Risk Allocation and the Costs and Benefits of Public-Private Partnerships*

Elisabetta Iossa† and David Martimort‡

May 26, 2011

Abstract. We study the agency costs of delegated public service provision, focusing on the link between organizational forms and uncertainty at project implementation. We consider a dynamic multitask moral hazard environment where the mapping between effort and performance is ex ante uncertain but new information may come along during operations. Our analysis points out at the efficiency gains that bundling planning and implementation - as under Public Private Partnerships - can bring in terms of better project design and lower operational costs. Bundling also results in increasingly better performance as uncertainty is reduced by growing experience in the sector. Bundling should instead be viewed with caution when the private sector seeks to radically innovate on public service provision or to introduce new services but lacks the knowledge and expertise to anticipate the impact of the innovative design/procedure/technology on the cost of operations. The compounding of asymmetric information ex post plus moral hazard and renegotiation may generate diseconomies of scope in agency costs which, for high operational risk, can make unbundling optimal. In this context, the use of private finance can help re-establishing the benefit of bundling only if lenders have sufficient expertise to help assessing project risks.

1 Introduction

Background. The realization of most public projects requires two main sequential stages: the planning stage where the project is designed, and the implementation stage where the project is executed. Under a public-private partnership (hereafter abbreviated as PPP), the supplier takes responsibility for both the building of the infrastructure and its managing and maintenance. The DBFO model ('Design', 'Build' 'Finance' and 'Operate'), the BOT model ('Build', 'Operate' and 'Transfer') or the BOO ('Build', 'Own' and 'Operate') are all common contractual modes that feature bundling of building and operation in a single contract with a single firm (or consortium of firms). PPPs are used across Europe, Canada, the U.S. and a number of developing countries for the provision

*For useful comments or discussions, we with to thank Malin Arve, John Bennett, Bernard Caillaud, Jérôme Pouyet, Emile Quinet, Timo Välillä and seminar participants at ESNIE (Ajaccio), at the Congrès de l' Association Française d'Economie (Paris-Sorbonne), Université de Franche-Comté and at the UBC P3 Project (Vancouver). Jim Hosek and two referees were instrumental in helping us to reshape our ideas on this project. We thank Perrin Lefebvre for exceptional research assistance. All errors are ours.

†SEFEMEQ, University of Rome Tor Vergata and Brunel University, CMPO, EIEF. Email: Elisabetta.Iossa@uniroma2.it. This author wishes to thank financial support from the Ministry of Education, University and Research (MIUR), PRIN grant 2008.

‡Paris School of Economics-EHESS. Email: david.martimort@parisschoolofeconomics.eu.

of public infrastructures and services in sectors such as transport, energy, water, IT, prisons, waste management, schools, hospitals and others. Transport projects for toll roads, ports and rail typically follow the 'concession model', where the private provider recoups (part or all of) its initial investment through charges to final users. In PPP hospitals, schools and prisons, instead, where users typically do not pay; it is the public sector that pays the private sector party for the service that it provides to users. This is known as the 'PFI model'.

When principals delegate to agents the tasks of planning and implementing projects, both moral hazard and adverse selection problems may arise. At planning stage, the agent may have to be motivated to efficiently design the project. At implementation stage, the agent may have to be motivated to exert effort in managing the project and to use efficiently the information that comes along during operations. In this paper, we study how uncertainty at implementation stage affects the design of compensation schemes for the agents, and we derive implications for the optimality of PPP agreements. Uncertainty is important not just because the agent maybe risk averse but also because it affects the degree of contract incompleteness.

Consider the changes that we have witnessed in the last twenty years to the way public services are provided. A central theme in both developed and developing countries has been an increasing level of delegation to the private sector for the provision of public services. Not only have traditional services such as transport, energy and gas been increasingly privatized, but new or more complex services have been contracted out for which there is no precedent in the private sector. Europe for example is experiencing with the delegation of prisons operations, of school design and maintenance, of new methods for waste disposal such as recycling plants, and with an increasingly use of transport concessions for toll roads, rail, ports and bridges. When the UK Government first outsourced the electronic monitoring services for the Home Detention Curgew (HDC) scheme in the mid 1990s the service was so new that there was no precedent for the service model (Serco Institute 2007).[1] When the UK National Audit Office (NAO) reviewed the success of the three highly complex PPP deals to finance and manage London's Underground system, two years after the first contract was signed, and in spite of a period of 'shadow running' to review arrangements and make adjustments, they found it *"hard to determine whether [the measures were] easy or difficult to achieve, and whether they [were] sufficiently aggressive"* (NAO, 2004).

Overview. We consider a simple model of procurement in a dynamic multitask moral hazard environment where the mapping between effort at design stage and performance at implementation stage is ex ante uncertain but new information may come along during operations.[2,3] We compare a setting where only one agent is in charge of both planning and implementation with one where these two tasks are allocated to two different agents. We start by considering the case of projects where uncertainty is pervasive in the sense that neither the public authority nor the contractor can predict ex ante all contingencies (productivity shocks) that may arise during operations. Incentives are provided through revenues sharing agreements when users pay (concession model), or through service standards agreements when users do not pay (PFI model), and through an infrastructure quality index. We show that design effort is always greater under bundling because the firm takes into account the positive externality that a better design generates on the marginal cost of operational effort. And with a

[1] The HDC scheme allowed prisoners to live outside prison custory under curfew rules; they would be fitted with electronic monitoring devices that will initiate an alert if they broke their curfew conditions.

[2] This modeling device distinguishes our approach from others in the literature. See Aghion and Tirole (1994), Tirole (1999) and Rosenkranz and Schmitz (2003) for incomplete contracting models of innovation which rely on the perfect knowledge of the mapping between research and development.

[3] See Holmström and Milgrom (1991) for the seminal model on multitasking in a moral hazard context. Multitasking is one feature that makes the analysis of PPPs arrangements quite specific compared with the existing more general literature on privatization. This literature analyzes the agency cost of delegation to the private sector in a framework where a single task has to be performed. See the seminal papers by Sappington and Stiglitz (1987) and Shapiro and Willig (1990) and the overviews by Shleifer (1998) and Martimort (2006).

lower cost, operational effort increases to raise the revenues from the provision of the service, or to raise service standards for services where users do not pay. Therefore, operational costs are also lower under bundling. Bundling has the greatest benefits in terms of innovative design and lower costs when demand risk is low, when service standards can be measured precisely, when the infrastructure quality index is imprecise, and when the technology is very flexible, in the sense that it easily adapts to unforeseen circumstances.

We then extend the analysis to projects where uncertainty is less pervasive and more complete contracting is possible, productivity shocks being verifiable ex post. This can be thought of as a metaphor for a second stage of private provision of public services, where sufficient past evidence is available to inform the parties of what may arise during the operational stage. The contract will then compensate the contractor also for adverse events beyond his control. This insurance improves the noise-to-signal ratio at the operating stage and reduces the risk premium due to provide incentives. With a more detailed and efficient risk allocation, the advantages of bundling in terms of greater design effort and lower operational costs are boosted.

The lesson of these findings is that bundling certainly dominates when contracting takes place under symmetric (although possibly incomplete) information over productivity shocks. Significantly different results are obtained for complex projects where productivity risks are high but their realizations are private information of the firm at the operating stage. With asymmetric information, compensating the firm for adverse events beyond its control becomes more complex, as the operator always wants to claim to be hit by a negative shock in order to receive a compensation. Under bundling, the problem is exacerbated since the firm may also undersupply the first-stage design effort and always claim to be hit by a bad shock. Preventing this "double deviation" calls for even less insurance at operating stage and a contract that is not time-consistent. With renegotiation becoming ex post attractive, ex ante incentives are further weakened. We then show that when uncertainty on productivity shocks is high, compounding asymmetric information ex post, moral hazard and renegotiation results in such lack of insurance that the gain at design stage due to the internalization of externalities becomes insufficient to compensate for excessive risk taking. Unbundling then dominates.

In DBFO projects, the bundling of design and implementation is accompanied by the use of private finance, which typically covers about 90% of project costs, the remaining 10% being equity. Whilst private finance is in general more costly than public finance, the due diligence carried out by the contractor's banks may sometimes stop contractors taking on too much inappropriate risk. The lender may for example have the expertise to assess risks carefully, improving the quality of the business plans, and ensuring that contingency plans are in place to manage risks appropriately. Our paper further shows that should the financier have such expertise and should the government have not, private finance increases the efficiency of the PPP, and in the presence of high risks this improvement can justify the higher cost of private capital. Practitioners do acknowledge that the record of conventional public sector procurement in assessing risk in not good (Leahy, 2005), and rating agencies such as MOODY's (2007) explicitly recognize the importance of the quality of the contractor's financial model - typically audited by lenders - to calculate operational risk in PPP projects.[4]

Evidence. Empirical observations confirm the importance of externalities across the design and the implementation stage of a project. The innovative design of PFI prisons in the UK was found for example to have brought cost saving of 30% (NAO 2003). The innovative design in PFI motorway projects combined with high modulus road bases and stone mastic asphalt (SMA) was found to reduce traffic disruption from maintenance, as well as lower costs, time and raw materials. In line with our predictions, PPPs have also found to result in increasingly better performance as uncertainty

[4]Lenders in PPP projects (directly or through independent consultants) typically undertake an audit of the financial model of the contractor, for example to ensure that the financial model has been prepared within accepted accountancy standards and that the calculations have been accurately formulated and are fully consistent with the assumptions specified in the project documents.

is reduced by growing PPP experience in a sector, and by appropriate training of the public officials involved in the PPP contract design and management. In the UK PPP housing sector, for example problems initially arose from the complexities in obtaining planning permissions for multiple sites and the difficulties in estimating costs for the refurbishment of properties at the outset of procurement but lessons were learnt from pathfinder projects and the injection of public sector expertise (HM Treasury 2006). Knowledge accumulation has also been favored by the setup of dedicated PPP government units in various countries to create centres of expertise on PPP projects (e.g. Central PPP Policy Unit in the Department of Finance in Ireland, the Unita' Tecnica della Finanza di Progetto in Italy, Partnership UK in the UK), and by the diffusion of best practices on risk assessment and risk allocation.

Our results also explain the skepticism on the use of PPPs for complex projects where risks are high and uncertainty is pervasive. In complex IT projects, the link between the effort the operator puts in handling the system and the system performance is certainly more uncertain than in conventional IT systems. Given the peculiarities of each and every scheme, there is a significant informational advantage that the operator gets over the public authority. With the public authority unable to verify why software customization was more time consuming than anticipated, the contractor will have to be responsible also for events beyond its control. For very complex projects, risk transfer becomes so costly that PPPs turn unsuitable. The UK has indeed recommended against using PPPs for complex IT projects. Also concerns have repeatedly been expressed on the performance of PPPs in new sectors such as waste and recycling, where the cost of disposal can vary dramatically and no past experience is available to inform the parties, and for complex and innovative clinical services where technology changes fast and risks are high.

Our results suggest that renegotiation plays an important role in determining the cost and benefit of PPPs. In practice, numerous instances have been recorded where PPP contracts have been renegotiated. In the UK, renegotiations occurred in 33% of Central Government Departments PFI projects signed between 2004 and 2006. The changes amounted to a value of over £4m per project per year equivalent to about 17% of the value of the project (NAO, 2007). Illustrative is also the case of specialized IT provision where the appropriate use of the facility involves continuous adaptation. Following performance failure and costly contract renegotiation, the HM Treasury in the UK now recommends against the use of PPPs for IT projects (see HM Treasury 2006). Renegotiation by the government of concession contracts in Latin American and Caribbean Countries is also widespread. Considering a compiled data set of more than 1,000 concessions granted during 1985–2000, Guash (2004) showed that 30 % of the concessions were renegotiated (see also Guash, Laffont and Straub, 2008).

Literature review. Our model belongs to a burgeoning literature that discusses the costs and benefits of bundling tasks in contexts plagued with transaction costs, agency costs or contract incompleteness taken separately or in tandem. That literature can roughly be decomposed into two trends which, although different in motivation and scope, share the common finding that bundling tasks arise when contracting costs exhibit scope economies.

The first subset of the literature builds on the incomplete contracts and property rights paradigms. Hart (2003) provides a model of PPPs where the sole source of incentives is asset ownership. A builder can perform two kinds of investment (productive and unproductive) which may both reduce operating costs, although only the productive investment raises also the benefit of providing the service. Under traditional procurement, the builder cannot internalize the impact of his effort neither on benefits nor on costs and, as a result, implements too little of the productive investment but the right amount of the unproductive one. Under PPP, the builder partly internalizes the impact of his productive investment whereas he also exerts too much of the unproductive one.[5] Bennett and Iossa (2006)

[5]Turning to the case where ownership concerns a public good and still using the property rights approach, Besley and Ghatak (2001) show that ownership should lie in the hands of that player with the highest valuation for the public good. On similar issues, see also King and Pitchford (2001) and Francesconi and Muthoo (2006).

study the desirability of bundling different phases of a project and of giving ownership to the investor. Innovations are non-contractible ex ante but verifiable ex post. Ownership of the asset gives control right to the owner to decide whether to implement quality enhancing or cost-reducing innovation proposed by the investor. The hold-up problem is less severe under PPP, compared with traditional procurement, when there is a positive externality between the building and managing stages. With a negative externality the opposite can hold.

Externalities between various production stages are also the focus of a branch of the agency literature, which beyond the specific case of PPPs, addresses organizational issues related to task bundling in moral hazard environments. Martimort and Pouyet (2008) build a model for PPPs with moral hazard. The agency costs due to the trade-off between insurance and incentives are lower under a PPP when there is a positive externality between building and managing assets; a better design reduces costs at the implementation stage. The argument is standard from the multitask literature:[6] providing incentives on one task makes incentives on the other cheaper.[7] Iossa and Martimort (2008) build on these insights, extending the analysis to cover other issues in PPPs, from contract duration, to private finance and regulatory risk.

We significantly depart from the analysis in Martimort and Pouyet (2008)and Iossa and Martimort (2008) by focusing on an environment where costs are non-verifiable and introducing an uncertain synergy between design and implementation.

Other benefits of bundling allowing for reciprocal effort supervisions, improved risk-sharing or improved incentives in ongoing relationships have been found in various other contexts by Che and Yoo (2001), Choi (1993), Macho-Stadler and Perez-Castrillo (1993), Itoh (1994), and Ramakrishnan and Thakor (1991).[8] Certainly closer to us is Schmitz (2005) who studies how the control of sequential tasks should be allocated in a pure moral hazard problem but with limited liability as a source of the agency problems. He considers the possibility that the second-stage project can be more or less successful or more or less costly depending on the first-stage effort while keeping the assumption of a common knowledge mapping between design and implementation. The prospects of getting liability rent in the second-stage are a powerful engine for first-stage incentives.[9] Benz, Grout and Halonen (2001) present a model of PPP which mixes elements of private information (whoever builds an asset knows how efficient it can be to provide the service) and moral hazard (service can be improved by effort) but in very different ways than ours. Under bundling, the adverse selection rent is again used to provide cheaper effort incentives. Bennett and Iossa (2010) study the scope for PPPs when not-for-profit firms (NPs) are involved. Using an incomplete contract approach, they show that bundling helps reduce overinvestment in building quality by NPs. Auriol and Picard (2011) compare traditional procurement with PPPs, focusing on the role of the shadow costs of public funds when the information structure changes with the organizational form.

Finally, Lewis and Sappington (1997), Khalil and al. (2006) and Krähmer and Strausz (2009) investigate the value of bundling planning and implementation in a context where the agent needs to acquire information. Such models feature also some interaction between an ex ante moral hazard and an ex post revelation problems but different from ours.[10]

[6]Holmström and Milgrom (1991).

[7]Martimort and Pouyet (2008) show also that ownership is an imperfect way of aligning incentives. To a large extent, the important issue they stress is not who owns the asset but instead whether tasks are bundled or not.

[8]The benefits for an organization of bundling tasks have also investigated in pure adverse selection environments. Those papers sometimes allow collusion between agents, delegated sub-contracting, various asymmetries and information structures, and imperfect substitutability/complementarity between tasks (Baliga and Sjöström, 1998, Baron and Besanko, 1992, 1999, Gilbert and Riordan, 1995, Laffont and Martimort, 1998, Dequiedt and Martimort, 2004, Mookherjee and Tsugamari, 2004, Severinov, 2008). Complementarity between tasks whose costs are privately known by agents creates scope economies in agency costs. In a nutshell, each piece of private information creates a wedge between marginal benefit and marginal costs which induces some inefficiency. Moving from having two agents each dealing with one task to a merger avoids a "double marginalization".

[9]On this, see also Laux (2001) although this paper is not concerned with organizational design per se.

[10]Also related is the work of Gromb and Martimort (2007) who analyze the organization of experts acquiring infor-

Organization of the paper. Section 2 presents our model. Section 3 analyzes optimal incentive schemes under either unbundling or bundling in our basic scenario where uncertainty is pervasive and productivity shocks that may arise during the operational stage cannot be predicted ex ante. This section provides also our main result on the benefits of bundling tasks. Section 4 considers more complete contracting, studying the case where shocks are verifiable and thus can be contracted upon, showing that the benefits of bundling are boosted. In Section 5, the analysis covers the central case where the firm has private information on the productivity shocks and shows that now bundling may generate diseconomies in agency costs. Section 6 introduces a new player in the model, the financier, and studies the costs and benefits of private finance. Section 7 summarizes our findings and provides alleys for further research. Proofs are relegated to an Appendix.

2 The Model

We consider the following public procurement context: A public authority (the principal) relies on a private firm or consortium (the agent) to provide a public service. Examples of such delegation include of course transportation, water production and sanitation, waste disposal, clinical and educational services etc. In such settings, the provision of the service requires also to build and design an infrastructure. This delegation of services to the private sector is modeled as a multi-task problem.

In the sequel, we discuss the rationale for relying on a PPP (bundling) rather than adopting a more traditional procurement model. With such traditional contracting, the principal first buys the infrastructure from a given builder and then selects an operator. We thus investigate whether the two tasks of respectively designing/building the assets and operating them should be bundled and performed by the same contractor (a consortium) or instead whether they should be unbundled and undertaken by two separate firms (a builder and a separate operator).

Technology. Investing in the infrastructure initially costs I. In a first building stage, the builder chooses a design. By exerting effort a (or, by making some investment in the quality of the infrastructure), the contractor improves the quality of the project and raises its social value. Investing a has a monetary cost $\frac{a^2}{2}$ which is borne by the firm.

For services where users pay (concession model), the revenues from the service provision are stochastic and defined as

$$R = e + \zeta$$

where e is the operating effort and ζ is a shock normally distributed with zero mean and variance η^2. The effect of e on R captures for example the higher demand from users of transport services when service reliability, on-the train services, or the efficiency of the ticketing system are higher. One can interpret ζ as a demand shock. In transport concessions, for example, even when there is a reasonable level of confidence in forecasts, demand can be dramatically affected by competition from other modes or facilities, from the conditions affecting the wider network, such as economic activity levels or tourism demand, and from the price of inputs (e.g. fuels).

To simplify exposition, we shall use R interchangeably to denote also the benefit from the service, for those services where users do not pay (PFI model). Here, e captures for example the operational efficiency of hospitals, the attitude of prisons staff towards inmates, or the quality of education and management in schools, whilst ζ captures the uncertainty affecting social benefits. In school projects, for example, uncertainty comes from socioeconomic conditions and cultural factors affecting the social value of education, the availability of private education, and birth rates affecting the school population.

The operating effort e improves benefits but its provision is costly for the operator. This cost can be a disutility of effort counted in monetary terms or a true operating cost. It is expressed as

mation in a first stage and revealing it in a second one.

$\frac{\mu}{2}(e - a - \theta)^2$ where a is the design effort, θ is a productivity shock and μ is a parameter that reflects the flexibility of the technology at the operating stage; μ being greater as the technology becomes less flexible. Except in Sections 4 and 5, we shall assume that θ is also normally distributed with zero mean and variance ϑ^2. The operator chooses the effort e *ex post*, i.e., once he already knows θ. This simple formulation allows us to capture how the firm can adapt its second-stage efforts to productivity shocks in the environment.

Some justifications are in order. First, a greater design effort a reduces the cost of operating effort. For example, better project design simplifies operations, better transport infrastructure reduces maintenance effort, and good school facilities (labs, gyms, theaters etc.) make it easier to motivate students and improve educational attainments. Second, the random variable θ in the cost of operating effort captures the uncertainty that may characterize the mapping between design and operation. When a project introduces a new service or an innovative project design greatly modifies existing procedures, the changes may be so dramatic that no established point of reference may be available to predict performance.[11] Third, the variance of θ, ϑ^2, captures operational (cost) risk, namely the effect on the cost of operations of exogenous factors, such as force majeure events, inflation, or changes in specific input prices.

Contracts and organizational forms. The revenue R is verifiable and can thus be contracted upon. In transport projects, for example, revenues can be verified through electronic ticket systems, whilst in energy projects they can be specified through computerized billing systems, and revenue sharing agreements between the public authority and the contractor are widely employed. For projects where users do not pay (PFI model), service standards (R) are verified using key performance indicators and audit, complaints received, and user satisfaction surveys.[12] Whilst R is verifiable, the government cannot disentangle the impact of outside variables ζ and the effect of operational effort e on revenues.[13]

Costs incurred both at the design and operation stages are nonverifiable.[14] However, the principal has also at his disposal a contractible quality index Q to check whether the quality of the infrastructure is good enough. This quality index is directly related to the first-stage effort and writes as

$$Q = a + \epsilon$$

where the error term ϵ is again normally distributed with zero mean and a variance σ^2 which captures the precision of the quality index.

Contracts based on revenues and quality index can be used to regulate the services.[15] Following Holmström and Milgrom (1987, 1991), we assume that contracts are linear in the contracting variables and we refer to the slopes of those contracts as the incentives intensities.

[11] The extensive use of pilot projects to experiment with new approaches before deciding whether to adopt them more broadly provides an additional example.

[12] See the study on operational PPP contracts by PUK (2008). Contracts specify a number of performance measures (typically 30/40 but sometimes up to 200; Serco, 2007) and the number of points that each underperformance/overperformance episode will attract. The accrued number of performance points is then used to adjust the contractor's payment (the so-called "availability payment") at the end of a period, usually either a year or a quarter.

[13] For example, the punctuality of trains and transport services is affected by weather or traffic conditions and the operational efficiency of contractors but the effects cannot be perfectly disentangled; the spreading of infections in hospitals is affected in non-deterministic way by both the distance between beds and the cleanliness of the hospitals. In prison services, one of the key performance indicators is the occurrence of re-offending, which is affected by factors outside the control of the prison contractors but tend to improve with the attitude of prison staff towards inmates.

[14] The non-verifiability of costs is a key feature of complex projects like PPPs. This is especially true in developing countries as noticed by Laffont (2005) and Estache and Wren-Lewis (2009).

[15] For instance, infrastructure quality indicators in the contract for London Underground included: (i) capability, a measure of the capacity of the infrastructure, capturing the average journey time; (ii) availability, a measure of the reliability of rolling stock, signaling, track, and station-based equipment; (iii) ambience, a measure of the quality of the environment for passengers, including the condition and cleanliness of trains and stations and the provision of passenger information (NAO, 2004).

Under *unbundling*, we assume that the operator does not observe the builder's effort but perfectly anticipates its value at equilibrium. The builder is rewarded as a function of the realized quality index and operates under a linear scheme based on a quality index only:[16]

$$t_B(Q) = \alpha_B + \beta Q.$$

The operator keeps a share of the operating revenue and operates under the following linear scheme:

$$t_O(R) = \alpha_O + \gamma R.$$

The fixed payments α_B and α_O can be interpreted as the upfront fee bid at the tender stage for the right to provide the service (in which case they are negative), or as ex post payments made by the government (in which case they are positive). We will use both interpretations in what follows. α_O can also be interpreted as an availability payment paid by the government for services where users do not pay. The incentives intensities of these contracts are captured by the parameters β and γ.

Remark 1 *Those schemes exhaust all contracting possibilities. Following Holmström (1979)'s "Informativeness Principle", and since the first-stage investment does not affect revenues, there is no point rewarding the builder as a function of revenues or the operator according to the quality index.*

Consider now the case of *bundling.* A conglomerate operates now under the linear scheme

$$t(Q, R) = \alpha + \beta Q + \gamma R.$$

Under all circumstances below, the outside opportunity of the firm(s) gives an exogenous payoff normalized at zero. Contracts are also designed ex ante, i.e., before productivity shocks realize.

Objectives. The principal is risk-neutral and maximizes his share of revenues net of the costs of paying for the services and the infrastructure. Formally, under unbundling, this objective is

$$R - t_B(Q) - t_B(R) - I$$

whereas, under bundling, it becomes

$$R - t(Q, R) - I.$$

Those expressions presume that the principal pays by himself the investment and appropriates all revenues. Of course, it is only an accounting convention to have the builder invest I, and then be paid $t(Q, R) + I$.

Firm(s) are (is) risk-averse with constant degree of risk-aversion $r \geq 0$, and we denote $v(x) = 1 - exp(-rx)$ this utility function.[17] Firms care about the expected net return. Under unbundling, the builder and the operator respectively maximize

$$E\left(v\left(t_B(Q) - \frac{a^2}{2}\right)\right) \text{ and } E\left(v\left(t_R(R) - \frac{\mu}{2}(e - a - \theta)^2\right)\right).$$

[16]In theory the builder could also be paid according to the realized revenues from the facility. However, as Remark 1 below emphasizes, this would have no effect on the builder's incentives in our context. Moreover, in practice, such payments would be difficult to implement as the operational contracts extend to many years after the facility is built.

[17]The assumption of risk-aversion for the firm(s) captures the fact that a PPP project might represent a large share of the firms' activities so that they can hardly be viewed as being diversified. Instead, assuming risk-neutrality for the government gives a simple benchmark: In the absence of moral hazard, optimal risk allocation requires the public sector to bear all risks. This assumption may be questioned in the case of small local governments whose PPP projects may represents a significant share of their overall budget. For a large country, the existing deadweight loss in the cost of taxation may as well introduce a behavior towards risk if the PPP project were to represent a large share of the budget. Lewis and Sappington (1995) and Martimort and Sand-Zantman (2007) analyze the consequences of having risk-averse governments in procurement settings.

Under bundling, the conglomerate's objective becomes

$$E\left(v\left(t(Q,R)-\frac{a^2}{2}-\frac{\mu}{2}(e-a-\theta)^2)\right)\right)$$

where $E(\cdot)$ is the expectation operator.

Contractible efforts and shocks. When the design and operating efforts are both contractible and all shocks perfectly insurable, the efficient levels of effort maximize social welfare. Whether bundling or unbundling is chosen is clearly irrelevant and the firm is fully insured against all shocks. Let $(a^*, e^*(\theta))$ be those efficient levels of effort. They solve:

$$\max_a \left\{-\frac{a^2}{2}+E_\theta\left(\max_e e-\frac{\mu}{2}(e-a-\theta)^2\right)-I\right\}.$$

We find

$$a^*=1 \text{ and } e^*(\theta)=a^*+\frac{1}{\mu}+\theta$$

Of course, the operating effort depends on the design effort but is biased upward, and the less so as the technology is less flexible. In the limiting case where $\mu \to \infty$, the design fully determines the operating effort up to the ex post shock θ. This highlights the existing *positive externality* between the two stages: improving design also increases operating revenues.[18]

Finally, only projects with positive value are undertaken, which imposes an upper bound I^* on the feasible investments:

$$I \leq I^*=\frac{1}{2}\left(1+\frac{1}{\mu}\right).$$

Remark 2 *Our framework could be modified to capture the possibility of a negative externality between design and operation.*[19] *Suppose that the disutility of effort at the operating stage writes as*

$$\frac{\mu}{2}(e+a-\theta)^2.$$

The operating effort moves thus in opposite directions with the design effort. Keeping the same expression of revenues as above, there would be no benefit in investing in a good design. Alternatively, if revenues were directly an increasing function of the first-stage effort and, for instance, write as $R=\lambda a+e+\epsilon$ *with* $\lambda>1$*, or if a good design was creating positive externality whose social value to the principal was* $S=\lambda a$*,*[20] *the optimal first-stage effort would remain positive.*[21]

3 Organizational Forms under Incomplete Contracting

In public service projects, uncertainty may be pervasive. The service may have typically been provided by the public sector and at the time of delegation no data about delivery capabilities and past

[18]Interestingly, once the operator knows θ, his expected revenues $a+\theta$ depend on both the first-stage effort and the productivity shock. This randomness will also be present when efforts are non-verifiable. Anticipating on our findings in the sequel, since only a share of the revenues is kept by the operator, the fluctuations of his payoff will not fully reflect fluctuations of productivity shocks.

[19]For example, an innovative design of a hospital, using recently-developed materials, may lead to improved lighting and air quality, and therefore better clinical outcomes, but may also increase maintenance costs.

[20]For example, an innovative design of a school using open plans may lead to a more welcoming, inclusive learning environment. An innovative design of a prison, with a gym, a workshop or an education block, can help inmates train for a job and reduce the percentage of re-offenders.

[21]Because of a space constraint, we leave the study of those more complex patterns of externalities for future research.

performance may be available.[22] The project may also introduce new services and procedures that make it difficult to foresee all contingencies that may arise at the implementation stage.[23]

To analyze the benefits of bundling for such projects, let us first assume that the productivity shocks θ cannot be contracted upon ex ante. Only ex post, will the realization of θ become common knowledge. At that stage, the operator adapts its second-stage effort to its realization, but the payment is not adjusted to capture the incompleteness of the contract.[24]

Unbundling. Consider first the builder's incentive constraint. The design effort maximizes the certainty equivalent of the builder's profit:

$$a = \arg\max_{\tilde{a}} \alpha_B + \beta\tilde{a} - \frac{\tilde{a}^2}{2} - \frac{r\sigma^2\beta^2}{2} \equiv \beta. \tag{1}$$

The builder's participation constraint written in terms of certainty equivalents becomes:

$$U_B = \alpha_B + \left(1 - r\sigma^2\right)\frac{\beta^2}{2} \geq 0. \tag{2}$$

Consider now the operator. As he knows θ and perfectly anticipates the builder's first-stage effort level, his incentive constraint writes as

$$e(\theta, \gamma, a) = \arg\max_{\tilde{e}} \alpha_O + \gamma\tilde{e} - \frac{\mu}{2}(\tilde{e} - a - \theta)^2 - \frac{r\eta^2\gamma^2}{2} \equiv a + \theta + \frac{\gamma}{\mu}. \tag{3}$$

The second-stage effort increases with the productivity shock θ, and the incentives intensity γ. This expression also shows that the operator's effort responds to productivity shocks and design as in the first-best.[25] However, since the operator only keeps a share of revenues, only a fraction of the productivity risk impacts on his expected payoff. Using (3), the certainty equivalent of this payoff in state θ becomes $\alpha_O + \gamma e(\theta, \gamma, a) - \left(\frac{1}{\mu} + r\eta^2\right)\frac{\gamma^2}{2}$. Taking into account the extra risk-premium coming from uncertainty on productivity shocks, the operator's participation constraint becomes (where we make explicit the dependence on a)

$$U_O(a) = \alpha_O + \gamma a + \left(\frac{1}{\mu} - r(\eta^2 + \vartheta^2)\right)\frac{\gamma^2}{2} \geq 0. \tag{4}$$

Proposition 1 *Suppose that θ is non-verifiable. The optimal scheme under unbundling entails the following incentive intensities and first-stage effort level:*

$$\gamma_U = \frac{\frac{1}{\mu}}{\frac{1}{\mu} + r(\eta^2 + \vartheta^2)} < 1 \text{ and } \beta_U = a_U = \frac{1}{1 + r\sigma^2} < 1. \tag{5}$$

Those results are quite standard but their implications in our context are interesting. The incentives intensity γ_U captures the transfer of demand risk to the operator. Higher demand risk transfer

[22] Focusing on such uncertain environments fits well with the observation made by Bajari and Tadelis (2001) that, in many procurement contexts, buyers and sellers face the same uncertainty on costs and demand conditions beforehand.

[23] This uncertainty often justifies for example the use of pilot projects, for the public sector to experiment with a new organizational form or a new project before committing its policy to it.

[24] This assumption is justified even if there is a renegotiation ex post as long as the principal keeps all bargaining power at this stage.

[25] See Backer and Johnson (2003) for another interesting moral hazard model with the agent choosing his effort after the realization of some shock.

(higher γ) raises the operator's incentives, but at the cost of a higher risk premium ($r(\eta^2 + \vartheta^2)\gamma^2/2$) to compensate the operator for all the risks he bears. This includes uncertainty on revenues and on the productivity shock θ. When demand risk increases (η^2 higher) or when operational uncertainty is higher (ϑ^2 higher), transferring demand risk becomes more costly (the risk premium increases) so that less risk transfer and weaker incentives become optimal. Similar insights apply with regard to the power of the incentives scheme for the builder. As the precision of the quality index decreases (σ^2 increases), incentives are reduced to decrease the risk premium ($r\sigma^2\beta^2/2$). A less flexible technology (μ greater) also corresponds to a lower powered revenues sharing agreement. It becomes indeed harder to provide incentives at the operating stage when the corresponding effort is almost fully determined by the first-stage design.

Bundling. Consider now the case of bundling. The second-stage incentive constraint remains as in (3). Turning now to the first-stage incentive constraint, the conglomerate anticipates the impact of design on second-stage effort and revenues. More precisely, the first-stage effort now solves:

$$a = \arg\max_{\tilde{a}} \alpha + \beta\tilde{a} - \frac{\tilde{a}^2}{2} + E_{\tilde{\theta}}\left(\gamma e(\theta,\gamma,a) - \frac{\mu}{2}(e(\theta,\gamma,a) - a - \theta)^2\right) - \frac{r(\eta^2+\vartheta^2)\gamma^2}{2} - \frac{r\sigma^2\beta^2}{2}$$

where the last terms represent the risk premiums associated to all risks borne by the conglomerate.

Using $e(\theta,\gamma,a)$ from (3) yields another expression of the first-stage incentive constraint as

$$a = \arg\max_{\tilde{a}} \alpha + (\beta+\gamma)\tilde{a} - \frac{\tilde{a}^2}{2} + \left(\frac{1}{\mu} - r(\eta^2+\vartheta^2)\right)\frac{\gamma^2}{2} - \frac{r\sigma^2\beta^2}{2} \equiv \beta + \gamma. \tag{6}$$

Incentives intensities at both stages contributes to raise the first-stage effort.

The conglomerate's participation constraint written in terms of certainty equivalents becomes:

$$U(a) = \alpha + (\beta+\gamma)a - \frac{a^2}{2} + \left(\frac{1}{\mu} - r(\eta^2+\vartheta^2)\right)\frac{\gamma^2}{2} - \frac{r\sigma^2\beta^2}{2} \geq 0. \tag{7}$$

Proposition 2 *Suppose that θ is non-verifiable. The optimal scheme under bundling entails the following properties.*

- *Incentives intensities on profits (resp. quality) are greater (resp. lower) than under unbundling:*

$$1 > \gamma_B = \frac{\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2}}{\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2} + r(\eta^2+\vartheta^2)} > \gamma_U, \tag{8}$$

$$1 > \beta_U > \beta_B = \frac{1}{1+r\sigma^2}\frac{r(\eta^2+\vartheta^2)}{\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2} + r(\eta^2+\vartheta^2)}. \tag{9}$$

- *The first-stage effort a_B is greater than under unbundling but still lower than the first-best:*

$$a_B = \frac{1+\gamma_B r\sigma^2}{1+r\sigma^2} \in (a_U, 1). \tag{10}$$

Under bundling, the principal can rely on both β and γ to provide incentives for first-stage effort. The first instrument is *direct*, whereas the second is *indirect*. A revenues sharing scheme indeed rewards the merged entity for his operational effort which is indirectly positively linked to the first stage effort. Because of this positive externality, agency costs exhibit economies of scope. The power of incentives is greater under bundling and the agent keeps a larger share of revenues. At the same time, relying on a quality index becomes less attractive which highlights a substitutability between direct and indirect rewards for first-stage effort. Keeping a greater share of revenues in turn boosts the design effort because the firm takes now into account the impact of a better design on operational costs. And with lower operational costs, the second-stage effort increases which raises revenues. From the definition of operational efforts under both organizational forms, we indeed obtain:

$$e(\theta, \gamma_B, a_B) > e(\theta, \gamma_U, a_U) \quad \forall\theta. \tag{11}$$

Importantly, the incentives intensities β and γ move in opposite directions. A more noisy quality index means that the principal relies less on that index to provide first-stage incentives. But as a substitute, the firm should keep more of its revenues. On the contrary, when revenues are more risky, there is less point in using revenues to provide first-stage incentives. By the same token, as the technology becomes more flexible (μ decreases), γ_B increases whereas β_B decreases. The principal relies more on revenues sharing since the conglomerate's operational effort becomes more responsive to second-stage incentives.[26]

Welfare comparison. The comparison between organizational forms is now straightforward.

Proposition 3 *Bundling is always the optimal organizational form.*

Related results have been found in the existing agency literature although in slightly different contexts (see our review of the literature). Our novel finding is that this welfare comparison takes now into account uncertainty in the mapping between first-stage and second-stage effort at the time of making the choice on the optimal organizational form. Nevertheless, the strong positive relationship between investment and operational efforts justifies an unambiguous choice of bundling. This suggests a potential gain from PPPs during a first stage of delegation to the private sector, where uncertainty is pervasive and insufficient past evidence is available to inform parties of what may arise later at the operational stage.

Bundling has the greatest benefits when $\sigma^2 = +\infty$, i.e., with no useful quality index. Indeed, with separate entities, there is no way to incentivize the builder. Instead, the conglomerate takes into account the impact of design on revenues and exerts a positive first-stage effort. Similarly, when μ goes to zero and the technology is very flexible, the positive externality across stages is large. The revenues sharing scheme is high powered (γ close to one), and again bundling has the greatest benefits.

The gains from bundling are also high when demand risk is low, service standards can be measured precisely (η^2 is low), or when operational risk is low (ϑ^2 is low) so that the firm keeps most revenues and better internalizes externality across stages (γ is high). To illustrate, PPPs may be beneficial for standard transport projects, such as toll roads and motorways, where traffic information is well documented. Social houses and public accommodations also show low demand risk although there reliable infrastructure quality indexes might also be available. Traditional procurement with unbundling may instead be preferable for new toll road projects in less developed areas, or for innovative transport projects, where both demand and operational risks are high.[27]

[26] When the technology becomes non-flexible (the limiting case where $\mu \to +\infty$), revenues are just a noisy measure of first-stage effort, namely:

$$R = a + \theta + \zeta.$$

Holmström (1979)'s "Informativeness Principle" then calls for using that observable to also reward the firm for its first-stage effort on top of the quality index. In that limiting case with no flexibility, revenues do not provide any incentives for the operational effort which is fully determined by the design stage.

[27] For example high uncertainty in demand forecast characterized the highly innovative Channel Tunner Rail Link connecting England to France (see NAO 2007).

Finally, bundling facilitates initial investment into the project. Indeed, the condition for a positive net present value project under bundling becomes:

$$I \leq I_B = \frac{1}{2(1+r\sigma^2)} + \frac{\left(\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2}\right)^2}{2\left(\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2} + r(\eta^2 + \vartheta^2)\right)} < I^*.$$

Instead, under unbundling the condition is more stringent:

$$I \leq I_U = \frac{1}{2(1+r\sigma^2)} + \frac{1}{2\mu^2\left(\frac{1}{\mu} + r(\eta^2 + \vartheta^2)\right)} < I_B.$$

That comparison explains the pervasive use of PPP contracts in European countries at the time governments' budget constraints were tight. In the U.K., for example, PPP were highly promoted by the Labour Government on efficiency grounds, though critics also suggested that behind its political support lied also the desire to promote private infrastructure investment at the time where limits imposed by the Maastricht Treaty were tight (IPPR 2001). Following Engel and al. (2007), we know that there is no justification for PPPs from freeing public funds per se. Our analysis shows that PPPs may nevertheless relax budget constraints and facilitate investment just based on efficiency grounds.

4 The Mature Phase of Delegation: Verifiable Productivity Shocks

So far we have modeled uncertain environments where productivity shocks cannot be described ex ante and remain non-contractible. That setting, as we mentioned, was suitable to capture the initial phase of delegation of public service provision to the private sector, where little past experience exists to assess operational risks. Only incomplete contracts that are not contingent on the productivity shocks were feasible. We now consider less uncertain environments where productivity shocks can be foreseen ex ante and thus described in a more complete contract. Ex post, contracting partners will have common knowledge of those shocks. Such setting can be thought of representing a more mature phase of delegation, where some past experience exists to inform parties as to what might or not happen as events unfold during the operational stage.

Information structure. For simplicity, we assume from now on that θ has still zero mean but can only take two values $\bar{\theta} = (1-\nu)\Delta\theta > 0$ and $\underline{\theta} = -\nu\Delta\theta < 0$ with respective probabilities ν and $1-\nu$. We denote $\Theta = \{\underline{\theta}, \bar{\theta}\}$. Let also $E_\theta(\cdot)$ be the expectation operator with respect to θ. For further references, notice that the first-best operational effort in the good productivity state $\bar{\theta}$ is greater than in the bad state $\underline{\theta}$.[28, 29]

Remark 3 *When the productivity shock θ is ex post verifiable, one could a priori think that the principal can provide insurance against this shock rather easily. This will indeed be immediate if shocks directly affect revenues. For instance, revenues could write as $R = e + \theta + \epsilon$ and "uninsurable" revenues would just be $\tilde{R} = R - \theta = e + \epsilon$ so that all relevant incentive contracts would be contingent on this latter variable. However, in our setting, productivity shocks affect the operator's non-verifiable*

[28] This monotonicity property will carry over to the case where θ is private information.

[29] The variance of θ can now be expressed as $\vartheta^2 = \nu(1-\nu)(\Delta\theta)^2$. With this expression at hands, we could "extrapolate" a bit and adapt the results of Section 3 to describe optimal contracts and organizational forms had such binary shock been non-verifiable. Indeed, it is straightforward to check that, with CARA utility functions, the formula for the incentives intensities that we already derived are also valid when $\Delta\theta$ is small enough even if θ is no longer normally distributed.

cost. As such their impact on revenues is only indirect and takes into account the ex post choice of the operational effort. As we will see some insurance is nevertheless still possible.

Unbundling. Consider first the case where the operator and builder are kept as separated entities. Only the operator suffers from the productivity risk and must be somewhat insured against those shocks. At the same time, the builder is still rewarded as a function of the realized quality index and operates under a linear scheme $t_B(Q) = \alpha_B + \beta Q$. This contract is independent of the productivity shock since this shock has not yet realized at the time the builder is paid.

Let us now turn to the description of the operator's contract. The largest class of contracts stipulates a menu of revenues sharing schemes that depend on the realized shock parameter θ since that shock is common knowledge and contractible. Such menu is designed with the dual objective of providing insurance to the operator against productivity shocks but also inducing operational effort. For further references, we denote such mechanism as the collection $\{(\alpha_O(\theta), \gamma(\theta))\}_{\theta\in\Theta}$.

Abusing slightly our previous notations, we define the certainty equivalent of the operator's payoff in state θ as

$$U_O(\theta, a) = \max_{\tilde{e}} \alpha_O(\theta) + \gamma(\theta)\tilde{e} - \frac{\mu}{2}(\tilde{e} - a - \theta)^2 - \frac{r\eta^2\gamma^2(\theta)}{2}.$$

This expression makes it clear that the risk borne *in fine* by the operator depends on the revenues sharing agreement. Unless the principal chooses to always leave all revenues to the firm (i.e., $\gamma(\theta) \equiv 1$ which is very bad for insurance), the fluctuations of the operator's payoff only partially reflect those of the productivity shock.

Using the definition of the optimal operational effort given from (3), we finally obtain

$$U_O(\theta, a) = \alpha_O(\theta) + \gamma(\theta)(a + \theta) + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2(\theta)}{2}. \tag{12}$$

Because θ is still unknown at the time of contracting, the risk-averse operator's participation constraint becomes:

$$E_\theta(v(U_O(\theta, a))) \geq 0. \tag{13}$$

Optimal incentive schemes. We can characterize the optimal contract as follows.

Proposition 4 *Suppose that θ is common knowledge ex post and verifiable. The optimal scheme under unbundling entails the following properties.*

- *The operator is fully insured against productivity shocks; fixed payments being lower in state $\bar{\theta}$:*

$$U_O(\bar{\theta}, a_B^*) = U_O(\underline{\theta}, a_B^*) = 0 \text{ and } \alpha_O^*(\underline{\theta}) > \alpha_O^*(\bar{\theta}).$$

- *The builder operates under the same scheme as when θ is non-verifiable and exerts the same first-stage effort:*

$$\beta_U^* = \beta_U \text{ and } a_U^* = a_U.$$

- *The share of revenues kept by the operator does not depend on the productivity shock but is greater than when productivity shocks are non-verifiable, and operational effort is higher:*

$$\gamma_U^*(\bar{\theta}) = \gamma_U^*(\underline{\theta}) = \gamma_U^* > \gamma_U \text{ and } e(\theta, \gamma_U^*, a_U^*) > e(\theta, \gamma_U, a_U) \quad \forall\theta \in \Theta.$$

Contracts between the public and private sectors for the provision of a service often entail fixed payments to reflect external conditions affecting the operator's profits (see e.g. HM Treasury 2007). When a negative shock realizes during operations (for example, input prices increase, or a national strike slows down production or there is a change in legislation that increases the cost of operations), the firm receives a compensation by the authority. When instead a positive shock realizes (for example, lending conditions improve so that the firm can re-finance its debt reducing its cost of capital), the firm shares the benefit with the authority. This also holds in our context where the fixed payment fully compensates the operator for the lost revenues:

$$0 < \alpha_O^*(\underline{\theta}) - \alpha_O^*(\bar{\theta}) = \gamma_U^* \Delta\theta < \Delta\theta.$$

As the shocks are exogenous and thus outside the firm's control, providing for these "compensation events" does not weaken incentives. In fact, by reducing the noise-to-signal ratio, this insurance reduces the risk premium and thus the cost of incentives.[30] Incentives for operating effort are thus boosted by letting the firm keep more revenues.

An important aspect of the analysis is that the share of revenues kept by the operator does not depend on the productivity shock. In other words, there is a dichotomy between solving moral hazard which is done with a fixed incentives intensity and providing insurance which is obtained by modifying the fixed payment. The reasoning is familiar from agency theory. Making the incentives intensity γ depend on θ would add further noise to the firm's overall compensation without inducing cheaper incentives since the operator's effort does not affect the distribution of θ.[31]

Bundling. Again the conglomerate can be induced to exert first-stage effort by rewarding its performances according to the quality index but now a revenue-sharing agreement can also be contingent on the productivity shock to provide better insurance. To keep symmetry with the unbundling case, we assume that rewards on quality index are made just after the design stage and cannot be contingent on the future productivity shock. A mechanism is thus now a collection $\{(\alpha(\theta), \gamma(\theta))\}_{\theta \in \Theta}$ cum a fixed incentive reward on quality β.[32] Overall, the conglomerate operates under a linear scheme that depends a priori of the ex post realization of θ and writes as:

$$t(Q, R, \theta) = \alpha(\theta) + \beta Q + \gamma(\theta) R.$$

The second-stage incentive constraint is kept unchanged and remains as in (3). To study the firm's first-stage incentive constraint, it is useful to rewrite its overall expected payoff as

$$E_\theta \left(v \left(\beta a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + U_O(\theta, a) \right) \right)$$

For each realization of θ, the conglomerate's profit is the sum of its payoff as a builder and as an operator. In particular, the profit from operating assets $U_O(\theta, a)$ is still defined as in (12). This expression makes it also clear how the conglomerate anticipates the impact of design on second-stage effort and revenues. More precisely, the first-stage effort now solves:

$$a = \arg\max_{\tilde{a}} E_\theta \left(v \left(\beta\tilde{a} - \frac{\tilde{a}^2}{2} - \frac{r\sigma^2\beta^2}{2} + U_O(\theta, \tilde{a}) \right) \right). \tag{14}$$

[30] To see why, let us come back on the case where θ cannot be contracted upon and only a pooling revenues sharing scheme that does not depend on θ can be offered. In this case, and supposing that $\Delta\theta$ is small enough, the extra risk-premium $r\vartheta^2\beta^2/2$ where $\vartheta^2 = \nu(1-\nu)(\Delta\theta)^2$ must be offered to ensure the operator's participation. We are back to the same analysis as in Section 3.

[31] The proof of this result is however somewhat more involved than what such rough intuition suggests. Indeed, in this model, the operational effort is chosen after productivity shocks realize and the principal might want to transfer wealth across those different states to reduce agency costs. The absence of income effect with CARA utility functions nevertheless ensures that such strategy is suboptimal.

[32] Without loss of generality, we incorporate the fixed payment α_B into $\alpha(\hat{\theta})$.

On the other hand, inducing the conglomerate's participation requires now that

$$E_\theta\left(v\left(\beta a-\frac{a^2}{2}-\frac{r\sigma^2\beta^2}{2}+U_O(\theta,a)\right)\right)\geq 0. \tag{15}$$

Proposition 5 *Suppose that θ is common knowledge ex post and verifiable. The optimal scheme under bundling entails the following properties.*

- *The firm is fully insured against productivity shocks. Again, fixed payments are lower in state $\bar{\theta}$*

$$\beta_B^* a_B^* - \frac{(a_B^*)^2}{2} - \frac{r\sigma^2(\beta_B^*)^2}{2} + U_O(\theta, a_B^*) = 0 \quad \forall\theta \text{ and } \alpha_B^*(\bar{\theta}) < \alpha_B^*(\underline{\theta}).$$

- *The share of revenues kept by the firm does not depend on the productivity shock but is greater than when productivity shocks are non-verifiable,*

$$\gamma_B^*(\theta) = \gamma_B^* > \gamma_B \quad \forall\theta\in\Theta.$$

- *Infrastructure quality is less rewarded than when productivity shocks are non-verifiable*

$$\beta_B^* < \beta_B.$$

- *Efforts are greater than when productivity shocks are non-verifiable:*

$$a_B^* > a_B \text{ and } e_B^*(\theta, \gamma_B^*, a_B^*) > e_B(\theta, \gamma_B, a_B^*) \quad \forall\theta\in\Theta.$$

As under unbundling, providing the firm with full insurance against productivity shocks reduces the risk premium and makes it more attractive to rely on revenues as an incentive tool to boost first-stage effort. Now, however, relying on the quality index is also less attractive. The quality index provides incentives for design effort, but under bundling, revenues sharing also provides such incentives - albeit indirectly. When θ is common knowledge and used to adjust the fixed payment, improving the noise-to-signal ratio at the operating stage raises the relative effectiveness of revenues sharing, compared to the quality index, as an instrument to provide incentives for design effort. Overall, operational efforts is boosted directly, as more of the revenues are kept by the firm than when productivity shocks are non-verifiable, and indirectly, as first-stage effort is greater.

These results suggest that projects that should rely more on the revenues from the service (or on service standards in cases where users do not pay) are those where risks that the contractor cannot control can be well identified ex ante (so as to be accounted for in the contract) and ex post (so that compensations can be paid) and then kept within the public sector.

Welfare comparison. The comparison between organizational forms is straightforward. The risk reduction that more complete contracts allows does not change the dominant organizational form. But with risk at operating stage being shared more efficiently (in the sense that the noise-to-signal ratio decreases), revenue sharing becomes a more effective instrument to provide incentives, and the relative gain of bundling increases.[33]

Proposition 6 *Suppose that θ is common knowledge ex post and verifiable. Bundling is again the optimal organizational form. Compared to the case where θ is non-verifiable, the gains of bundling tasks are greater.*

[33] In the simple case where the quality index cannot be used to contract but only revenues sharing schemes are available, the builder has no incentives under unbundling and only bundling can induce some first-stage effort. However, with more complete contracts, leaving more revenues to the conglomerate firm boosts its first-stage effort.

When the firm can use its past experience to predict events that may realize during operations, and when the authority also has such experience, they can agree that the firm will not be responsible for events beyond its control. Risk identification and risk allocation in PPP contracts is indeed key. Standardized PPP contracts typically contain a "risk matrix" that spells out, for each event that can be predicted, how the risk should be allocated between the private and the public sector, where risk sharing considerations in best international practices are in line with our suggestions.[34] In these circumstances, our analysis shows that the benefits of bundling are the greatest. PPPs become particularly effective at incentivizing innovative approaches (a increases) to public service provision and containing the cost of operations (e increases). As it is reasonable to assume that the shocks θ become verifiable ex post as more and more experience in the sector accumulates, our analysis suggests that the benefits from PPPs, whilst positive also under incomplete contracting, will increase as a mature phase of delegation develops.

Finally, it is interesting to observe that even more projects can now be financed under bundling. The condition for a positive net present value becomes:

$$I \leq I_B^* = \frac{1}{2(1+r\sigma^2)} + \frac{\left(\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2}\right)^2}{2\left(\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2} + r\eta^2\right)} \text{ where } I_B^* > I_B.$$

5 The Cost of Innovation: Asymmetric Information on Productivity Shocks

The complexity of the operations undertaken by a contractor may vary greatly from one sector to the other and within the same sector. Complexity is low when the contractor is responsible for simple, low-skilled repetitive tasks that are not critical to the service operation (e.g. routine janitorial services of simple buildings), or for facilities management services considered routine for buildings and basic civil infrastructure. Complexity is higher for performing technically demanding services that require specialist skill sets or demanding project management skills, such as operating complex medical or process engineering equipment. Complex projects are also the ones introducing a new design, changing existing procedures and approaches, experimenting with a new technology or providing new services. The degree of complexity in a project is important since it affects project risk.[35]

With complex projects, the specialized skills required to manage the service and the changes introduced by the innovations typically make past information within the public sector of little use to assess current risks. The greater uncertainty at operational stage that this complexity brings about is then combined with an informational advantage of the operator on the productivity shocks that affect operations. In this section, we analyze the role of PPPs in such scenarios, where complexity, risk, and asymmetric information come together.

With asymmetric information, the contract that was offered to the operator when θ was common knowledge is certainly no longer feasible. Whether bundling or unbundling has been chosen, an operator having faced a high productivity shock wants to report a low productivity in order to receive a higher payment ($\alpha(\bar{\theta}) < \alpha(\underline{\theta})$). There is now a tension between truthtelling and insurance.

Whatever the choice of organizational form, a menu of screening mechanisms stipulates revenues sharing agreements that depend now on the operator's announcement $\hat{\theta}$ on the realization of the shock parameter θ that he has privately observed. By the *Revelation Principle*,[36] there is no loss of generality

[34] Risk matrixes may contain hundreds of different entries, one for each identifiable risk. See the risk allocation guidelines and the risk matrix preparred by the PPP unit in Australia, http://www.partnerships.vic.gov.au/CA25708500035EB6/0/9AFD5F0EA938027FCA2570C0001A8570?OpenDocument

[35] For instance, Moody's (2007) identifies different categories of project complexity and assigns a higher degree of risk to more complex projects.

[36] Myerson (1982).

in considering schemes of the form $\{(\alpha_O(\theta), \gamma(\theta))\}_{\theta\in\Theta}$ that induce truthful revelation. Such menu is now designed with the triple goals of providing the builder with insurance against productivity shocks, inducing efforts at both stages and inducing truthtelling ex post.

Unbundling. Consider first the operator's incentive problem. Let now define the operator's ex post information rent (i.e., once informed on θ) as:

$$U_O(\theta, a) = \max_{(\tilde{e},\hat{\theta})} \alpha_O(\hat{\theta}) + \gamma(\hat{\theta})\tilde{e} - \frac{\mu}{2}(\tilde{e} - a - \theta)^2 - \frac{r\eta^2\gamma^2(\hat{\theta})}{2} = \max_{\hat{\theta}} \alpha_O(\hat{\theta}) + \gamma(\hat{\theta})(a + \theta) + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2(\hat{\theta})}{2}$$

where a is the equilibrium first-stage effort that is anticipated by the operator. This expression encompasses the usual incentive compatibility constraints that are both necessary and sufficient to induce truthful revelation once θ is known. In particular, the constraint that prevents an operator having faced a shock $\bar{\theta}$ to pretend having faced a more averse shock $\underline{\theta}$ can be expressed as:

$$U_O(\bar{\theta}, a) - U_O(\underline{\theta}, a) \geq \Delta\theta\gamma(\underline{\theta}). \tag{16}$$

This constraint is of course the only relevant one (binding) in our framework with two productivity shocks.[37] By under-reporting his productivity, the operator offers a lower estimate of revenues and receives a greater fixed payment. With such strategy, the high-productivity firm appropriates an extra rent worth $\Delta\theta\gamma(\underline{\theta})$. This rent is strictly positive unless $\gamma(\underline{\theta}) = 0$. Setting $\gamma(\underline{\theta}) \equiv 0$ would of course remove the incentives to lie but it would also destroy all incentives to exert effort at the operating stage following a bad shock.

Proposition 7 *Suppose that the operator has private information on θ. The optimal scheme under unbundling entails the following properties.*

- *The firm is only partially insured against productivity shocks:*
$$\hat{U}_O(\bar{\theta}, \hat{a}_U) > 0 > \hat{U}_O(\underline{\theta}, \hat{a}_U).$$
- *The share of revenues kept by the operator is greater and the operating effort is higher when a good productivity shock hits,*
$$\hat{\gamma}_U(\bar{\theta}) = \gamma_U^* > \hat{\gamma}_U(\underline{\theta}) > 0 \text{ and } \hat{e}_U(\bar{\theta}) = e_U^* > \hat{e}_U(\underline{\theta}).$$
- *Incentives intensity on quality and first-stage effort are the same as when productivity shocks are verifiable:*
$$\hat{\beta}_U = \beta_U^* \text{ and } \hat{a}_U = a_U^*.$$

The important insight of Proposition 7 is that solving the asymmetric information problem requires to impose an endogenous risk on the operator (now $\hat{U}_O(\bar{\theta}, \hat{a}_U) > \hat{U}_O(\underline{\theta}, \hat{a}_U)$), which in turn calls for less powerful incentives at the operating stage. On top, the optimal contract exhibits three important features. First, a revenue sharing rule dependent on the productivity shock is necessary to ensure truthtelling. If γ were kept independent of θ (as when shocks are verifiable) the firm would always report a negative shock and retain the unreported revenues.

Second, the truthtelling incentive constraint is relaxed by making the revenues sharing agreement in state $\underline{\theta}$ less powerful whereas the agreement following $\bar{\theta}$ remains unchanged compared to the case

[37] Laffont and Martimort (2002, Chapter 2) for instance.

where productivity shocks are contractible; a standard "no-distortion at the top" result which is familiar from the screening literature.

Third, compared to the case where θ is common knowledge, the firm's exposure to risk is greater and the cost of incentives is higher. To reduce the risk premium, it is then optimal to reduce the operator's share of revenues in state $\underline{\theta}$, although this comes at the cost of reducing the operator's incentives to exert effort. The operator hit by a low productivity shock is instead on very low-powered incentives. In the Appendix, we show that such distortions are very large in the limit when $\Delta\theta$ is itself large and $\gamma_U(\underline{\theta})$ converges towards zero in that case.

Because he is not concerned with productivity risk, the builder is kept under the same scheme than when productivity shocks are verifiable. This property will no longer hold with bundling.

Bundling. A mechanism is still a collection $\{(\alpha(\theta), \gamma(\theta))\}_{\theta\in\Theta}$ cum a fixed incentive reward on quality β. Of course, such scheme has now to be incentive compatible. Slightly abusing our previous notations, we again define the firm's certainty equivalent payoff from operating assets in state θ as as

$$U_O(\theta, a) = \max_{(\tilde{e},\hat{\theta})} \alpha(\hat{\theta}) + \gamma(\hat{\theta})\tilde{e} - \frac{\mu}{2}(\tilde{e} - a - \theta)^2 - \frac{r\eta^2\gamma^2(\hat{\theta})}{2} = \max_{\hat{\theta}} \alpha(\hat{\theta}) + \gamma(\hat{\theta})(a+\theta) + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2(\hat{\theta})}{2}.$$

By a standard argument, such truthtelling constraints imply that the monotonicity condition $\gamma(\bar{\theta}) \geq \gamma(\underline{\theta})$ must hold.

The analysis of the firm's first-stage incentive constraint nevertheless requires some care. Even though the menu $\{(\alpha(\theta), \gamma(\theta))\}_{\theta\in\Theta}$ can be designed to induce truthful revelation given that the firm exerts the equilibrium level of effort a, truthful revelation is no longer guaranteed had the firm deviated to an alternate effort $\tilde{a}$. To see that more clearly, suppose that the principal wants to implement an effort level a that just leaves the firm indifferent between lying or not on the state of nature, namely (16) is binding, with on top a separating allocation such that $\gamma(\bar{\theta}) > \gamma(\underline{\theta})$:

$$U_O(\bar{\theta}, a) = \alpha(\bar{\theta}) + \gamma(\bar{\theta})(a + \bar{\theta}) + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2(\bar{\theta})}{2} = \alpha(\underline{\theta}) + \gamma(\underline{\theta})(a + \bar{\theta}) + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2(\underline{\theta})}{2}$$

$$= U_O(\underline{\theta}, a) + \Delta\theta\gamma(\underline{\theta}).$$

If the agent deviates towards a greater level of effort than that conjectured in equilibrium, $\tilde{a} > a$, truthtelling is preserved. Indeed, we have:

$$U_O(\bar{\theta}, \tilde{a}) = \alpha(\bar{\theta}) + \gamma(\bar{\theta})(\tilde{a} + \bar{\theta}) + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2(\bar{\theta})}{2} = U_O(\bar{\theta}, a) + \gamma(\bar{\theta})(\tilde{a} - a)$$

$$> U_O(\underline{\theta}, a) + \Delta\theta\gamma(\underline{\theta}) + \gamma(\underline{\theta})(\tilde{a} - a) = \alpha(\underline{\theta}) + \gamma(\underline{\theta})(\tilde{a} + \bar{\theta}) + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2(\underline{\theta})}{2}.$$

Instead, if the agent were to deviate towards a lower level of effort $\tilde{a} < a$, that last inequality would be reversed so that, in state $\bar{\theta}$, the firm would prefer to pretend being hit by a bad shock. Intuitively, with asymmetric information, the principal wants the conglomerate to bear more risk to induce truthtelling. However, the firm can reduce its risk exposure by reducing its first-stage effort and always choose thereafter the contract targeted for a low risk. Preventing such "double deviation" is now an issue that was absent under unbundling.

To solve the agency problem in that context, we first describe the form that the truthtelling incentive constraint must take to also control for this double deviation.

Proposition 8 *Suppose that the principal wants to implement an effort a. The following two conditions are necessary to induce truthtelling ex post and no deviation (either downwards or locally upward*[38]*) away from the first-stage effort a at minimal agency cost.*

1. *The truthtelling constraint must write as:*

$$U_O(\bar{\theta}, a) - U_O(\underline{\theta}, a) = \Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma) \tag{17}$$

where $\Delta\gamma = \gamma(\bar{\theta}) - \gamma(\underline{\theta}) \geq 0$ *and* $\xi(\gamma(\bar{\theta}), \Delta\gamma)$ *is defined as the (unique) non-negative solution to*

$$\xi(\gamma(\underline{\theta}), \Delta\gamma) = \frac{2\nu(\Delta\gamma)^2}{\nu + (1-\nu)exp(r(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma)))}. \tag{18}$$

2. *The first-stage effort must satisfy:*

$$a = \beta + \gamma(\underline{\theta}) + \frac{\nu\Delta\gamma}{\nu + (1-\nu)exp(r(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma)))}. \tag{19}$$

The new *generalized incentive constraint* (17) encompasses both the possibility of ex ante moral hazard and ex post adverse selection deviations in a rather tractable way. To better provide insurance, the principal needs to screen the firm according to its productivity shock which means implementing revenues sharing schemes that satisfy $\Delta\gamma > 0$. But preventing then the "double deviation" requires the firm to bear so much risk ex post that the truthtelling constraint is actually slack. The term $\xi(\gamma(\bar{\theta}), \Delta\gamma)$ represents how much extra slack is needed; it captures the new agency cost that arises under asymmetric information. There is no explicit formula for that extra slack but it is immediate to observe that it increases with $\Delta\gamma > 0$ and decreases with $\gamma(\underline{\theta})$. As a result of the extra risk borne at the operational stage, providing incentives on design is more costly than when θ is verifiable as captured by the incentive constraint (19) which generalizes (6) to this more complex environment.[39]

Our analysis demonstrates that the compounding of first-stage moral hazard and ex post asymmetric information makes it costly to offer different revenues sharing schemes for different realizations of θ. Although such screening is attractive to solve the asymmetric information side of the problem and provide insurance, it weakens incentives for efforts. Our next result unveils a condition under which this trades-off makes pooling optimal under bundling, i.e., $\Delta\gamma = 0$.

Proposition 9 *Suppose that the firm has private information on θ and that $\Delta\theta$ is small enough. Then, the optimal pooling mechanism (i.e., such that the revenues sharing agreement is independent of the productivity shocks) is such that:*

$$\hat{\gamma}_B(\bar{\theta}) = \hat{\gamma}_B(\underline{\theta}) = \hat{\gamma}_B \in (\hat{\gamma}_U, \gamma_B^*).$$

This pooling mechanism implements an effort $\hat{a}_B$ such that

$$\hat{a}_B < a_B^*.$$

[38]The conditions (17) and (19) below are only necessary and not sufficient because they are obtained by considering either downward deviations or local upward deviations away from the level of effort a that one wants to implement. In particular, upward deviations are never so large that it could become attractive to have countervailing incentives, with the firm claiming to be in state $\bar{\theta}$ when instead $\underline{\theta}$ realized. In other words, some of the first-stage effort levels and rent profiles satisfying (17) and (19) may no longer be feasible when such large deviations are considered. The argument developed below is that the optimal effort within the (possibly) larger set of implementable allocations that we so describe might indeed be achieved with a contract -and, anticipating on the findings in Proposition 9, by a pooling contract- that is actually immune to such non-local upward deviations.

[39]Observe that, when the firm becomes more risk averse (r increases), the revenues sharing agreement in the worst state $\underline{\theta}$ matters more when choosing first-period effort. Incentives to under-supply that first-period effort are exacerbated.

This mechanism cannot be improved upon by slightly increasing $\hat{\gamma}_B(\bar{\theta})$ above $\hat{\gamma}_B(\underline{\theta})$.[40]

When uncertainty on productivity shocks is small, screening the firm with different state-contingent contracts does not help much but it would require to keep a costly slack on the truthtelling constraint. The principal prefers to pool over revenues sharing schemes to limit the first-stage effort distortion even though this destroys insurance against productivity shocks.

Renegotiation. Suppose instead that $\Delta\theta$ is large enough so that, the optimal mechanism under bundling might have $\Delta\gamma > 0$. Such mechanism is not immune to the possibility of an interim renegotiation which would take place just after the design stage but still before the firm learns the productivity shock and chooses accordingly its operational effort. Since the first-stage effort is now sunk under such scenario, a renegotiation towards a contract with incentives intensity $\gamma^R(\theta)$, fees $\alpha^R(\theta)$ and payoffs $U_O^R(\theta, a)$ would reduce the risk borne at the operating stage to its minimal amount consistent with truthful revelation of the productivity shock θ. There is no longer any reason to leave a slack on the truthtelling constraint which now becomes:

$$U_O^R(\bar{\theta}, a) - U_O^R(\underline{\theta}, a) = \Delta\theta\gamma^R(\underline{\theta}).$$

Anticipating such risk-reducing renegotiation,[41] the firm finds it attractive to reduce its first-stage effort, as we have seen above.

Insisting on (deterministic) contracts that remain renegotiation-proof thus forces the principal to offer pooling allocations which are immune to such risk reduction:

$$\gamma^R(\bar{\theta}) = \gamma^R(\underline{\theta}) \text{ and } \alpha^R(\bar{\theta}) = \alpha^R(\underline{\theta}).$$

Mutatis mutandis, Proposition 9 applies again for those renegotiation-proof contracts for any $\Delta\theta$. We can state:

Proposition 10 *Suppose that the firm has private information on θ and that renegotiation-proof (deterministic) contracts are offered. Then, the optimal renegotiation-proof mechanism under bundling is such that the revenues sharing agreement is independent of the productivity shocks:*

$$\hat{\gamma}_B(\bar{\theta}) = \hat{\gamma}_B(\underline{\theta}) = \hat{\gamma}_B \in (\hat{\gamma}_U, \gamma_B^*).$$

This pooling mechanism implements an effort $\hat{a}_B$ such that

$$\hat{a}_B < a_B^*.$$

Welfare comparison. As we have seen above, screening possibilities disappear under bundling when either $\Delta\theta$ is small enough or renegotiation is a concern. From a welfare point of view, the trade-off between organizational forms is then reduced to comparing the gains from bundling tasks, which better incentivizes the first-stage effort, and the costs of keeping pooling allocations at the operating stage, which weakens incentives for operating effort. The next proposition unveils that the trade-off can indeed go now against bundling.

Proposition 11 *The following properties hold.*

[40] Due to the highly nonlinear objectives function, we content ourselves with checking the local optimality of a pooling contract. Actually, from Proposition 11 below, this local optimal pooling contract suffices to prove the benefits of bundling.

[41] Note that unbundling is of course and by definition immune to the threat of such renegotiation.

- *When $\Delta\theta$ is small enough, bundling dominates.*
- *When $\Delta\theta$ is large enough and renegotiation is a concern under bundling, unbundling dominates.*

Under bundling, the compounding of asymmetric information ex post, moral hazard and renegotiation forces to use pooling contracts that do not screen according to productivity shocks. For large operational risks, such pooling contracts are close to fixed payments which induce no operational effort. Under unbundling, where it is easier to extract the information necessary to provide insurance, the revenues sharing scheme is higher powered following a good shock. Unbundling thus still allows screening along productivity shocks and preserves some incentives at the operational stage when a high productivity shock hits. Bundling is then suboptimal.

Instead, for small operational risks, asymmetric information is not a big deal. The possibility of screening the operator according to productivity shocks under unbundling does not yield much gains compared to the internalization of externalities that is possible under bundling, even though a pooling contract is offered in that case.

Our results suggest that bundling of project phases into a single contract is not always positive, despite the presence of a positive externality across stages. For complex projects where risk is high, PPPs may refrain innovation and result in higher operational costs compared to unbundling. Innovation in complex projects brings about a level of risk at operational stage that neither the authority nor the contractor wishes to bear. Asymmetric information on the events that unfold at operational stage makes it difficult to compensate the firms for events beyond its control. Risk is then shared inefficiently and incentives at operating stage are weakened. The fear that innovative approaches at design stages create excessive risk bearing at operating stage will also refrain the firm from innovating.

6 The Value and Cost of Private Finance

In DBFO models, the bundling of design and implementation is accompanied by the use of private finance, whilst in BOT model it is not: the facility is financed by the public sector and remains in public ownership throughout the contract. In practice, the interest of lenders in financing infrastructure projects for the provision of public services is explained by the stable returns, to a large extent uncorrelated with the market, that these investments are known to generate. But the interest of governments in private finance is less clear-cut. Existing evidence suggests that private finance has been sometimes employed more because of the urge to overcome budget and Maastricht Treaty constraints, or to gain political consent through strategic infrastructure programs, than because of efficiency considerations (IPPR, 2001).[42] And since private finance is typically more costly than public finance, the cost of this distortion has often been high. However, an argument also often heard is that, with significant funds at risk, lenders and equity investors provide a level of rigor and due diligence to assessing project risks that may turn beneficial also to the public sector, especially when the government lacks the knowledge and expertise to evaluate project risks rigorously (Leahy 2005).

In this section, we analyze the potential benefits and costs of private finance when there is bundling of design and implementation, considering the case where the financier observes the ex post realization of the productivity shock whilst the government does not.[43] This assumption implies that the government's contract cannot be made contingent on the shock, whilst the financier can ask for reimbursements that explicitly depend on realized profits and thus on productivity shocks.

[42] The Eurostat made a decision (news release 18/2004) on the accounting of PPPs to ensure homogeneity across member states and limit accounting tricks made to comply with the rules of the Stability and Growth Pact.

[43] De Bettignies and Ross (2009) also discuss the benefit of private finance in PPPs. In their model, information is symmetric but private finance may lead to the efficient termination of bad projects, while public developers may sustain such projects for political reasons.

Suppose that the government and the financier are two different entities taking moves sequentially. The firm must respond to the contracts offered by those two principals. The government offers first a contract to the firm, which stipulates how much the firm should be paid for providing the services. Then, the firm turns to its financier to cover the initial investment I that is needed to build the infrastructure.

For simplicity, we shall also assume that $\sigma^2 = +\infty$, so that no quality index is available (equivalently, $\beta \equiv 0$). Note that in this case, we have $\gamma_B^* = \frac{1+\frac{1}{\mu}}{1+\frac{1}{\mu}+r\eta^2}$. Bundling has then its greatest benefits. In such context, the government can still use a profit sharing scheme of the form $t^*(R) = \alpha_B + \gamma_B^* R$ and keep the residual $R - t^*(R)$. The fee α_B will be soon described but we should already notice that this contract is seemingly incomplete because it does not provide any insurance to the firm against productivity shocks. Note also that such scheme preserves the choice of operational efforts that remain equal to their optimal values under all circumstances below.

Given such a scheme, the financier asks for state-dependent loan reimbursements of the form

$$z(\theta) = \delta_B + \gamma_B^* \theta$$

so that the firm's net return in state θ, when it exerts effort a, writes now as

$$\alpha_B - \delta_B + \gamma_B^* a - \frac{a^2}{2} + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma_B^{*2}}{2}.$$

This quantity is independent of θ so that the firm is thus insured against productivity shocks. Clearly such reimbursement rule is optimal from the financier's viewpoint. Notice also that, given such payoff, the firm also chooses the second-best first-stage effort $a_B^* = \gamma_B^*$.

Given that the financier has all bargaining power in setting up reimbursement rules, he extracts as much as possible from the firm and gets an expected payoff worth

$$\delta_B - I = \alpha_B + \gamma_B^* a_B^* - \frac{a_B^{*2}}{2} + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma_B^{*2}}{2} - I.$$

Moving backwards, the government initially chooses α_B so that, in expectations, the financier breaks even and, finally,

$$\alpha_B = I - \left(1 + \frac{1}{\mu} - r\eta^2\right)\frac{\gamma_B^{*2}}{2}.$$

This implementation is not unique. The government could as well ask the firm for an entry fee α_B' and let the financier not only provide insurance but also solve the moral hazard problem by himself. Given that this entry fee only shifts the firm's payoff by a constant, it has no incentive effect. The financier's problem is thus formally identical to that of the government in our main scenario were both the responsibilities for incentivizing the firm and proving insurance were in its own hands. Therefore, the financier chooses the same incentives intensity γ_B^* and his overall payoff ends up being:

$$a_B^* - \frac{a_B^{*2}}{2} + \frac{1}{\mu}\left(\gamma_B^* - \frac{\gamma_B^{*2}}{2}\right) - \frac{r\eta^2\gamma_B^{*2}}{2} - I - \alpha_B'$$

where again $a_B^* = \gamma_B^*$. Again, the government can take advantage of his first-mover advantage in setting up the fee α_B' to just let the financier break-even. This yields:

$$\alpha_B' = a_B^* - \frac{a_B^{*2}}{2} + \frac{1}{\mu}\left(\gamma_B^* - \frac{\gamma_B^{*2}}{2}\right) - \frac{r\eta^2\gamma_B^{*2}}{2} - I \equiv \frac{\left(1+\frac{1}{\mu}\right)^2}{2\left(1+\frac{1}{\mu}+r\eta^2\right)} - I.$$

This entry fee is of course just equal to the social value of the project that could have been achieved with the government directly providing insurance on top of inducing efforts. In other words, the financier is made indirectly residual claimant for the implementation of the optimal incentive package.

Proposition 12 *Suppose that the financier has the expertise to observe θ whilst the government does not. Then, bundling dominates unbundling.*

Using private finance brings additional costs over and above conventional funding. The cost of the debt, which in the UK before the current credit crisis was about 1 percentage point (60-150 basis points) above the nominal cost of Government borrowing, by 2009 was estimated to have risen to around 140 to 250 basis points (House of Lords, 2009) because of the credit crunch. The current cost of private finance raises issues about the benefit of PPP. If the private sector borrows capital at a higher interest rate than public authorities, the real cost of an investment I becomes $(1 + \rho)I$ for some $\rho > 0$. Notwithstanding, Proposition 12 suggests that the benefits may still outweigh this cost if project risk is high ($\Delta\theta$) and the financier has the expertise to evaluate project risks to an extent that public authorities cannot do. With its expertise, the financier can help to improve upon risk allocation and re-establish the benefit of bundling.

7 Conclusion

We have studied the agency costs of delegated project management, focusing on the link between the organizational form and the uncertainty that characterizes the project implementation. Our analysis has pointed at the efficiency gains that bundling of project planning and implementation can bring to a public sector seeking to delegate the provision of public services, but it has also emphasized how unbundling may be preferred when operational risks are high and informational asymmetries can create an undue advantage to the firm. In public procurement, our results suggests that contracting out through PPPs can yield potentially the highest benefit for services where uncertainty is limited or where sufficient past experience exists to inform the parties as to what may happen during operations so that an efficient risk allocation can be achieved. This also points out to the suitability of PPPs for services that have been traditionally provided in house under low-powered incentives but where the institutional context has changed creating scope for modernizing approaches, risks however remaining low. Examples include prison services and educational services.

Caution should instead be exerted when the public sector seeks to radically innovate on public service provision or to introduce new services but it lacks the knowledge or expertise to anticipate the impact of the innovative design/procedure/technology on the cost of operations. PPPs are less likely to deliver efficiency gains for highly innovative and complex services where risks are high and it is difficult for the government to commit to transfer such high risks to the private sector. However, to the extent that it is easier to predict contingencies that may arise during operations once the service has been already contracted out to the private sector, and thus that asymmetric information reduces over time, our analysis suggests that improvements in PPP contracting and performance should be observed over time.

An important issue that has been left out of the analysis is related to the procurement process for PPPs and how the uncertainty that characterizes service provision affects the costs of participating to the tender and thus the level of competitiveness in the contract award. Most PPP or PFI contracts are too complex to use the open or restricted procedure. In most PPPs, the contracting authority is unable to determine the technical specifications and the appropriate price level in advance. Therefore, until now the negotiated procedure has been the preferred solution for procuring PPP or PFI contracts. Current experience shows however that the procurement process for PPPs has been costly and time consuming. Albeit with differences between sectors, it has been estimated that PPP tendering periods last an average of 34 months (NAO, 2007) and that procurement costs can reach 5-10%

of the capital cost of a project (Yescombe, 2007).[44] Recently the European Commission introduced the Competitive Dialogues, a new procedure for PPPs contracts (EU Directive 2004/18/EC) but doubts have emerged as to its suitability because of it being a complex and time consuming procedure.

References

Aghion, P. and J. Tirole (1994), "The Management of Innovation," *Quarterly Journal of Economics,* 109: 1185-1209.

Auriol, E. and P.M. Picard (2011), "A Theory of BOT Concession Contracts," , *IDEI Working Paper, n. 667.*

Baker, G. and B. Jorgensen (2003), "Volatility, Noise and Incentives," Mimeo Harvard University.

Bajari, P. and S. Tadelis (2001), "Incentives versus Transaction Costs: A Theory of Procurement Contracts," *The RAND Journal of Economics*, 32: 387-407.

Baliga, S. and T. Sjöström (1998), "Decentralization and Collusion," *Journal of Economic Theory*, 104: 525-536.

Baron, D. and D. Besanko (1992), "Information, Control and Organizational Structure," *Journal of Economics and Management Strategy*, 1: 237-275.

Baron, D. and D. Besanko (1999), "Informational Alliance," *Review of Economic Studies*, 66: 743-768.

Bennett, J. and E. Iossa (2006), "Building and Managing Facilities for Public Services," *Journal of Public Economics*, 90: 2143-2160.

Bennett, J. and E. Iossa (2010), "Contracting Out Public Service Provision to Not-for-profit Firms," *Oxford Economic Papers,* 62: 784-802.

Benz, A., P. Grout, and M. Halonen (2001), "What Should the State Buy?" CMPO Working Paper No. 01/40, University of Bristol.

Besley, T. and M. Ghatak (2001), "Government versus Private Oownership of Public Goods," *Quarterly Journal of Economics*, 116: 1343-1372.

Che, Y. K. and S. Yoo (2001), "Optimal Incentives for Teams," *American Economic Review*, 91: 525-541.

Choi, Y. (1993), "Managerial Incentive Contracts with a Production Externality," *Economics Letters,* 42: 37-42.

De Bettignies, J.E.and T.W. Ross (2009), "Public–Private Partnerships and the Privatization of Financing: An Incomplete Contracts Approach," *International Journal of Industrial Organization*, 27: 358-368.

Dequiedt, V. and D. Martimort (2004), "Delegated Monitoring versus Arm's-Length Contracting," *International Journal of Industrial Organization*, 22: 951-981.

Engel, E., R. Fisher and A. Galetovic (2007), "The Basic Public Finance of Public-Private-Partnerships," forthcoming in *The Journal of the European Economic Association.*

[44]These transaction costs are also to a large extent independent of the size of a project, which suffices to make PPP unsuitable for low capital value projects. The HM Treasury (2006) in the UK currently considers PFI projects for less than £20m as poor value for money.

Estache, A. and B. Wren-Lewis (2009), "Towards a Theory of Regulation for Developing Countries: Following Jean-Jacques Laffont's Lead," *Journal of Economic Literature,* 47: 730-771.

Francesconi, M. and A. Muthoo (2006), "Control Rights in Complex Partnerships," IZA Discussion Paper.

Gilbert, R. and M. Riordan (1995), "Regulating Complementary Products: A Comparative Institutional Analysis," *The RAND Journal of Economics*, 26: 243-256.

Gromb, D. and D. Martimort (2007), "Collusion and the Organization of Delegated Expertise," *Journal of Economic Theory*, 137: 271-299.

Guash, J.L. (2004), "Granting and Renegotiating Infrastructure Concessions: Doing it Right," *WBI Development studies, the World Bank.*

Guasch, J.L., J.J. Laffont and S. Straub (2008), "Renegotiation of Concession Contracts in Latin America: Evidence from the Water and Transport Sectors," *International Journal of Industrial Organization*, Special Issue on PPPs , 26: 421-442.

Hart, O. (2003), "Incomplete Contracts and Public Ownership: Remarks and an Application to Public-Private Partnerships," *Economic Journal*, 119: 69-76.

HM Treasury (2006), *Strengthening Long-Term Partnerships.* London: HM Treasury.

HM Treasury (2007), *Standardisation of PFI Contracts, Version 4.* London: HM Treasury.

Holmström, B. (1979), "Moral Hazard and Observability," *Bell Journal of Economics*, 10: 74-91.

Holmström, B. and P. Milgrom (1987), "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica* 55: 303-328.

Holmström, B. and P. Milgrom (1991), "Multi-Task Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization*, Special Issue, 24-52.

IPPR, Institute of Public Policy Research (2001), *Building Better Partnerships.* London: Institute for Public Policy Research.

Khalil, F., D. Kim and D. Shin (2006), "Optimal Task Design: To Integrate or Separate Planning and Implementation?" *Journal of Economics and Management Strategy*, 15: 457-47.

King, S. and R. Pitchford (2001), "Private or Public? A Taxonomy of Optimal Ownership and Management Regimes," Mimeo University of Melbourne.

Krähmer, D. and R. Strausz (2009), "Optimal Procurement Contracts with Pre-Project Planning,"Mimeo University of Berlin.

Laffont, J.J. (2005), *Regulation and Development*, Cambridge University Press.

Laffont, J.J. and D. Martimort (1998), "Collusion and Delegation," *The RAND Journal of Economics*, 29: 280-305.

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

Laux, C. (2001), "Limited Liability and Incentive Contracting with Multiple Projects," *The RAND Journal of Economics*, 32: 514-526.

Leahy, P. (2005), "Lessons from the Private Finance Initiative in the United Kingdom," European Investment Bank Papers.

Lewis, T. and D. Sappington (1995), "Optimal Capital Structure in Agency Relationships," *The RAND Journal of Economics*, 26: 343-361.

Lewis, T. and D. Sappington (1997), "Information Management in Incentive Problems," *Journal of Political Economy*, 105: 786-819.

Macho-Stadler, I. and D. Perez-Castrillo (1993), "Moral Hazard with Several Agents: The Gains from Cooperation," *International Journal of Industrial Organization*, 11: 73-100.

Martimort, D. (2006), "An Agency Perspective on the Costs and Benefits of Privatization," *Journal of Regulatory Economics*, 30: 5-44.

Martimort, D. and J. Pouyet (2008), "Build It Not: Normative and Positive Theories of Public-Private Partnerships," *International Journal of Industrial Organization*, Special Issue on PPPs , 26: 393-411.

Martimort, D. and W. Sand-Zantman (2007), "Signalling and the Design of Delegated Management Contracts for Public Utilities," *The RAND Journal of Economics*, 37: 763-782.

Myerson, R., (1982), "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10: 67-81.

MOODY'S (2007), "Operating Risk in Privately-Financed Public Infrastructure (PFI/PPP/P3) Projects."

Mookherjee, D. and M. Tsumagari (2004), "The Organization of Supplier Networks: Effects of Delegation and Intermediation," *Econometrica*, 72: 1179-1219.

NAO, National Audit Office (2003), *Operational Performance of Prisons.* London: National Audit Office, HC700.

NAO, National Audit Office (2004), *London Underground: Were They Good Deals?* London: National Audit Office, HC 645.

NAO, National Audit Office (2007), *Improving the PFI Tendering Process.* London: National Audit Office, HC149.

PUK, Partnership UK (2008), *Investigating the Performance of Operational PFI contracts,* a report by Ipsos MORI. London.

Ramakrishnan, R. and A. Thakor (1991), "Cooperation versus Competition in Agency," *Journal of Law, Economics and Organization,* 7: 248-283.

Rosenkranz, S. and P. Schmitz (2003), "Optimal Allocation of Ownership Rights in Dynamic RD Alliances," *Games and Economic Behavior*, 43: 153-173.

Sappington, D. and J. Stiglitz (1987), "Privatization, Information and Incentives," *Journal of Policy Analysis and Management*, 6: 567-582.

Schmitz, P. (2005), "Allocating Control in Agency Problems with Limited Liability and Sequential Hidden Actions," *The RAND Journal of Economics,* 36: 318-336.

SERCO Institute (2007), "What Gets Measured: Contracting for Delivery," The Serco Institute.

Severinov, S. (2008), "The Value of Information and Optimal Organization," *The RAND Journal of Economics*, 39: 238-265.

Shapiro, C. and R. Willig (1990), "Economic Rationales for the Scope of Privatization," in E. Suleiman and J. Waterbury eds. *The Political Economy of Public Sector Reform and Privatization*, 55-87.

Shleifer, A. (1998), "State versus Private Ownership," *Journal of Economic Perspectives*, 12: 133-150.

Tirole, J. (1999), "Incomplete Contracts: Where Do We Stand," *Econometrica*, 67: 741-781.

Appendix

Proof of Proposition 1. Under unbundling, the principal's expected payoff is

$$E_\theta \left((1-\gamma)e(\theta,\gamma,a)\right) - \beta a - \alpha_B - \alpha_O - I.$$

Taking into account the expressions of of the operator's and builder's payoffs, the principal's problem consists in maximizing expected net surplus minus risk premiums and any surplus left to the builder and the operator:

$$\max_{(U_B, U_O(\cdot), \beta, \gamma)} E_\theta \left(e(\theta,\gamma,a) - \frac{\mu}{2}(e(\theta,\gamma,a) - a - \theta)^2 \right) - \frac{a^2}{2} - \frac{r(\eta^2+\vartheta^2)\gamma^2}{2} - \frac{r\sigma^2\beta^2}{2} - U_B - U_O(a) - I$$

subject to (1), (2), (3) and (4).

The two participation constraints (2) and (4) are of course binding. Inserting $U_B = U_O(a) = 0$ and the expressions of efforts from (1) and (3) into the maximand, we must solve:

$$\max_{(\beta,\gamma)} \mathcal{W}_U(\beta,\gamma) \equiv \beta - \frac{\beta^2}{2} + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r(\eta^2+\vartheta^2)\gamma^2}{2} - \frac{r\sigma^2\beta^2}{2} - I.$$

Optimizing immediately yields (5). ■

Proof of Proposition 2. Under bundling, the principal's expected payoff is

$$E_\theta \left((1-\gamma)e(\theta,\gamma,a)\right) - \beta a - \alpha - I.$$

Taking into account the expressions of the conglomerate's payoff, we obtain the following expression of the principal's problem:

$$\max_{(U(a), \beta, \gamma)} \quad E_\theta \left(e(\theta,\gamma,a) - \frac{\mu}{2}(e(\theta,\gamma,a) - a - \theta)^2 \right) - \frac{a^2}{2} - \frac{r(\eta^2+\vartheta^2)\gamma^2}{2} - \frac{r\sigma^2\beta^2}{2} - U(a) - I$$

subject to (3), (6) and (7).

Inserting the expressions of efforts at both stages given by (3) and (6) into the maximand and taking into account that (7) is binding, this optimization problem boils down to:

$$\max_{(\beta,\gamma)} \mathcal{W}_B(\beta,\gamma) \equiv \beta + \gamma - \frac{(\beta+\gamma)^2}{2} + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r(\eta^2+\vartheta^2)\gamma^2}{2} - \frac{r\sigma^2\beta^2}{2} - I.$$

First-order conditions for optimality are:

$$1 - a = 1 - \beta - \gamma = r\sigma^2\beta,$$

$$1 - a + \frac{1}{\mu}(1-\gamma) = r(\eta^2+\vartheta^2)\gamma.$$

From this, the optimal incentives intensities and first-stage effort given in (8), (9) and (10) follow. ■

Proof of Proposition 3. First, observe that

$$\max_{\beta} \mathcal{W}_U(\beta,\gamma) = \frac{1}{2(1+r\sigma^2)} + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r(\eta^2+\vartheta^2)\gamma^2}{2} - I$$

whereas

$$\max_{\beta} \mathcal{W}_B(\beta,\gamma) = \gamma - \frac{\gamma^2}{2} + \frac{(1-\gamma)^2}{2(1+r\sigma^2)} + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r(\eta^2+\vartheta^2)\gamma^2}{2} - I.$$

Since $\gamma \in [0,1]$, we thus get:

$$\Delta\mathcal{W}(\gamma) = \max_{\beta} \mathcal{W}_B(\beta,\gamma) - \max_{\beta} \mathcal{W}_U(\beta,\gamma) = \frac{r\sigma^2}{1+r\sigma^2}\left(\gamma - \frac{\gamma^2}{2}\right) \geq 0. \tag{A1}$$

From this inequality, we immediately derive

$$\max_{(\beta,\gamma)} \mathcal{W}_B(\beta,\gamma) \geq \max_{\beta} \mathcal{W}_B(\beta,\gamma_U) > \max_{\beta} \mathcal{W}_U(\beta,\gamma_U) = \max_{(\beta,\gamma)} \mathcal{W}_U(\beta,\gamma)$$

where the strict inequality follows from the fact that $\gamma_U \in (0,1)$. ■

Proof of Proposition 4. Under unbundling, the principal's expected payoff writes now as:

$$E_\theta\left((1-\gamma(\theta))e(\theta,\gamma(\theta),a) - \alpha_O(\theta)\right) - \beta a - \alpha_B - I.$$

Using the expressions of U_B and $U_O(\theta,a)$ given in the text, the optimal incentive package solves:

$$\max_{(a,U_B,U_O(\cdot),\beta,\gamma(\cdot))} E_\theta\left(e(\theta,\gamma(\theta),a) - \frac{\mu}{2}(e(\theta,\gamma(\theta),a)-a-\theta)^2 - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} - \frac{r\eta^2\gamma^2(\theta)}{2} - U_O(\theta,a)\right) - U_B - I$$

subject to (1), (2), (3) and (13).

Both participation constraints (2) and (13) are binding at the optimum. (2) binding means $U_B = 0$. From (13) binding and the strict concavity of $v(\cdot)$, it immediately follows that $U_O(\bar{\theta},a) = U_O(\underline{\theta},a) = 0$. Inserting those expressions of the operator's and builder's payoffs into the maximand, we obtain:

$$\max_{(\beta,\gamma(\cdot))} \beta - (1+r\sigma^2)\frac{\beta^2}{2} + E_\theta\left(\frac{1}{\mu}\left(\gamma(\theta) - \frac{\gamma^2(\theta)}{2}\right) - \frac{r\eta^2\gamma^2(\theta)}{2}\right) - I.$$

Optimizing with respect to $\gamma(\theta)$ and β finally yields

$$\gamma_U^*(\bar{\theta}) = \gamma_U^*(\underline{\theta}) = \gamma_U^* \equiv \frac{\frac{1}{\mu}}{\frac{1}{\mu} + r\eta^2} > \gamma_U,$$

$$\beta_B^* = \beta_U \text{ and } a_B^* = a_U.$$

Because the operator is fully insured against productivity shocks, we immediately obtain

$$\alpha_O^*(\underline{\theta}) - \alpha_O^*(\bar{\theta}) = \gamma_U^*\Delta\theta > 0.$$

For future reference (see below the proof of Proposition 6), we define also the principal's expected payoff when choosing a mechanism with incentives intensities β and $\gamma(\underline{\theta}) = \gamma(\bar{\theta}) = \gamma$ as

$$\mathcal{W}_U^*(\beta,\gamma) = \beta - (1+r\sigma^2)\frac{\beta^2}{2} + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r\eta^2\gamma^2}{2} - I.$$

■

Proof of Proposition 5. We first prove the following lemma.

Lemma 1 *Under bundling, there is no gain to make γ depend on θ:*

$$\gamma(\bar{\theta}) = \gamma(\underline{\theta}) = \gamma. \tag{A2}$$

Proof. First, observe that the operator's payoff with a state-dependent contract writes a priori as

$$U_O(\theta, a) = \alpha(\theta) + \gamma(\theta)(a+\theta) + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2(\theta)}{2}. \tag{A3}$$

Using the CARA specification, the conglomerate's participation constraint (15) can be rewritten as

$$U_O(\underline{\theta}, a) - \frac{1}{r} ln\left(1 - \nu + \nu exp(-r(U_O(\bar{\theta}, a) - U_O(\underline{\theta}, a)))\right) \geq 0$$

Denoting $U_O(\theta, a) = \gamma(\theta)a + \tilde{U}_O(\theta)$, $\Delta\gamma = \gamma(\bar{\theta}) - \gamma(\underline{\theta})$ and $\Delta\tilde{U}_O = \tilde{U}_O(\bar{\theta}) - \tilde{U}_O(\underline{\theta})$, this condition becomes:

$$(\beta + \gamma(\underline{\theta}))a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + \tilde{U}_O(\underline{\theta}) - \frac{1}{r} ln\left(1 - \nu + \nu exp(-r(a\Delta\gamma + \Delta\tilde{U}_O))\right) \geq 0. \tag{A4}$$

Observe that the right-hand side above is strictly concave in a. The first-order approach leads us thus to the following expression of the conglomerate's incentive constraint in the first stage:

$$a = \beta + \gamma(\underline{\theta}) + \frac{\nu\Delta\gamma exp(-r(a\Delta\gamma + \Delta\tilde{U}_O))}{1 - \nu + \nu exp(-r(a\Delta\gamma + \Delta\tilde{U}_O))}. \tag{A5}$$

The principal's expected payoff writes now:

$$E_\theta\left((1 - \gamma(\theta))e(\theta, \gamma(\theta), a) - \alpha(\theta)\right) - \beta a - I.$$

Using (3) and (A3), this expression can be rewritten as:

$$(1-\beta)a + E_\theta\left(\frac{1}{\mu}\left(\gamma(\theta) - \frac{\gamma^2(\theta)}{2}\right) - \frac{r\eta^2\gamma^2(\theta)}{2} - U_O(\theta, a)\right) - I \tag{A6}$$

Recalling the definition of $U_O(\theta, a)$, we can rewrite the optimization problem as:

$$\max_{(a,\beta,\gamma(\cdot),\tilde{U}(\cdot))} (1-\beta)a + E_\theta\left(\frac{1}{\mu}\left(\gamma(\theta) - \frac{\gamma^2(\theta)}{2}\right) - \frac{r\eta^2\gamma^2(\theta)}{2}\right) - \nu(a\Delta\gamma + \Delta\tilde{U}_O) - \gamma(\underline{\theta})a - \tilde{U}_O(\underline{\theta}) - I$$

subject to (A4) and (A5).

Clearly, the participation constraint (A4) is binding. Inserting the expression of $\tilde{U}_O(\underline{\theta})$ so obtained into the maximand, we can rewrite the optimization problem as

$$\max_{(y,a,\beta,\gamma(\cdot))} a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + E_\theta\left(\frac{1}{\mu}\left(\gamma(\theta) - \frac{\gamma^2(\theta)}{2}\right) - \frac{r\eta^2\gamma^2(\theta)}{2}\right) + \varphi(y) - I$$

$$\text{subject to } a = \beta + \gamma(\underline{\theta}) + \frac{\nu\Delta\gamma exp(-ry)}{1 - \nu + \nu exp(-ry)} \tag{A7}$$

where we have introduced the new optimization variable

$$y = a\Delta\gamma + \Delta\tilde{U}_O$$

and defined the function

$$\varphi(y) = -\nu y - \frac{1}{r} ln\left(1 - \nu + \nu exp(-ry)\right).$$

Note that $\varphi(0) = 0$, $\varphi'(y) = -\frac{\nu(1-\nu)(1-exp(-ry))}{1-\nu+\nu exp(-ry)}$, with in particular $\varphi'(0) = 0$.

Denoting by λ the multiplier of the incentive constraint (A7),the optimality conditions with respect to a, β, $\gamma(\bar{\theta})$, $\gamma(\underline{\theta})$ and y are respectively given by:

$$a = 1 - \lambda, \tag{A8}$$

$$-r\sigma^2\beta + \lambda = 0, \tag{A9}$$

$$\frac{1}{\mu}(1 - \gamma(\bar{\theta})) - r\eta^2\gamma(\bar{\theta}) + \frac{\lambda exp(-ry)}{1 - \nu + \nu exp(-ry)} = 0, \tag{A10}$$

$$\frac{1}{\mu}(1 - \gamma(\underline{\theta})) - r\eta^2\gamma(\underline{\theta}) + \frac{\lambda}{1 - \nu + \nu exp(-ry)} = 0, \tag{A11}$$

$$\varphi'(y) - \frac{r\nu(1-\nu)\lambda\Delta\gamma exp(ry)}{((1-\nu)exp(ry) + \nu)^2} = 0. \tag{A12}$$

For $y = 0$, (A10) and (A11) imply that $\gamma(\bar{\theta}) = \gamma(\underline{\theta}) = \gamma$ and thus $\Delta\gamma = 0$. But then (A12) also holds at $y = 0$ which shows that we have found the optimal allocations for $y = 0$ and that (A2) holds.
■

From Lemma 1, we get the simpler expression for $U_O(\theta, a)$ as

$$U_O(\theta, a) = \alpha(\theta) + \gamma(a + \theta) + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2}{2}.$$

Again from Lemma 1, and in particular the fact that $y = 0$ is optimal, the participation constraint (A4) implies that $\tilde{U}_O(\bar{\theta}) = \tilde{U}_O(\underline{\theta}) = -\gamma a$. Thus $U_O(\bar{\theta}, a) = U_O(\underline{\theta}, a)$ which is given by

$$\beta a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + U_O(\theta, a) = 0 \quad \forall\theta \in \Theta, \tag{A13}$$

or to put it differently

$$\alpha(\theta) + \gamma\theta + (\beta + \gamma)a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + \left(\frac{1}{\mu} - r\eta^2\right)\frac{\gamma^2}{2} = 0 \quad \forall\theta \in \Theta.$$

From which, we immediately deduce that

$$\alpha(\underline{\theta}) - \alpha(\bar{\theta}) = \gamma\Delta\theta > 0.$$

The first-order optimality conditions (A8), (A10) and (A11) immediately yield:

$$1 - a = 1 - \beta - \gamma = r\sigma^2\beta,$$

$$1 - a + \frac{1}{\mu}(1 - \gamma) = r\eta^2\gamma.$$

From this, we derive the optimal contract parameters:

$$\beta_B^* = \frac{1}{1 + r\sigma^2}\frac{r\eta^2}{\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2} + r\eta^2}, \text{ and } \gamma_B^* = \frac{\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2}}{\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2} + r\eta^2}.$$

The optimal investment level is now given by

$$a_B^* = \frac{1+\gamma_B^* r\sigma^2}{1+r\sigma^2}.$$

The comparison of those variables with their values when θ is non-observable is straightforward.

For further reference (see below the proof of Proposition 6), note that we may also rewrite the principal's optimization problem after having inserted the expression of $U_O(\theta, a)$ from (A13) as:

$$\max_{(a,\beta,\gamma)} a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r\eta^2\gamma^2}{2} - I \text{ subject to (6)}$$

which amounts to

$$\max_{(\beta,\gamma)} \mathcal{W}_B^*(\beta,\gamma) \equiv \beta + \gamma - \frac{(\beta+\gamma)^2}{2} + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r\eta^2\gamma^2}{2} - \frac{r\sigma^2\beta^2}{2} - I.$$

■

Proof of Proposition 6. The proof is identical to that of Proposition 3. In particular $\Delta W^*(\gamma) \equiv \max_{(\beta,\gamma)} \mathcal{W}_B^*(\beta,\gamma) - \max_{(\beta,\gamma)} \mathcal{W}_U^*(\beta,\gamma)$ is as in expression (A1). ■

Proof of Proposition 7. To induce participation from the operator before he learns the realization of θ, the ex ante participation constraint (13) must again be satisfied. At the same time, it is worth noticing that the first-stage effort still satisfies the incentive constraint (1). Under unbundling and asymmetric information, the optimal contract therefore solves the following maximization problem:

$$\max_{(a,U_B,U_O(\cdot),\beta,\gamma(\cdot))} E_\theta\left(e(\theta,\gamma(\theta),a) - \frac{\mu}{2}(e(\theta,\gamma(\theta),a) - a - \theta)^2 - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} - \frac{r\eta^2\gamma^2(\theta)}{2} - U_O(\theta,a)\right) - U_B - I$$

subject to (1), (2), (3), (13) and (16).

Both participation constraints (2) and (13) and the truthtelling constraint (16) are binding at the optimum. (2) binding means $U_B = 0$. The fact that both (13) and (16) are binding allows us to solve explicitly this system of equations for $(U_O(\bar{\theta},a), U_O(\underline{\theta},a))$ thanks to the CARA specification. We get the following expressions of the operator's payoff in each state of nature:

$$U_O(\bar{\theta},a) = \Delta\theta\gamma(\underline{\theta}) + \frac{1}{r} ln\left(1-\nu+\nu exp(-r\Delta\theta\gamma(\underline{\theta}))\right),$$

$$U_O(\underline{\theta},a) = \frac{1}{r} ln\left(1-\nu+\nu exp(-r\Delta\theta\gamma(\underline{\theta}))\right).$$

Inserting those expressions of the builder's and the operator's payoffs into the maximand leads to the more compact expression of expected social welfare:

$$\max_{(\beta,\gamma(\cdot))} \hat{W}_U(\beta,\gamma) \equiv \beta - (1+r\sigma^2)\frac{\beta^2}{2} + E_\theta\left(\frac{1}{\mu}\left(\gamma(\theta) - \frac{\gamma^2(\theta)}{2}\right) - \frac{r\eta^2\gamma^2(\theta)}{2}\right) + \varphi(\Delta\theta\gamma(\underline{\theta})) - I$$

Optimizing with respect β, we find $\hat{\beta}_U = \beta_U^*$. Therefore,

$$\hat{a}_U = \hat{\beta}_U^* = a_U^* = \frac{1}{1+r\sigma^2}, \tag{A14}$$

Optimizing with respect to $\gamma(\bar{\theta})$, we find:

$$\hat{\gamma}_U(\bar{\theta}) = \gamma_U^* = \frac{\frac{1}{\mu}}{\frac{1}{\mu} + r\eta^2}. \tag{A15}$$

This finally yields the following expression of the principal's expected payoff under unbundling (where we use the notation $\gamma = \gamma(\underline{\theta})$):

$$\max_{\beta} \hat{W}_U(\beta, \gamma) = \frac{1}{2(1 + r\sigma^2)} + \frac{\nu}{2\mu(1 + \mu r\eta^2)} + (1 - \nu)\left(\frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r\eta^2\gamma^2}{2}\right) + \varphi(\Delta\theta\gamma) - I. \tag{A16}$$

Optimizing with respect to γ, we finally get:

$$\hat{\gamma}_U(\underline{\theta}) = \Phi(\Delta\theta, \hat{\gamma}_B(\underline{\theta})) < \gamma_B^* \tag{A17}$$

where the function $\Phi(\Delta\theta, \gamma) = \frac{\frac{1}{\mu} - \frac{\nu\Delta\theta(1-exp(-r\Delta\theta\gamma))}{1-\nu+\nu exp(-r\Delta\theta\gamma)}}{\frac{1}{\mu}+r\eta^2}$ is strictly decreasing in γ. Since we have $\Phi(\Delta\theta, 0) > 0$ and $\Phi(\Delta\theta, 1) < \frac{\frac{1}{\mu}}{\frac{1}{\mu}+r\eta^2} < 1$, (A17) has a unique solution $\hat{\gamma}_U(\underline{\theta}) \in (0, 1)$. ■

Proof of Proposition 8. Take a level of first-stage effort a that the principal wants to implement. Suppose that the principal wants to implement this effort a with revenues sharing schemes such that $\Delta\gamma > 0$. (The case where the principals wants to implement this effort a with a contract such that $\Delta\gamma = 0$ at minimal agency costs is easy and will lead also to conditions (17) and (19) but rewritten under those specific circumstances.) From (17), we know that incentive compatibility implies

$$U_O(\bar{\theta}, a) - U_O(\underline{\theta}, a) = \Delta\theta\gamma(\underline{\theta}) + \xi \tag{A18}$$

for some $\xi > 0$. Suppose that the agent was to deviate to an alternate effort $\tilde{a}$. When knowing ex post $\bar{\theta}$, the agent tells the truth whenever

$$U_O(\bar{\theta}, a) + \gamma(\bar{\theta})(\tilde{a} - a) \geq U_O(\underline{\theta}, a) + \gamma(\underline{\theta})(\tilde{a} - a) + \Delta\theta\gamma(\underline{\theta})$$

or alternatively when

$$\tilde{a} \geq a - \frac{\xi}{\Delta\gamma}.$$

Let us turn to the first-stage moral hazard incentive constraint. We are considering only deviations that do not induce a bad firm to lie, those deviations are thus sufficiently close to a to avoid the countervailing problem mentioned in Footnote 38. We are thus describing a priori a larger set of implementable allocations.

In full generality, the first-stage moral hazard incentive constraint can still be written as in (14). This incentive constraint can be written in two different ways, depending on whether the first-stage effort deviation induces a lie or not in state $\bar{\theta}$.

• Consider first deviations such that $\tilde{a} \geq a - \frac{\xi}{\Delta\gamma}$. These deviations are dominated when the payoff of choosing a satisfies

$$E_\theta\left(v\left(\beta a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + U_O(\theta, a)\right)\right)$$
$$\geq \max_{\tilde{a} \geq a - \frac{\xi}{\Delta\gamma}} E_\theta\left(v\left(\beta\tilde{a} - \frac{\tilde{a}^2}{2} - \frac{r\sigma^2\beta^2}{2} + U_O(\theta, a) + \gamma(\theta)(\tilde{a} - a)\right)\right).$$

Taking into account (A18) gives us the following condition written in terms of certainty equivalents

$$\beta a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + U_O(\underline{\theta}, a) - \frac{1}{r}ln\left(1 - \nu + \nu exp(-r(\Delta\theta\gamma(\underline{\theta}) + \xi))\right) \geq$$

$$\max_{\tilde{a}\geq a-\frac{\xi}{\Delta\gamma}} \beta\tilde{a}-\frac{\tilde{a}^2}{2}-\frac{r\sigma^2\beta^2}{2}+\gamma(\underline{\theta})(\tilde{a}-a)+U_O(\underline{\theta},a)-\frac{1}{r}ln\left(1-\nu+\nu exp(-r(\Delta\gamma(\tilde{a}-a)+\Delta\theta\gamma(\underline{\theta})+\xi))\right)$$

or simplifying

$$\beta a-\frac{a^2}{2}-\frac{1}{r}ln\left(1-\nu+\nu exp(-r(\Delta\theta\gamma(\underline{\theta})+\xi))\right)\geq$$

$$\max_{\tilde{a}\geq a-\frac{\xi}{\Delta\gamma}} \beta\tilde{a}-\frac{\tilde{a}^2}{2}+\gamma(\underline{\theta})(\tilde{a}-a)-\frac{1}{r}ln\left(1-\nu+\nu exp(-r(\Delta\gamma(\tilde{a}-a)+\Delta\theta\gamma(\underline{\theta})+\xi))\right). \tag{A19}$$

The maximand on the right-hand side is concave in $\tilde{a}$ since $\Delta\gamma\geq 0$ (a standard monotonicity condition coming from adding up truthtelling constraints in state $\bar{\theta}$ and $\underline{\theta}$.[45]) The maximum must also be achieved at a by definition. The first-order condition for optimality evaluated at $\tilde{a}=a$ is

$$a=\beta+\frac{\nu\gamma(\bar{\theta})+(1-\nu)\gamma(\underline{\theta})exp(r(\Delta\theta\gamma(\underline{\theta})+\xi))}{\nu+(1-\nu)exp(r(\Delta\theta\gamma(\underline{\theta})+\xi))}. \tag{A20}$$

• Consider now deviations such that $\tilde{a}\leq a-\frac{\xi}{\Delta\gamma}$. These deviations are dominated when

$$E_\theta\left(v\left(\beta a-\frac{a^2}{2}-\frac{r\sigma^2\beta^2}{2}+U_O(\theta,a)\right)\right)$$

$$\geq \max_{\tilde{a}\leq a-\frac{\xi}{\Delta\gamma}} E_\theta\left(v\left(\beta\tilde{a}-\frac{\tilde{a}^2}{2}-\frac{r\sigma^2\beta^2}{2}+U_O(\theta,a)+\gamma(\theta)(\tilde{a}-a)\right)\right).$$

Taking again into account (A18), this gives us the following condition

$$\beta a-\frac{a^2}{2}\geq \max_{\tilde{a}\leq a-\frac{\xi}{\Delta\gamma}} \beta\tilde{a}-\frac{\tilde{a}^2}{2}+\gamma(\underline{\theta})(\tilde{a}-a). \tag{A21}$$

Suppose first that

$$\beta+\gamma(\underline{\theta})<a-\frac{\xi}{\Delta\gamma}. \tag{A22}$$

Then the maximum on the right-hand side of (A21) is achieved for $\beta+\gamma(\underline{\theta})$ but the inequality in (A21) then cannot hold.

Suppose *a contrario* that

$$\beta+\gamma(\underline{\theta})\geq a-\frac{\xi}{\Delta\gamma}. \tag{A23}$$

Then, the maximum on the right-hand side of (A21) is achieved for $a-\frac{\xi}{\Delta\gamma}$ and (A21) implies that

$$(\beta+\gamma(\underline{\theta}))a-\frac{a^2}{2}\geq(\beta+\gamma(\underline{\theta}))\left(a-\frac{\xi}{\Delta\gamma}\right)-\frac{1}{2}\left(a-\frac{\xi}{\Delta\gamma}\right)^2.$$

Hence, preventing downward deviations requires to set

$$\frac{\xi}{2\Delta\gamma}\geq a-\beta-\gamma(\underline{\theta}). \tag{A24}$$

which also implies (A23) as posited.

[45]See for instance Laffont and Martimort (2002, Chapter 2).

Let us turn to the principal's problem. The objective writes again as in (A6) so that principal's problem becomes thus

$$\max_{(a,U_O(\cdot),\beta,\gamma(\cdot),\xi)} \quad (1-\beta)a + E_\theta\left(\frac{1}{\mu}\left(\gamma(\theta) - \frac{\gamma^2(\theta)}{2}\right) - \frac{r\eta^2\gamma^2(\theta)}{2} - U_O(\theta, a)\right) - I$$

subject to (13), (A18), (A20) and (A24).

Both the participation constraint and (13) and the truthtelling constraint (A18) are binding at the optimum. This allows us to solve explicitly this system of equations for $(U_O(\bar{\theta}, a), U_O(\underline{\theta}, a))$ thanks to the CARA specification. We get the following expressions of the firm's payoff in each state of nature:

$$\beta a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + U_O(\bar{\theta}, a) = \Delta\theta\gamma(\underline{\theta}) + \xi + \frac{1}{r} ln\left(1 - \nu + \nu exp(-r(\Delta\theta\gamma(\underline{\theta}) + \xi))\right),$$

$$\beta a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + U_O(\underline{\theta}, a) = \frac{1}{r} ln\left(1 - \nu + \nu exp(-r(\Delta\theta\gamma(\underline{\theta}) + \xi))\right).$$

Inserting those expressions of the into the maximand leads to more compact expression:

$$\max_{(\beta,\gamma(\cdot),\xi)} a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + E_\theta\left(\frac{1}{\mu}\left(\gamma(\theta) - \frac{\gamma^2(\theta)}{2}\right) - \frac{r\eta^2\gamma^2(\theta)}{2}\right) + \varphi(\Delta\theta\gamma(\underline{\theta}) + \xi) - I$$

subject to (A20) and (A24).

The above objective is decreasing in ξ. Moreover, from (A20), a is also decreasing in ξ and so is $a - \frac{a^2}{2}$ as long as $a < 1$, a condition that is satisfied on the solution as we will see. Henceforth, (A24) is binding. Inserting the expression of a coming from (A20) into ξ finally yields (18) and (19). ■

Proof of Proposition 9. Under bundling, the principal's optimization problem can be rewritten as:

$$\max_{(\beta,\gamma(\underline{\theta}),\Delta\gamma\geq 0)} a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + \nu\left(\frac{1}{\mu}\left(\gamma(\underline{\theta}) + \Delta\gamma - \frac{(\gamma(\underline{\theta}) + \Delta\gamma)^2}{2}\right) - \frac{r\eta^2(\gamma(\underline{\theta}) + \Delta\gamma)^2}{2}\right)$$

$$+(1-\nu)\left(\frac{1}{\mu}\left(\gamma(\underline{\theta}) - \frac{\gamma^2(\underline{\theta})}{2}\right) - \frac{r\eta^2\gamma^2(\underline{\theta})}{2}\right) + \varphi(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma)) - I \text{ subject to (19).}$$

We are going to show that a pooling contract is locally optimal when $\Delta\theta$ is small enough.

Inserting a from (19) into the maximand yields an expression in $(\beta, \gamma(\underline{\theta}), \Delta\gamma)$, namely

$$\hat{W}_B(\beta, \gamma(\underline{\theta}), \Delta\gamma) = a(\beta, \gamma(\underline{\theta}), \Delta\gamma) - \frac{a^2(\beta, \gamma(\underline{\theta}), \Delta\gamma)}{2} - \frac{r\sigma^2\beta^2}{2} + \nu\left(\frac{1}{\mu}\left(\gamma(\underline{\theta}) + \Delta\gamma - \frac{(\gamma(\underline{\theta}) + \Delta\gamma)^2}{2}\right) - \frac{r\eta^2(\gamma(\underline{\theta}) + \Delta\gamma)^2}{2}\right)$$

$$+(1-\nu)\left(\frac{1}{\mu}\left(\gamma(\underline{\theta}) - \frac{\gamma^2(\underline{\theta})}{2}\right) - \frac{r\eta^2\gamma^2(\underline{\theta})}{2}\right) + \varphi(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma)) - I.$$

First, observe that

$$\frac{\partial\xi}{\partial\Delta\gamma} = \frac{\frac{4\nu\Delta\gamma}{\nu + (1-\nu)exp(r(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma)))}}{1 + \frac{2\nu(1-\nu)r(\Delta\gamma)^2 exp(r(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma)))}{(\nu + (1-\nu)exp(r(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma))))^2}} \Rightarrow \frac{\partial\xi}{\partial\Delta\gamma}|_{\Delta\gamma=0} = 0,$$

and

$$\frac{\partial\xi}{\partial\gamma(\underline{\theta})} = \frac{-\frac{2\nu(1-\nu)r(\Delta\gamma)^2\Delta\theta exp(r(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma)))}{(\nu + (1-\nu)exp(r(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma))))^2}}{1 + \frac{2\nu(1-\nu)r(\Delta\gamma)^2 exp(r(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma)))}{(\nu + (1-\nu)exp(r(\Delta\theta\gamma(\underline{\theta}) + \xi(\gamma(\underline{\theta}), \Delta\gamma))))^2}} < 0 \Rightarrow \Delta\theta + \frac{\partial\xi}{\partial\gamma(\underline{\theta})} > 0 \text{ and } \frac{\partial\xi}{\partial\gamma(\underline{\theta})}|_{\Delta\gamma=0} = 0.$$

Those expressions help us to find in particular the first-order derivatives of $\hat{W}_B(\beta,\gamma(\underline{\theta}),\Delta\gamma)$ with respect to β, $\Delta\gamma$ and $\gamma(\underline{\theta})$ evaluated at $\Delta\gamma=0$. Taking into account that $\xi(\gamma(\underline{\theta}),0)=0$, we get first:

$$\frac{\partial \hat{W}_B}{\partial \beta}(\beta,\gamma(\underline{\theta}),0) = 1 - a(\beta,\gamma(\underline{\theta}),0) - r\sigma^2\beta, \tag{A25}$$

$$\frac{\partial \hat{W}_B}{\partial \Delta\gamma}(\beta,\gamma(\underline{\theta}),0) = \nu\left(\frac{1-a(\beta,\gamma(\underline{\theta}),0)}{\nu+(1-\nu)exp(r\Delta\theta\gamma(\underline{\theta}))} + \frac{1-\gamma(\underline{\theta})}{\mu} - r\eta^2\gamma(\underline{\theta})\right), \tag{A26}$$

$$\frac{\partial \hat{W}_B}{\partial \gamma(\underline{\theta})}(\beta,\gamma(\underline{\theta}),0) = 1 - a(\beta,\gamma(\underline{\theta}),0) + \frac{1-\gamma(\underline{\theta})}{\mu} - r\eta^2\gamma(\underline{\theta}) - \frac{\nu(1-\nu)\Delta\theta(1-exp(-r\Delta\theta\gamma(\underline{\theta})))}{1-\nu+\nu exp(-r\Delta\theta\gamma(\underline{\theta}))}. \tag{A27}$$

For an interior solution in $\gamma(\underline{\theta})$ corresponding to $\Delta\gamma=0$, we must have $\frac{\partial \hat{W}_B}{\partial \gamma(\underline{\theta})}(\beta,\gamma(\underline{\theta}),0)=0$. Using (A27), we can simplify the expression of $\frac{\partial \hat{W}_B}{\partial \Delta\gamma}(\beta,\gamma(\underline{\theta}),0)$ in (A26) and obtain

$$\frac{\partial \hat{W}_B}{\partial \Delta\gamma}(\beta,\gamma(\underline{\theta}),0) = \frac{\nu(1-\nu)(1-exp(-r\Delta\theta\gamma(\underline{\theta})))}{1-\nu+\nu exp(-r\Delta\theta\gamma(\underline{\theta}))}(\nu\Delta\theta - (1-a(\beta,\gamma(\underline{\theta}),0))). \tag{A28}$$

Hence, $\frac{\partial \hat{W}_B}{\partial \Delta\gamma}(\beta,\gamma(\underline{\theta}),0) < 0$ when $\Delta\theta$ is small enough (notice indeed that (A26) and the analysis below of the pooling case $\gamma=\gamma(\bar{\theta})=\gamma(\underline{\theta})$ altogether imply that $1-a(\beta,\gamma(\underline{\theta}),0)$ is close to $1-a_B^* = \frac{r\sigma^2}{1+r\sigma^2}(1-\gamma_B^*)>0$ so that the right-hand side of (A28) is strictly positive when $\Delta\theta$ is small enough). Then, $\Delta\gamma=0$ is a local optimum.

With such pooling contract and denoting $\gamma=\gamma(\bar{\theta})=\gamma(\underline{\theta})$ (notice that this pooling case is also relevant for the case of renegotiation-proof contracts in which case $\Delta\theta$ is arbitrary), the principal's problem becomes:

$$\max_{(\beta,\gamma)} a - \frac{a^2}{2} - \frac{r\sigma^2\beta^2}{2} + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r\eta^2\gamma^2}{2} + \varphi(\Delta\theta\gamma) - I \text{ subject to (6).}$$

Inserting the constraint into the maximand, we define

$$\hat{W}_B(\beta,\gamma) = \beta+\gamma - \frac{(\beta+\gamma)^2}{2} - \frac{r\sigma^2\beta^2}{2} + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r\eta^2\gamma^2}{2} + \varphi(\Delta\theta\gamma) - I.$$

Optimizing first with respect to β, we find an optimal value

$$\beta = \frac{1-\gamma}{1+r\sigma^2}.$$

Note that

$$\max_{\beta} \hat{W}_B(\beta,\gamma) = \frac{1}{2(1+r\sigma^2)} + \frac{r\sigma^2}{1+r\sigma^2}\left(\gamma - \frac{\gamma^2}{2}\right) + \frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r\eta^2\gamma^2}{2} + \varphi(\Delta\theta\gamma) - I. \tag{A29}$$

Optimizing with respect to γ finally yields the following implicit definition of $\hat{\gamma}_B$:

$$\hat{\gamma}_B = \Psi(\Delta\theta,\hat{\gamma}_B) \tag{A30}$$

where $\Psi(\Delta\theta,\gamma) = \frac{\frac{1}{\mu}+\frac{r\sigma^2}{1+r\sigma^2}-\frac{\nu(1-\nu)\Delta\theta(1-exp(-r\Delta\theta\gamma))}{1-\nu+\nu exp(-r\Delta\theta\gamma)}}{\frac{1}{\mu}+\frac{r\sigma^2}{1+r\sigma^2}+r\eta^2}$ is decreasing in γ. Since we have $\Psi(\Delta\theta,0)>0$ and $\Psi(\Delta\theta,1) < \frac{\frac{1}{\mu}+\frac{r\sigma^2}{1+r\sigma^2}}{\frac{1}{\mu}+\frac{r\sigma^2}{1+r\sigma^2}+r\eta^2} < 1$, (A30) has a unique solution $\hat{\gamma}_B\in(0,1)$. It is immediate to check that

$$\hat{\gamma}_B < \gamma_B^*.$$

Since $\Psi(\Delta\theta, \gamma) > \Phi(\Delta\theta, \gamma)$, the implicit solutions to (A17) and (A30) are such that

$$\hat{\gamma}_B > \hat{\gamma}_U.$$

The optimal effort is finally

$$\hat{a}_B = \frac{1 + r\sigma^2 \hat{\gamma}_B}{1 + r\sigma^2} < a_B. \tag{A31}$$

For further references (see below the proof of Proposition 11), observe that $\Psi(\Delta\theta, \gamma)$ is strictly decreasing in $\Delta\theta$. Hence, $\hat{\gamma}_B$ is a decreasing function of $\Delta\theta$ that converges towards 0 as $\Delta\theta$ goes to ∞. Provided that $\Delta\theta\hat{\gamma}_B$ converges also towards zero as $\Delta\theta$ goes to ∞, a Taylor approximation of the right-hand side of (A30) yields

$$\hat{\gamma}_B \underset{\Delta\theta \to +\infty}{\approx} \frac{\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2} - r\nu(1-\nu)(\Delta\theta)^2\hat{\gamma}_B}{\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2} + r\eta^2}.$$

Finally, we obtain the following approximation $\hat{\gamma}_B \underset{\Delta\theta \to +\infty}{\approx} \dfrac{\frac{1}{\mu} + \frac{r\sigma^2}{1+r\sigma^2}}{r\nu(1-\nu)\Delta\theta^2}$. In particular,

$$\lim_{\Delta\theta \to +\infty} \Delta\theta\hat{\gamma}_B = 0. \tag{A32}$$

■

Proof of Proposition 10. Immediate from the text. ■

Proof of Proposition 11. Remember that, under bundling and when either $\Delta\theta$ is small enough or if one insists on renegotiation-proof contracts, the principal's welfare expressed as a function of $\gamma = \gamma(\underline{\theta})$ writes as (A29) whereas it is (A16) under unbundling. From this it immediately follows that

$$\Delta\hat{\mathcal{W}}(\gamma) = \max_\beta \hat{\mathcal{W}}_B(\beta,\gamma) - \max_\beta \hat{\mathcal{W}}_U(\beta,\gamma) = \frac{r\sigma^2}{1+r\sigma^2}\left(\gamma - \frac{\gamma^2}{2}\right) + \nu\left(\frac{1}{\mu}\left(\gamma - \frac{\gamma^2}{2}\right) - \frac{r\eta^2\gamma^2}{2} - \frac{1}{2\mu(1+\mu r\eta^2)}\right).$$

This quadratic function is negative in the neighborhood of $\gamma = 0$ and positive when γ is close to γ_U^*. It has a unique root in $(0, \gamma_U^*)$. When $\Delta\theta$ is small enough, $\hat{\gamma}_U$ is close to γ_U^* and thus $\Delta\hat{\mathcal{W}}(\hat{\gamma}_U) > 0$. This in turn implies $\max_{(\beta,\gamma)} \hat{\mathcal{W}}_B(\beta,\gamma) > \max_{(\beta,\gamma)} \hat{\mathcal{W}}_U(\beta,\gamma)$ and bundling dominates.

When instead $\Delta\theta$ is large enough, (A32) implies that

$$\lim_{\Delta\theta \to +\infty} \Delta\hat{\mathcal{W}}(\hat{\gamma}^B) = -\frac{\nu}{2\mu(1+\mu r\eta^2)} < 0$$

and unbundling dominates. ■