purpose, and a similar role is played by the financial comptrollers placed by the MoF within LMs. Such controls are particularly extensive in the Francophone Treasury system (Bouley, Fournel, and Leruth, 2002) but also exist in the Anglophone system (Diamond, 2002). As observation suggests, however, these control techniques are not perfect, partly because agents are very active in trying to bypass them (Lienert, 2003; Moussa, 2004).

We hereafter interpret setting up ex ante controls implemented by the MoF as increasing the LM's cost of cheating.[45] We endogenize the cost of cheating as a decision variable of the MoF and first consider a model without ex post audit.

Assume c represents the cost of the ex ante controls. It may be interpreted as the cost of implementing and running controls, but also as the economic cost (sometimes heavy) of procedures that may complicate the expenditure process and reduce the LM's absorptive capacity.[46] The LM's cost of cheating $\eta(c)$ is now endogenous and depends on the controls implemented by the MoF. We assume $\eta_c > 0$, $\eta_{cc} < 0$, $\eta(0) = 0$, and $\lim_{c\to\infty} \eta(c) = \infty$.[47] We limit our analysis to the specification of incentive-compatible schemes (thus, where the MoF deters cheating). The MoF will decide on the levels of ex ante control, transfer, and effort so as to maximize its expected output, as follows:

$$\underset{e_H, e_L, t_H, t_L, c}{Max} E(X) = q\left[\alpha(\theta_H, e_H) - t_H\right] + (1-q)\left[\alpha(\theta_L, e_L) - t_L\right] - c$$

Subject to *IR(L), IR(H)* (which now has the form: $t_H - \psi(e_H) \geq 0$), and

---

[44] As an exception to this statement, Strausz (2006) compares the value of monitoring versus auditing. He argues that if both supervision techniques are equally efficient, auditing is (weakly) superior when the principal can commit to a verification strategy—as auditing can bring out additional information. However, when the principal's verification behavior is noncontractible, monitoring may be optimal. This result is explained because auditing requires steeper incentives, which may be suboptimal if the agent is risk averse and/or due to higher rents. Our analysis differs from Strausz's in that we do not assume that ex post audit and ex ante controls rely on the same technology, and we have assumed (in the cheating-proof regime) that the principal could commit to an audit strategy.

[45] The literature on collusion adopts a similar approach when it acknowledges that hiring a collusive auditor is still useful, because it makes shirking costly for the agent, as he will have to pay a bribe to falsify the report (e.g., Kofman and Lawarrée, 1996).

[46] For instance, in Senegal, the procedures for disbursing the Health Decentralization Fund are such that it takes on average 10 months for the resources to be at the disposal of the providers. This leaves only two months to the facility to absorb those resources (World Bank, 2004a).

[47] The more effective controls are, the higher $\eta(c)$ for any $c > 0$.

$$t_H - \psi(e_H) \geq t_L - \psi(\tilde{e}_L) - \eta(c) \qquad IC(H)$$

Relying on ex ante controls is different from ex post audits because: (i) the performance of ex ante controls is "intrinsic" (depending on η), and does not depend on external factors like the level of penalty; (ii) the cost of controls is incurred ex ante, whatever the state of nature, while the cost of audit is incurred, if at all, only when a low output is observed; and (iii) the decision parameter of the MoF is bounded in the case of ex post audits while, in theory, the MoF could increase ex ante controls infinitely (although it would not be efficient to do so).

The results and proofs related to this discussion are presented in Appendix III. Once again, the problem remains to respect *IC(H)* at the lowest cost, hence the importance of the shadow cost (Lagrance multiplier) of that constraint. We first show that it is not efficient to increase controls above the point where $\left[1/\eta_c(c)\right] \geq q$. If the cost of deterring cheating by controls alone is too high (i.e., when the minimum level of control that would be necessary to deter cheating, $\eta(c) = t_L - \psi(\tilde{e}_L)$, would be such that $\left[1/\eta_c(c)\right] > q$), the MoF must also use other means to deter cheating, including rents and productive distortions. For instance, the LM could be granted a rent when $i = H$ (the level of this rent decreases with the amount of control by the MoF) and be required to produce the second-best level of effort when $i = L$. In that case, the MoF's expected output corresponds to the second-best solution.

When ex ante controls are sufficiently effective to allow the MoF to deter cheating by raising controls (until a point such that $\left[1/\eta_c(c)\right] < q$), the LM gets no rent. However, it may be profitable for the MoF to distort the required effort, if it can save by reducing the level of controls. The MoF would then choose the level of control and the effort required when $i = L$ by comparing their cost (i.e., so as to relax *IC(H)* at the lowest cost). The trade-off here is between increased efficiency and the cost of control. The agency cost would then be equal to $AC^{AC} = (1-q)\left\{x_L^* - x_L^{AC} - \left[\psi(e_L^*) - \psi(e_L^{AC})\right]\right\} + c^*$ (where the superscript "AC" stands for "ex ante controls").

The results do not fundamentally differ from the cheating-proof regime with ex post audits. Both types of control are assessed with regard to their ability to deter cheating, and if their cost is too high relative to their benefits, the MoF has to use other means (rents and distortions). Yet, one can compare the relative value of both types of controls on the basis of the shadow cost of each problem's *IC(H)* constraint, i.e., the Lagrange multiplier of those constraints at equilibrium. These are respectively given by $\lambda_{AC} = 1/\eta_c(c)$ for ex ante controls, and $\lambda_{EP} = \left[(1-q)z + \lambda_\gamma\right]/\sigma P$ for ex post audits, where λ_γ represents the shadow cost of the constraint on the maximum audit probability $\gamma = 1$, (where the subscript "EP" stands for "ex post audits").[48] Both measure the difficulty of deterring cheating and determine the agency costs (cost of control, rents, and production distortions) associated with each regime. For instance, it is not profitable for the MoF to increase the level of ex ante controls

[48] Under the random audit feature, $\lambda_\gamma = 0$.

when $\left[1/\eta_c(c)\right] > q$, nor to use ex post audits when $\left[(1-q)z+\lambda_\gamma\right]/\sigma P > q$. An analysis of shadow costs reveals that both thresholds depend on the probability that $i = H$. It also reveals that the higher the probability of high output, the higher the incentive to use controls to guarantee it. Second, the effect of ex post audits on relaxing the constraint *IC(H)* is mitigated by the size of the expected penalty. As already mentioned, penalties are low and/or not enforced in developing countries, and ex post audits may not be able to deter cheating, making ex ante controls more effective. Third, if penalties are a credible threat, ex post audits may prove to be effective because their cost is incurred only when the principal observes a low output (i.e., with probability $(1-q)$ under a cheating-proof regime) contrary to ex ante controls which are imposed in a nondiscriminatory manner.

C. Integrating Ex Ante and Ex Post Controls

We now briefly consider (without explicitly solving) the case in which the MoF can deter cheating through a combination of ex ante controls and ex post audits.[49] Limiting ourselves to incentive compatible schemes, the MoF's problem is to maximize the expected output:

$$\underset{e_H, e_L, t_H, t_L, \gamma, c}{Max} E(X) = q\left[\alpha(\theta_H, e_H) - t_H\right] + (1-q)\left[\alpha(\theta_L, e_L) - t_L - \gamma z\right] - c$$

Subject to *IR(L), IR(H)* and:

$$t_H - \psi(e_H) \geq t_L - \psi\left(\tilde{e}_L\right) - \eta(c) - \gamma\sigma P \qquad IC(H)$$

We have already solved both components of this problem separately, through the optimization of the MoF's problem with respect to the audit probability and the control cost, respectively. Under a combined approach, the most interesting case occurs when the MoF simultaneously uses both types of controls. This may only take place when $\lambda = 1/\eta_c(c) = \left[(1-q)z+\lambda_\gamma\right]/\sigma P \leq q$. This may be interpreted as linking the cost-effectiveness of each type of control and comparing their costs (resp. c and $\left[(1-q)z+\lambda_\gamma\right]$) to their effectiveness in deterring cheating (which depends on η and σP). When the MoF uses both audits and controls in combination, it will thus equate the relative contribution of each type of control. Finally, the optimal level of ex ante controls and probability of ex post

[49] We could extend the analysis to a case in which the MoF would like to implement a mixed-strategy equilibrium with respect to ex post audit, while also using ex ante controls. Compared to the case presented in subsection IV.C., the audit probability that keeps the LM indifferent between cheating and complying would be reduced by the use of ex ante controls: $\gamma = \left[\psi\left(e_L^*\right) - \psi\left(\tilde{e}_L^*\right) - \eta(c)\right]/\sigma P$. However, the probability of cheating would not be influenced by the introduction of ex ante controls, as its equilibrium value is determined by keeping the MoF indifferent between auditing and not auditing: $m = \left[z(1-q)\right]/\left[(\sigma P - z)q\right]$. We know that the agency cost associated with the mixed strategy is such that the cost of audit offsets the benefits from penalty. This would still hold in this case, so that the cost of ex ante controls would not have any counterpart benefit, neither in terms of reduced cheating, nor in terms of collected penalty. Therefore, our model suggests that when it chooses to implement a mixed-strategy equilibrium, the MoF should not use ex ante controls at the same time.

audits, as well as the production distortion when $i = L$ and the possible rent when $i = H$, will be determined simultaneously, so as to satisfy the constraint *IC(H)* at the lowest cost.

VI. CONCLUSION

We have argued that the principal-agent theory offers a powerful analytical framework to better understand PEM systems and guide their design in developing countries. The model discussed in the paper equally applies to "managerial" systems relying on ex post audits (in the British tradition), or systems relying more on ex ante controls (in many Francophone African countries). It allows for comparisons between institutional settings (e.g., depending on whether or not the MoF is able to commit to a certain audit probability) and types of control (e.g., comparing the effectiveness of ex post audits and ex ante controls) by examining the cost-effectiveness of various tools available to the principal to deter cheating. However, this often entails some productive distortions, which result from a trade-off between economic efficiency, on the one hand, and the cost of control and/or an informational rent, on the other hand. Finally, we have interpreted corruption and the lack of governance as "rents" captured by the LMs at the expense of their principal as a result of the informational advantage. By assuming that the agent's effort encompasses productive effort, as well as negative actions such as those related to corruption, we have linked the model to the literature on collusion in organizations.

The model shows that several regimes can exist and their optimality depends on country-specific parameters, hence the importance of basing the choice of a PEM system on a detailed analysis. Nevertheless, it is possible to draw a few general lessons that can help PEM advisors address some important issues:

- **Ex post controls** (which we mostly assimilate to a court of audit in the paper) should be used up to the point where their marginal cost is equal to their return in terms of improved economic efficiency. This will depend on several parameters such as arbitrarily low or ineffective penalties (often the result of social pressure or a weak judicial system in developing countries). Rather than setting up a court of audit (they do not exist in all countries), it may then be profitable to rely on other tools such as external, private audit firms, which increase the cost of cheating for the LMs. In certain conditions, we also stress the importance of setting up a court of audit so as to make the commitment assumption credible and, in conjunction with better funding, increase its activities, thereby increasing the deterring aspect of the threat of punishment.

- The effectiveness of ***internal*** **controls** is similarly determined by cost-benefit considerations, but money spent on internal controls tends to be more effective than money spent on ex post controls in developing countries. An important parameter is the extent to which these controls can be bypassed, for example through the use of extraordinary procedures. The cost of internal controls should be assessed carefully, taking into account not only the cost of additional controllers or systems, but also the economic cost due to the resulting slow down of the expenditure process.

- In countries where the efficiency of both internal and external controls is dubious, theory recommends that the LM should be granted an **"informational rent"** in the form of a transfer above the compensation for the effort made. However, in practice, and beyond the difficulty of implementing such schemes in a public sector context in many countries, the efficiency of informational rents may be reduced if appropriate performance measures on which to base the contract between the MoF and the LM are unreliable or even unavailable.

- The model may also help **sequencing reforms**, although we do not develop this aspect in the paper. As causes for the poor performance of the PEM system are identified, it is possible to decide on what measures should take priority. For instance, if the MoF is not in a position to deter cheating by introducing internal controls, nor to grant informational rents, it is trapped in the so-called cheating-inducing regime. A first step could be to announce that private audit firms will be hired. In our model, this would relate to implementing a mixed-strategy equilibrium, which tends to be an easily implementable and cheaper solution (as it incurs no fixed cost) to reduce the extent of cheating. If the country lacks reliable fiscal data, ex post audits are not very effective and the priority should be to reinforce the accounting system, before reinvigorating the court of audit.

Finally, although the principal-agent theory provides very interesting insights for the design of PEM systems, we have only considered a few aspects and many more are worth exploring. For instance, a principal-agent analysis could be applied to the allocation of resources for control purposes between different LMs (for example because they have different probabilities of cheating). Future research could also focus on the dynamic aspects of PEM design and take into account the repeated interactions between the MoF and LMs at the time the contract is prepared. Although not easily tractable, the realities of the negotiation process between the MoF and the LMs are very complex, with some LMs being better informed than others.

Expenditure Tracking Indicators and Benchmarks

Public Expenditure Management Issue		Benchmark Description
Formulation	1. Composition of the budget entity	Very close fit to government finance statistics (GFS) definition of general government
	2. Limitations to the use of off-budget transactions	Extra (or off) budget expenditure is not significant
	3. Reliability of budget as guide to outturn	Level and composition of outturn is "quite close" to budget
	4. Data on donor financing	Donor-funded expenditures are included in budget or reports
	5. Classification of budget transactions	Functional and/or program information provided
	6. Identification of poverty-reducing expenditure	Identified through use of classification system
	7. Quality of multiyear expenditure projections	Projections are integrated into budget formulation
Execution	8. Level of payment arrears	Very few or no arrears accumulated
	9. Quality of internal audit	Effective internal audit function
	10. Use of expenditure tracking surveys	Tracking used on regular basis
	11. Quality of fiscal/banking data reconciliation	Satisfactory and timely reconciliation of fiscal and monetary data
Reporting	12. Timeliness of internal budget reports	Monthly expenditure reports provided within four weeks of end of month
	13. Classification used for tracking poverty-reducing expenditures	Good quality, timely functional reporting derived from classification system
	14. Timeliness of accounts closure	Accounts closed within two months of year-end
	15. Timeliness of final audited accounts	Audited accounts presented to legislature within one year
Procurement	16. Effective procurement	Procurement processes promote competition, transparency, and value-for-money

Source: IDA/IMF, 2005.

The Cheating-Proof Regime with Ex Post Audits[50]

We first present the general case before deriving some specific results from a numerical example.

1. General case

The optimal, incentive-compatible contract problem has the following Lagrangian:

$$L = q\left[\alpha(\theta_H, e_H) - t_H\right] + (1-q)\left[\alpha(\theta_L, e_L) - t_L - \gamma z\right]$$
$$+\lambda_1\left[t_H - \psi(e_H)\right] + \lambda_2\left[t_L - \psi(e_L)\right] + \lambda_3\left[t_H - \psi(e_H) - t_L + \psi(\tilde{e}_L) + \gamma\sigma P\right] + \lambda_4(1-\gamma)$$

with the additional non-negativity constraints.
The Kuhn-Tucker conditions for maximization are (directly considering the efforts and transfers are positive, so that the associated conditions hold with equality):

(2.1) $$\frac{\partial L}{\partial e_H} = q\alpha_e(\theta_H, e_H) - (\lambda_1 + \lambda_3)\psi_e(e_H) = 0 \;;$$

(2.2) $$\frac{\partial L}{\partial e_L} = (1-q)\alpha_e(\theta_L, e_L) - \lambda_2\psi_e(e_L) + \lambda_3 \frac{\partial \psi(\tilde{e}_L)}{\partial e_L} = 0 \;;$$

(2.3) $$\frac{\partial L}{\partial t_H} = -q + \lambda_1 + \lambda_3 = 0 \;;$$

(2.4) $$\frac{\partial L}{\partial t_L} = -(1-q) + \lambda_2 - \lambda_3 = 0 \;;$$

(2.5) $$\frac{\partial L}{\partial \gamma} = -(1-q)z + \lambda_3\sigma P - \lambda_4 \le 0 \text{ and } \gamma\frac{\partial L}{\partial \gamma} = 0 \;;$$

plus the constraints and their complementary slackness conditions.

From (2.5), audit is not efficient ($\gamma = 0$) when $\lambda_3 < \dfrac{(1-q)z}{\sigma P}$.
From (2.3), we know that $q = \lambda_1 + \lambda_3$, so that $\lambda_3 \le q$. Thus, from (2.5), audit is optimal ($\gamma > 0$) only if $\lambda_3 = \dfrac{(1-q)z + \lambda_4}{\sigma P} \le q$, and therefore $(1-q)z + \lambda_4 \le q\sigma P$.

[50] This analysis and its results are very similar to that of Kofman and Lawarrée (1993) with one truthful costly auditor, at the following exceptions: (i) we assume the principal cannot earn a penalty out of a complying agent; (ii) we assume a fixed penalty; (iii) our specification of the production and disutility of effort functions differ.

Several general results may be drawn:

- ***IC(H)* is binding**: either audit is not efficient, so that the second-best contract is chosen, and it is well-known that *IC(H)* is binding; or audit is efficient, and from (2.5) $\lambda_3 > 0$, so that *IC(H)* is also binding.
- ***IR(L)* is binding:** from (2.4), $\lambda_2 = (1-q) + \lambda_3 > 0$, so that **there is no rent when** $i = L$.
- **Production is efficient when** $i = H$: from (2.3), $q = \lambda_1 + \lambda_3$; replacing in (2.1), we find that $\alpha_e(\theta_H, e_H) = \psi_e(e_H)$.
- **There is underproduction when** $i = L$: from the Implicit Function Theorem, one has $\frac{d\tilde{e}_L}{de_L} = \frac{\alpha_e(\theta_L, e_L)}{\alpha_e(\theta_H, \tilde{e}_L)}$. From (2.4), we know $\lambda_2 = (1-q) + \lambda_3$. Replacing in (2.2) yields:
  $\left\{1 - q + \lambda_3\left[\psi_e(\tilde{e}_L)/\alpha_e(\theta_H, \tilde{e}_L)\right]\right\}\alpha_e(\theta_L, e_L) = [1 - q + \lambda_3]\psi_e(e_L)$.
  $\alpha_e(\theta_L, e_L)$ could be equal to $\psi_e(e_L)$ only if $\lambda_3 = 0$ (which is not possible: we have seen that *IC(H)* was always binding) or if $\psi_e(\tilde{e}_L) = \alpha_e(\theta_H, \tilde{e}_L)$. But we know that $\alpha_e(\theta_L, e_L) < \alpha_e(\theta_H, \tilde{e}_L)$ and $\psi_e(\tilde{e}_L) < \psi_e(e_L)$, so that it is inconsistent. Hence, $\psi_e(\tilde{e}_L) < \alpha_e(\theta_H, \tilde{e}_L)$ and $\alpha_e(\theta_L, e_L) > \psi_e(e_L)$.

Three cases can occur according to the value of the parameters:

1) If $q\sigma P < (1-q)z$, **audit is never optimal** and, from the Revelation Principle, the second-best contract is preferable.

2) If $q\sigma P = (1-q)z$, the MoF is indifferent between auditing and not auditing.

From (2.5), we know that $q\sigma P = (1-q)z \geq \lambda_3\sigma P - \lambda_4$, which is possible only if the latter equation holds with equality (otherwise $\gamma = 0$ and the second-best contract is optimal), and $\lambda_3 = q$ and $\lambda_4 = 0$ so that $0 < \gamma < 1$.
Replacing λ_3 in (2.2), one has: $\left\{1 - q + q\left[\psi_e(\tilde{e}_L)/\alpha_e(\theta_H, \tilde{e}_L)\right]\right\}\alpha_e(\theta_L, e_L) = \psi_e(e_L)$. The production distortion when $i = L$ is e_L^{SB} (second-best).
As $\lambda_3 = q$, we know from (2.3) that $\lambda_1 = 0$, so that there is a rent when $i = H$.
The problem is undetermined and the MoF is indifferent between auditing with probability $0 < \gamma^{EP} = \left[\psi(e_H^*) + \psi(e_L^{SB}) - \psi(\tilde{e}_L^{SB}) - t_H\right]/\sigma P < 1$ and leaving a rent equal to $t_H = \psi(e_H^*) + \psi(e_L^{SB}) - \psi(\tilde{e}_L^{SB}) - \gamma^{EP}\sigma P$ when $i = H$, or offering the second-best contract.
In this case, the MoF's expected output and the agency cost incurred correspond to the second-best solution.

3) If $q\sigma P > (1-q)z$, audit is efficient and the optimal contract may exhibit different types.

From (2.5), we know that $\lambda_3 \sigma P - \lambda_4 = (1-q)z$. Three cases may arise:[51]

- $\lambda_3 = q$ (therefore $\lambda_1 = 0$) and $\lambda_4 > 0$ (therefore $\gamma = 1$): "Rent extraction" (RE);
- $\lambda_3 < q$ (therefore $\lambda_1 > 0$) and $\lambda_4 > 0$ (therefore $\gamma = 1$): "Effort adjustment" (EA);
- $\lambda_3 < q$ (therefore $\lambda_1 > 0$) and $\lambda_4 = 0$ (therefore $\gamma < 1$): "Random audit" (RA).

We characterize the three contracts hereafter.

a. Systematic audit and "rent extraction" (RE)

When $\lambda_3 = q$ and $q\sigma P > (1-q)z$, we know from (2.5) that $\lambda_4 > 0$, which means that **the MoF audits with probability one:** $\gamma = 1$.
Replacing λ_3 in (2.2), we find that **the production distortion when** $i = L$ **is the same as at the second-best:** e_L^{SB}**.**
From (2.3), we also know that $\lambda_1 = 0$, so that *IR(H)* is not binding and **there is a rent when** $i = H$. The corresponding transfer is $t_H = \psi(e_H^*) + \psi(e_L^{SB}) - \psi(\tilde{e}_L^{SB}) - \sigma P$.
The MoF's expected output is thus:

$$E(X^{RE}) = q\left[\alpha(\theta_H, e_H^*) - \psi(e_H^*) - \psi(e_L^{SB}) + \psi(\tilde{e}_L^{SB}) + \sigma P\right] + (1-q)\left[\alpha(\theta_L, e_L^{SB}) - \psi(e_L^{SB}) - z\right].$$

The agency cost incurred is equal to $AC^{RE} = AC^{SB} - q\sigma P + (1-q)z$.

b. Systematic audit and "effort adjustment" (EA)

When $\lambda_3 < q$, we know from (2.3) that $\lambda_1 > 0$, so that *IR(H)* is binding and **there is no rent when** $i = H$: $t_H = \psi(e_H^*)$.
As long as $\lambda_4 > 0$, **the MoF audits with probability one:** $\gamma = 1$.
However, from (2.2) we know that
$\left\{1 - q + \lambda_3\left[\psi_e(\tilde{e}_L)/\alpha_e(\theta_H, \tilde{e}_L)\right]\right\}\alpha_e(\theta_L, e_L) = [1 - q + \lambda_3]\psi_e(e_L)$, which suggests that the

[51] These cases are similar to those found by Kofman and Lawarrée (1993), and we use the same labels as these authors.

productive distortion is smaller than at the second-best.[52] The latter is defined by *IC(H)* in such a way that $\psi\left(e_L^{EA}\right)=\psi\left(\tilde{e}_L^{EA}\right)+\sigma P$.

The MoF's expected output is:

$$E\left(X^{EA}\right)=q\left[\alpha\left(\theta_H,e_H^*\right)-\psi\left(e_H^*\right)\right]+(1-q)\left[\alpha\left(\theta_L,e_L^{EA}\right)-\psi\left(e_L^{EA}\right)-z\right].$$

The agency cost incurred is equal to

$$AC^{EA}=(1-q)\left\{\left[\alpha\left(\theta_L,e_L^*\right)-\alpha\left(\theta_L,e_L^{EA}\right)\right]-\left[\psi\left(e_L^*\right)-\psi\left(e_L^{EA}\right)\right]+z\right\}.$$

c. Random audit (RA)

When $\lambda_4=0$, **the MoF audits with probability below one** (random audit): $0<\gamma<1$.

From (2.5), we know that $\lambda_3=\dfrac{(1-q)z}{\sigma P}<q$. Replacing in (2.2) and dividing by $(1-q)$:

$$\left\{1+\frac{z}{\sigma P}\left[\psi_e\left(\tilde{e}_L\right)/\alpha_e\left(\theta_H,\tilde{e}_L\right)\right]\right\}\alpha_e\left(\theta_L,e_L\right)=\left[1+\frac{z}{\sigma P}\right]\psi_e\left(e_L\right).$$ This equality defines the productive distortion e_L^{RA}.

As $\lambda_1>0$, *IC(H)* is binding, and we may determine $\gamma^{RA}=\left[\psi\left(e_L^{RA}\right)-\psi\left(\tilde{e}_L^{RA}\right)\right]/\sigma P$.

In this case, the MoF's expected output is:

$$E\left(X^{RA}\right)=q\left[\alpha\left(\theta_H,e_H^*\right)-\psi\left(e_H^*\right)\right]+(1-q)\left[\alpha\left(\theta_L,e_L^{RA}\right)-\psi\left(e_L^{RA}\right)-\gamma^{RA}z\right].$$

The agency cost incurred is equal to

$$AC^{RA}=(1-q)\left\{\left[\alpha\left(\theta_L,e_L^*\right)-\alpha\left(\theta_L,e_L^{RA}\right)\right]-\left[\psi\left(e_L^*\right)-\psi\left(e_L^{RA}\right)\right]+\gamma^{RA}z\right\}.$$

The example below allows us to show that, with a consistent formulation of the productive and disutility of effort functions, the EA and RA productive distortions are smaller than at the second-best. We can also characterize the boundaries between the different regimes.

2. Example

Let $x=\theta^2e^{1/2}$ and $\psi(e)=e^3/3$, which respect our assumptions regarding the signs of the derivatives of both functions.

The Lagrangian is now:

[52] The general formulation of the productive and disutility of effort functions do not allow us to prove that the EA and RA productive distortions are always smaller than the second-best one. For this purpose, we use an example below.

$$L = q\left[(\theta_H)^2(e_H)^{\frac{1}{2}} - t_H\right] + (1-q)\left[(\theta_L)^2(e_L)^{\frac{1}{2}} - t_L - \gamma z\right]$$
$$+\lambda_1\left[t_H - \frac{(e_H)^3}{3}\right] + \lambda_2\left[t_L - \frac{(e_L)^3}{3}\right] + \lambda_3\left[t_H - \frac{(e_H)^3}{3} - t_L + \frac{(\theta_L/\theta_H)^{12}(e_L)^3}{3} + \gamma\sigma P\right] + \lambda_4(1-\gamma)$$

The Kuhn-Tucker conditions for maximization are now ((2.3), (2.4) and (2.5) are similar):

(2.1)' $$\frac{\partial L}{\partial e_H} = \frac{1}{2}q\left[(\theta_H)^2(e_H)^{-\frac{1}{2}}\right] - (\lambda_1 + \lambda_3)(e_H)^2 = 0$$

(2.2)' $$\frac{\partial L}{\partial e_L} = \frac{1}{2}(1-q)\left[(\theta_L)^2(e_L)^{-\frac{1}{2}}\right] - \lambda_2(e_L)^2 + \lambda_3\left[(\theta_L/\theta_H)^{12}(e_L)^2\right] = 0$$

The general results are similar to the general case: *IC(H)* and *IR(L)* are binding; production is efficient when $i = H$; there is underproduction when $i = L$. Moreover:

- From (2.1)', we find that the optimal level of effort is $e_H^* = \left[\frac{1}{2}(\theta_H)^2\right]^{2/5}$.
- From (2.2)', we find that $e_L = \left\{\frac{(1-q)(\theta_L)^2}{2\left[(1-q)+\lambda_3\left(1-(\theta_L/\theta_H)^{12}\right)\right]}\right\}^{2/5} < e_L^* = \left[\frac{1}{2}(\theta_L)^2\right]^{2/5}$ when $0 < \lambda_3 \le q$. In particular, the second-best distortion when $i = L$ is such that $e_L^{SB} = \left\{\frac{(1-q)(\theta_L)^2}{2\left[1-q(\theta_L/\theta_H)^{12}\right]}\right\}^{2/5}$.
- **The productive distortion under "effort adjustment" (EA) is smaller than at the second-best**: this results derives from the two expressions of the effort above, knowing that $\lambda_3 < q$ and $\theta_L < \theta_H$, so that $e_L^{EA} > e_L^{SB}$.
- More precisely, the productive distortion under "effort adjustment" is defined by *IC(H)* in such a way that $\psi\left(e_L^{EA}\right) = \psi\left(\tilde{e}_L^{EA}\right) + \sigma P$, i.e.

  $$\frac{1}{3}\left\{\frac{(1-q)(\theta_L)^2}{2\left[(1-q)+\lambda_3\left(1-(\theta_L/\theta_H)^{12}\right)\right]}\right\}^{6/5}\left[1-(\theta_L/\theta_H)^{12}\right] = \sigma P$$. From this equation, we can find the value of λ_3 and, replacing in e_L^{EA} , we find that $e_L^{EA} = \left[\frac{3\sigma P}{1-(\theta_L/\theta_H)^{12}}\right]^{1/3}$.
- From (2.5) and (2.2)', we find that the effort required when $i = L$ under "random audit" (RA) is: $e_L^{RA} = \frac{(\theta_L)^{4/5}}{\left\{2\left[1+\frac{z}{\sigma P}\left[1-(\theta_L/\theta_H)^{12}\right]\right]\right\}^{2/5}}$. As $q\sigma P > (1-q)z$ and $\theta_L < \theta_H$, it is

easy to show that $e_L^{RA} > e_L^{SB}$, which means that **the productive distortion under random audit is also smaller than at the second-best**.

- From *IC(H)*,we may determine the audit probability under the RA scheme:

$$\gamma^{RA} = \frac{(\theta_L)^{12/5}\left[1-(\theta_L/\theta_H)^{12}\right]}{3\sigma P\left\{2\left[1+\frac{z}{\sigma P}\left[1-(\theta_L/\theta_H)^{12}\right]\right]\right\}^{6/5}}.$$

This example allows us to determine the boundaries between the different regimes. Let $\frac{\left[(1-q)(\theta_L)^2\right]^{6/5}\left[1-(\theta_L/\theta_H)^{12}\right]}{3\left\{2\left[1-q(\theta_L/\theta_H)^{12}\right]\right\}^{6/5}} \equiv A$ and $\frac{(\theta_L)^{12/5}\left[1-(\theta_L/\theta_H)^{12}\right]}{3\left\{2\left[1+\frac{z}{\sigma P}\left[1-(\theta_L/\theta_H)^{12}\right]\right]\right\}^{6/5}} \equiv B$ (Note that $A < B$).

Proposition:
- The RE scheme takes place when $\sigma P \leq A$
- The EA scheme takes place when $A < \sigma P \leq B$
- The RA scheme takes place when $\sigma P > B$

Proof:

- To satisfy *IR(H)* under the RE scheme, one must have $\psi\left(e_L^{SB}\right)-\psi\left(\tilde{e}_L^{SB}\right)-\sigma P \geq 0$, and

  therefore $\sigma P \leq \frac{1}{3}\left\{\frac{(1-q)(\theta_L)^2}{2\left[1-q(\theta_L/\theta_H)^{12}\right]}\right\}^{6/5}\left[1-(\theta_L/\theta_H)^{12}\right] \equiv A$.

- To satisfy $\gamma^{RA} < 1$ under the RA scheme, one must have

$$\sigma P > \frac{(\theta_L)^{12/5}\left[1-(\theta_L/\theta_H)^{12}\right]}{3\left\{2\left[1+\frac{z}{\sigma P}\left[1-(\theta_L/\theta_H)^{12}\right]\right]\right\}^{6/5}} \equiv B.$$

- The EA scheme is the only one which corresponds to $A < \sigma P \leq B$.
  Under EA, the productive distortion is smaller than at the second-best ($e_L^{EA} > e_L^{SB}$), if $\sigma P > A$, and it is higher than at the RA ($e_L^{RA} > e_L^{EA}$) if $\sigma P < B$. It is easy to show that the EA scheme satisfies the Kuhn-Tucker conditions (2.1) and (2.2).

- Finally, the problem is continuous:
  - SB-RE: the productive distortion is similar, the difference arises from the fact that σP reduces the rent.
  - RE-EA: the productive distortion is similar ($e_L^{EA} = e_L^{SB}$) at $\sigma P = A$ and $\lambda_3 = q$.

- <u>EA-RA:</u> the productive distortion is similar ($e_L^{EA} = e_L^{RA}$) at $\sigma P = B$ and $\frac{(1-q)z}{\sigma P} = \lambda_3$.

The Cheating-Proof Regime with Ex Ante Controls

The optimal, incentive-compatible contract problem has the following Lagrangian:

$$L = q\left[\alpha\left(\theta_H, e_H\right) - t_H\right] + (1-q)\left[\alpha\left(\theta_L, e_L\right) - t_L\right] - c + \lambda_1\left[t_H - \psi\left(e_H\right)\right] + \lambda_2\left[t_L - \psi\left(e_L\right)\right]$$
$$+\lambda_3\left\{t_H - \psi\left(e_H\right) - \left[t_L - \psi\left(\tilde{e}_L\right) - \eta\left(c\right)\right]\right\}$$

with the additional non-negativity constraints.
The Kuhn-Tucker conditions for maximization are (directly considering the efforts and transfers are positive, so that the associated conditions hold with equality):

(3.1) $\dfrac{\partial L}{\partial e_H} = q\alpha_e\left(\theta_H, e_H\right) - \left(\lambda_1 + \lambda_3\right)\psi_e\left(e_H\right) = 0$;

(3.2) $\dfrac{\partial L}{\partial e_L} = \left(1-q\right)\alpha_e\left(\theta_L, e_L\right) - \lambda_2\psi_e\left(e_L\right) + \lambda_3\dfrac{\partial\psi\left(\tilde{e}_L\right)}{\partial e_L} = 0$;

(3.3) $\dfrac{\partial L}{\partial t_H} = -q + \lambda_1 + \lambda_3 = 0$;

(3.4) $\dfrac{\partial L}{\partial t_L} = -\left(1-q\right) + \lambda_2 - \lambda_3 = 0$;

(3.5) $\dfrac{\partial L}{\partial c} = -1 + \lambda_3\eta_c\left(c\right) \le 0$ and $c\dfrac{\partial L}{\partial c} = 0$;

plus the constraints and their complementary slackness conditions.

Several general results may be drawn:

- ***IC(H)* is binding:** either $c = 0$ and the second-best contract is chosen, or $c > 0$ and from (3.5) $\lambda_3 > 0$, so that *IC(H)* is binding in both cases.
- ***IR(L)* is binding:** from (3.4), one has $\lambda_2 > 0$, so that **there is no rent when** $i = L$.
- **Production is efficient when** $i = H$: from (3.3), one has $q = \lambda_1 + \lambda_3$; replacing in (3.1), we find that $\alpha_e\left(\theta_H, e_H\right) = \psi_e\left(e_H\right)$.
- **There is under-production when** $i = L$: by a similar argument as in Appendix II.

The optimal level of ex ante control is determined as follows:

First note that the minimum control cost necessary to deter cheating (possibly in combination with rents and productive distortions) is such that $\eta\left(c\right) = t_L - \psi\left(\tilde{e}_L\right) - t_H + \psi\left(e_H\right)$.

1. It is not profitable to increase control as long as $\left[1/\eta_c\left(c\right)\right] > q$.

Proof:
From (3.5), $c = 0$ when $\lambda_3 < 1/\eta_c(c)$. From (3.3), we know that $q = \lambda_1 + \lambda_3$, so that $\lambda_3 \leq q$. Thus, from (3.5) control is optimal only if $\lambda_3 = 1/\eta_c(c) \leq q$.

2. The maximal level until which the MoF can increase control is $c : [1/\eta_c(c)] = q$. The MoF is then indifferent between controlling and not controlling. If the cost of deterring cheating by controls alone is too high, the MoF must use a combination of incentives (including rents and productive distortions).[53]

Proof:
If the MoF opts for imposing controls, from (3.5) we know that $\lambda_3 = [1/\eta_c(c^*)] = q$, and from (3.3) that $\lambda_1 = 0$. *IR(H)* is not binding and **there is a rent when** $i = H$.
Replacing λ_3 in (3.2), we find that **the productive distortion when** $i = L$ **is equivalent to the second-best:** the LM is required to produce e_L^{SB}.
As *IC(H)* is binding, the transfer when $i = H$ must be equal to
$t_H = \psi(e_H^*) + \psi(e_L^{SB}) - \psi(\tilde{e}_L^{SB}) - \eta(c^*)$.
We know that $\lim_{c \to c^*} \frac{c}{\eta(c)} = \lim_{c \to c^*} \frac{1}{\eta_c(c)} = q$, so that $c^* = q\eta(c^*)$.
The problem is undetermined. The MoF is indifferent between offering the second-best contract on the one hand; and, on the other hand, controlling the MoF and receiving the following expected output:

$$E(X) = q\left[\alpha(\theta_H, e_H^*) - \psi(e_H^*) - \psi(e_L^{SB}) + \psi(\tilde{e}_L^{SB}) + \eta(c^*)\right] + (1-q)\left[\alpha(\theta_L, e_L^{SB}) - \psi(e_L^{SB})\right] - c^*$$

The agency cost incurred is the second-best.

3. If the cost of deterring cheating by controls alone is low enough, the MoF increases controls until $[1/\eta_c(c^*)] < q$ and leaves no rent to the LM.

Proof:
From (3.5), $\lambda_3 = [1/\eta_c(c^*)] < q$, so that $\lambda_1 > 0$. *IR(H)* is binding and there is **no rent**.
The MoF deters cheating by increasing control until $\eta(c^*) = \psi(e_L^{AC}) - \psi(\tilde{e}_L^{AC})$, where e_L^{AC} and $\tilde{e}_L^{AC}$ are determined by (3.2):

[53] This situation can occur when the minimum level of control necessary to deter cheating, $\eta(c) = t_L - \psi(\tilde{e}_L)$, is such that $[1/\eta_c(c)] > q$. In that case, the MoF must leave a rent to the LM when $i = H$.

$\left\{1-q+\left(1/\eta_c\left(c^*\right)\right)\left[\psi_e\left(\tilde{e}_L^{AC}\right)/\alpha_e\left(\theta_H,\tilde{e}_L^{AC}\right)\right]\right\}\alpha_e\left(\theta_L,e_L^{AC}\right)=\left[1-q+\left(1/\eta_c\left(c^*\right)\right)\right]\psi_e\left(e_L^{AC}\right)$.

This problem consists of two equations with two unknowns, c^* and e_L^{AC}.

Using the same argument as in the example of Appendix II, we can show that, under reasonable assumptions regarding the form of the production and effort functions, and as $\lambda_3 < q$, the productive distortion associated to this regime is lower than at the second-best: $e_L^* > e_L^{AC} > e_L^{SB}$.

The MoF's expected output is:

$E(X)=q\left[\alpha\left(\theta_H,e_H^*\right)-\psi\left(e_H^*\right)\right]+(1-q)\left[\alpha\left(\theta_L,e_L^{AC}\right)-\psi\left(e_L^{AC}\right)\right]-c^*$.

The agency cost incurred is $AC^{AC}=(1-q)\left\{\left[\alpha\left(\theta_L,e_L^*\right)-\alpha\left(\theta_L,e_L^{AC}\right)\right]-\psi\left(e_L^*\right)+\psi\left(e_L^{AC}\right)\right\}+c^*$.

<u>Illustration:</u>

We illustrate these results using $x=\theta^2 e^{\frac{1}{2}}$, $\psi(e)=\dfrac{e^3}{3}$, and $\eta(c)=c^{1/2}$ which respect our assumptions.

From (3.2), and replacing e_L^{AC} in $\eta\left(c^*\right)=\psi\left(e_L^{AC}\right)-\psi\left(\tilde{e}_L^{AC}\right)$, we find the following equation:

$$(3)^{5/6}\left(c^*\right)^{5/12}\left[2(1-q)\right]+12\left(c^*\right)^{11/12}\left[1-\left(\theta_L/\theta_H\right)^{12}\right]-(1-q)\left(\theta_L\right)^2\left[1-\left(\theta_L/\theta_H\right)^{12}\right]^{5/6}=0.$$

As an example, suppose that $q=0,7$; $\theta_L=2$; $\theta_H=4$. From the equation above, we find that $c^*\approx 0,0499625$, which respects the condition $1/\eta_c\left(c^*\right)\approx 0,447<q=0,7$.

Finally, we also find that $e_L^{AC}\approx 0,916$ and $\tilde{e}_L^{AC}\approx 0,000224$, which respects *IC(H)*.

REFERENCES

Ablo, Emmanuel, and Ritva Reinikka, 1998, "Do Budgets Really Matter? Evidence from Public Spending on Education and Health in Uganda," World Bank Policy Research Working Paper No. 1926 (Washington: World Bank).

Aghion, P., M. Dewatripont, and P. Rey, 1994, "Renegotiation Design with Unverifiable Information," *Econometrica*, Vol. 62, No. 2, pp. 257–82.

Ahmad, Ehtisham, and Leonardo Martinez, 2004, "On the Design of Targeted Expenditure Programs," IMF Working Paper 04/220 (Washington: International Monetary Fund).

———, Eivind Tandberg, and Ping Zhang, 2002, "On National and Supranational Objectives: Improving the Effectiveness of Targeted Expenditure Programs," IMF Working Paper 02/209 (Washington: International Monetary Fund).

Baiman, Stanley, John H. Evans, and Nandu J. Nagarajan, 1991, "Collusion in Auditing," *Journal of Accounting Research,* Vol. 29, pp. 1–18.

Baron, David, and David Besanko, 1984, "Regulation, Asymetric Information, and Auditing," *RAND Journal of Economics*, Vol. 15, No. 4, pp. 447–70.

Baron, David, and Roger Myerson, 1982, "Regulating a Monopolist with Unknown Costs," *Econometrica*, Vol. 50, No. 4, pp. 911–30.

Bayart, Jean-François, 1993, *The State in Africa: The Politics of the Belly* (New York: Longman).

Besley, Timothy, and John McLaren, 1993, "Taxes and Bribery: The Role of Wage Incentives," *Economic Journal*, Vol. 103, No. 416, pp. 119–41.

Bolton, Patrick, 1990, "Renegotiation and the Dynamics of Contract Design," *European Economic Review*, Vol. 34, Nos. 2–3, pp. 303–10.

Bouley, Dominique, Jerôme Fournel, and Luc Leruth, 2002, "How Do Treasury Systems Operate in Sub-Saharan Francophone Africa?" IMF Working Paper 02/58 (Washington: International Monetary Fund).

Chand, Sheetal K., and Karl O. Moene, 1999, "Controlling Fiscal Corruption," *World Development*, Vol. 27, No. 7, pp. 1129–40.

Dabla-Norris, Era, and Elisabeth Paul, "What Transparency Can Do When Incentives Fail: An Analysis of Rent Capture" IMF Working Paper 06/146 (Washington: International Monetary Fund).

Dewatripont, Mathias, and Eric Maskin, 1990, "Contract Renegotiation in Models of Asymmetric Information," *European Economic Review*, Vol. 34, pp. 311–21.

Dia, Mamadou, 1993, "A Governance Approach to Civil Service Reform in Sub-Saharan Africa," World Bank Technical Paper No. 225 (Washington: World Bank).

———, 1996, *Africa's Management in the 1990s and Beyond: Reconciling Indigenous and Transplanted Institutions*, (Washington: World Bank).

Diamond, Jack, 2002, "The Role of Internal Audit in Government Financial Management: An International Perspective," IMF Working Paper 02/94 (Washington: International Monetary Fund).

———, 2006, *Budget System Reform in Emerging Economies: The Challenges and the Reform Agenda*, Occasional Paper No. 245 (Washington: International Monetary Fund).

Eskeland, Gunnar S., and Henrik Thiele, 1999, "Corruption Under Moral Hazard," World Bank Policy Research Working Paper No. 2204 (Washington: World Bank).

Faure-Grimaud, Antoine, Jean-Jacques Laffont, and David Martimort, 1999, "The Endogenous Transaction Costs of Delegated Auditing," *European Economic Review*, Vol. 43, Nos. 4–6, pp. 1039–48.

Gupta, Sanjeev, and Marijn Verhoeven, 2001, "The Efficiency of Government Expenditure: Experiences from Africa," *Journal of Policy Modelling*, Vol. 23, No. 4, pp. 433–67.

International Development Association (IDA) and International Monetary Fund (IMF), 2002, "Actions to Strengthen the Tracking of Poverty-Reducing Public Spending in Heavily Indebted Poor Countries (HIPCs)" (March) (Washington).

———, 2005, "Update on the Assessments and Implementation of Action Plans to Strengthen Capacity of HIPCs to Track Poverty-Reducing Public Spending" (April) (Washington).

Kessler, Anke, 2000, "On Monitoring and Collusion in Hierarchies," *Journal of Economic Theory*, Vol. 91, No. 2, pp. 280–91.

Khalil, Fahad, 1997, "Auditing Without Commitment," *RAND Journal of Economics*, Vol. 28, No. 4, pp. 629–40.

———, and Jacques Lawarrée, 2006, "Incentives for Corruptible Auditors in the Absence of Commitment," *Journal of Industrial Economics*, Vol. 54, No. 2, pp. 269–91.

Kofman, Fred, and Jacques Lawarrée, 1993, "Collusion in Hierarchical Agency," *Econometrica*, Vol. 61, No. 3, pp. 629–56.

———, 1996, “On the Optimality of Allowing Collusion,” *Journal of Public Economics*, Vol. 61, No. 3, pp. 383–407.

Laffont, Jean-Jacques, and David Martimort, 2002, *The Theory of Incentives: The Principal-Agent Model* (Princeton: Princeton University Press).

Laffont, Jean-Jacques, and Jean-Charles Rochet, 1997, “Collusion in Organizations,” *Scandinavian Journal of Economics*, Vol. 99, No. 4, pp. 485–95.

Laffont, Jean-Jacques, and Jean Tirole, 1993, *A Theory of Incentives in Procurement and Regulation* (Cambridge, Massachusetts: MIT Press).

Lienert, Ian C., 2003, “A Comparison Between Two Public Expenditure Management Systems in Africa,” IMF Working Paper 03/2 (Washington: International Monetary Fund).

Mbaku, John Mukum, 1998, “Corruption and the Crisis of Institutional Reforms in Africa,” in *African Studies,* Vol. 47 (New York: Edwin Mellen Press).

Mookherjee, Dilip, and Ivan Png, 1989, “Optimal Auditing, Insurance, and Redistribution,” *Quarterly Journal of Economics*, Vol. 104, No. 2, pp. 399–415.

Moussa, Yaya, 2004, “Public Expenditure Management in Francophone Africa: A Cross-Country Analysis,” IMF Working Paper 04/42 (Washington: International Monetary Fund).

New South Wales (NSW) Treasury, 2000, *The Financial Management Framework for the General Government Sector*, by John Pierce (December).

Potter, Barry, and Jack Diamond, 1999, *Guidelines for Public Expenditure Management* (Washington: International Monetary Fund).

Reinikka, Ritva, and Jakob Svensson, 2004, “Local Capture: Evidence from a Central Government Transfer Program in Uganda,” *Quarterly Journal of Economics*, Vol. 119, No. 2, pp. 679–705.

République du Cameroun, Institut National de la Statistique, 2004, “Enquête sur le suivi des dépenses publiques et la satisfaction des bénéficiaires dans les secteurs de l’éducation et de la santé. Phase I : Volet Santé.” Rapport principal des résultats. (Version provisoire du 12 mars 2004).

Schick, Allen, 1997, *Modern Budgeting* (Paris: Organisation for Economic Co-operation and Development).

———, 1998, "Why Most Developing Countries Should Not Try New Zealand Reforms," *The World Bank Research Observer*, Vol. 13, No. 1, pp. 123–31.

Strausz, Roland, 2006, "Timing of Verification Procedures: Monitoring Versus Auditing," *Journal of Economic Behavior and Organization*, Vol. 59, No. 1, pp. 89–107.

Tanzi, V., 2000, "Rationalizing the Government Budget: Or Why Fiscal Policy Is So Difficult," in *Economic Policy Reform: The Second Stage*, ed. by Anne O. Krueger (Chicago: University of Chicago Press), pp. 435–52.

Tirole, Jean, 1986, "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics, and Organizations*, Vol. 2, No. 2, pp. 181–214.

World Bank, 1997, "Helping Countries Combat Corruption: The Role of the World Bank," Poverty Reduction and Economic Management (Washington: World Bank).

———, 2004a, Senegal Public Expenditure Review, Report No. 29357-SN (Washington: World Bank, PREM 4, Africa Region).

———, 2004b, "Supporting and Strengthening Supreme Audit Institutions: A World Bank Strategy," Financial Management Network, Operations Policy and Country Services (Washington: World Bank).

Ye, Xiao, and Sudharshan Canagarajah, 2002, "Efficiency of Public Expenditure Distribution and Beyond: A Report on Ghana's 2000 Public Expenditure Tracking Survey in the Sectors of Primary Health and Education," Africa Region Working Paper Series No. 31 (Washington: World Bank).