A Mechanism Design Approach to Climate Agreements*

David Martimort[†] and Wilfried Sand-Zantman[‡]

This Version: November 13, 2012

Abstract: We analyze international environmental agreements in contexts with asymmetric information, voluntary participation by sovereign countries and possibly limited enforcement. Taking a mechanism design perspective, we study how countries can agree on effort targets and compensations to take into account multilateral externalities. We highlight a trade-off between solving free riding in effort provision at the intensive margin for participating countries and free riding at the extensive margin to ensure participation of all countries. We also show that the optimal mechanism admits a simple approximation by menus with attractive implementation properties. Finally, we also highlight how limits on enforcement and commitment might strongly hinder performances of optimal mechanisms.

Keywords: public goods, asymmetric information, global warming, mechanism design.

JEL Codes: Q54, D82, H23.

*We thank workshop participants at *Paris School of Economics*, the *Paris Environmental and Energy Economics Seminar*, the *CIRPEE (UQAM/HEC Montréal, Laval University)'s annual conference*, *CREST-LEI Paris*, *Frankfort*, *GREQAM-Marseille*, the *Congress of the Canadian Economic Association* in Calgary, and the *Workshop on the Economics of Climate Change CDC Paris*, but also Jean-Marc Bourgeon, Renaud Bourlès, Gabrielle Demange, Pierre Fleckinger, Bard Harstad, Jérôme Pouyet, François Salanié and Alain Trannoy for helpful comments on an earlier draft. We also thank Daniel Coublucq and Perrin Lefèbvre for outstanding research assistance. All errors are ours.

[†]Paris School of Economics-EHESS. Email: david.martimort@parisschoolofeconomics.eu

[‡]Toulouse School of Economics (GREMAQ-IDEI). Email: wsandz@tse-fr.eu.

1 Introduction

Global warming has by now become an issue of tantamount importance. If the *"business as usual"* scenario (thereafter BAU) prevails in the near future, expected damage could reach up to 13,8 % of GDP by 2200.[1] Because the corresponding distribution of costs and (possible) benefits is non-trivial, reaching an agreement among sovereign countries over the design of environmental policies that would slow down this process is a formidable challenge.

Economists taking a Coasian perspective would nevertheless suggest a strikingly simple solution. Under strong assumptions - complete information, absence of transaction costs, perfect enforceability of contractual arrangements- efficient outcomes should emerge from environmental negotiations whatever the distribution of current emissions to which countries might be entitled. The record of recent negotiations from Montreal, to Kyoto, Copenhagen, Cancun and Rio meetings and their repeated failures demonstrate that efficiency remains in fact by and large out of reach. Uncertainty on the physical processes behind global warming, conflicting views and private information on its local costs and benefits, limited ability of countries to commit to a path of emissions reductions are all ingredients that might significantly hinder this optimistic Coasian scenario.

To understand how those impediments to efficient negotiations might be circumvented, this paper takes a mechanism design perspective. We study the optimal design of environmental treaties when sovereign countries have private information on their abatement costs and must voluntary abide to the agreement, possibly under enforcement or commitment constraints. This analysis offers a number of important and novel insights that have no counterpart in a Coasian setting.

Two free-riding problems. In the context of a multilateral externalities problem as global warming, private information is of course the source of free riding *at the intensive margin.* Indeed, each privately informed country that participates to an agreement may exaggerate abatements costs, undersupply effort towards depollution and leave most of the burden of cost abatements on others. Such free riding has already received much attention in the mechanism design literature, noticeably in pure public good settings, and it is well known that its source is the possible conflict between incentive compatibility, participation constraints and budget balance.[2]

However, as pointed out by Chander and Tulkens (2008), free riding also bites, and it is more specific to environmental negotiations, *at the extensive margin.* Indeed,

[1]See Stern (2006).

[2]See Laffont and Maskin (1982) and Mailath and Postlewaite (1990) in general environments. Rob (1989), Neeman (1999) and Baliga and Maskin (2003) have developed specific applications targeted to environmental economics.

sovereign countries may also opt out of the negotiation and still enjoy the benefits of the agreement ratified by others. Note that when considering such deviation, a country forms conjectures on how others react which in turn determines its payoff of not joining. Should the remaining coalition disband with all countries adopting their BAU emissions or should ratifying countries stick to some restricted and conditional treaty? Incentives to free-ride by not participating to an agreement certainly depend on those conjectures. From a mechanism design perspective, those conjectures impact thus on the participation constraints that apply to the design problem; an ingredient absent from the received mechanism design literature on public goods provision.

The central case that we will consider below is the BAU scenario, i.e. the situation where everyone conjectures that, if any country refuses to participate, others follow their BAU policy. When all countries must therefore agree to enforce a mechanism, and for a sufficiently significant externality, efficiency is obtained even under asymmetric information. Otherwise, effort levels always lie somewhere in between their BAU level and the first best in order to curb free riding at the intensive margin. At the same time, and still as a means to prevent incentives to free-ride, inefficient countries that choose emissions which are too close to their BAU level are also asked to contribute to a fund. This fund is then used to subsidize the most efficient countries which instead expand their effort so as to better internalize the externality they exert on others.

Contributions to this fund cannot be too large without inducing countries with the highest opportunity costs of effort to leave the agreement. This important insight points at the existence of a trade-off between solving free riding at the intensive and at the extensive margins. Inefficient countries end up being indifferent between joining in or not, in which case they exert their BAU efforts while paying the expected externality they enjoy from the greater effort exerted by the most efficient parties.

Commitment and Enforcement problems. Barrett (2003) reports that the Kyoto Protocol suffers from (at least) two flaws.[3] First, countries could refuse to ratify without further being punished. Second, the protocol did not incorporate any compliance mechanism for ratifying countries. This suggests that the design of an agreement should also account for two further constraints, namely the impossibility to credibly commit to punish non-ratifiers and the difficulty in enforcing the agreement for ratifiers.

Looking first at the commitment problem, we analyze different conjectures on the credibility of the punishments that the coalition may impose on non-participating countries. Two polar cases are studied. When the mechanism does not stipulate any punishment, free riding at the extensive margin takes an extreme form and no incentive compatible allocation may outperform the BAU outcome. On the contrary, if participants to the treaty can *minmax* non-ratifiers (which of course requires non-credible

[3] On this issue, see also Aldy and Stavins (2007).

threats) the first best can always be implemented, an optimistic albeit unrealistic scenario.

Turning to the enforcement problem, we argue that what can be achieved by a treaty depends on the collective ability to guarantee that each ratifying country abides to the rule of the game once accepted. This is true even though internal political pressures at reelections time, lobbying, and incentives to foster short-term growth may push governments to renege on international agreements. Introducing an explicit *enforcement constraint* (harder to satisfy than the pure participation constraints) exacerbates inefficiencies. The set of inefficient countries which are just indifferent between joining in or not expands and less funds are available to subsidize efforts by the most efficient ones. Difficulties in enforcement make the BAU option more attractive.

Approximate implementation. Proposing a handy set of instruments that could be put into place to implement good climate-change policies is high on the practitioners' agenda.[4] In this respect, we also investigate how the optimal mechanism can be implemented or, at least, approximated in practice. Our analysis reveals that a simple two-items menu that specifies either a fixed contribution or a Pigovian subsidy per unit of effort *cum* a stronger contribution may perform quite well to approximate the optimal mechanism. Countries are then split into two groups. Efficient ones take the incentive option while inefficient ones just contribute a fixed amount to the fund. Numerical simulations testify that this menu reaches most of the welfare gains under asymmetric information. This might leave an optimistic view on the possibility of solving the climate-change problem even in non-Coasean environments.

Literature review. The existing literature on climate negotiations has insisted on the possible failures in reaching global agreements. The focus is on conditions for reaching efficiency while at the same time requiring the worldwide coalition to be robust to secessions. To tackle those issues, Chandler and Tulkens (1995, 1997) introduce the notion of γ-core for economies with multilateral externalities. They defined the worth of a coalition, assuming that countries outside the coalition play individual best responses. They demonstrated that the grand-coalition is feasible despite individual incentives to free ride at the extensive margin. Under complete information, efficiency may be compatible with a worldwide coalition. We borrow from this contribution an important concern on the role played by conjectures on how stringent participation constraints are. Under asymmetric information, efficiency is nevertheless far less easy to reach. To the best of our knowledge, there has been almost no work addressing the multilateral externality problems in climate agreement taking a mechanism design perspective. An exception is Helm and Wirl (2011) who consider a two-country version of this problem where bargaining power is asymmetrically distributed and the uninformed country

[4] See for instance Bradford (2008) and Guesnerie (2008) among others.

designs a mechanism controlling collective emissions. Our paper significantly differs from this one by taking a more normative approach in a framework with multiple privately informed countries and a more even distribution of bargaining power.

Another important line of research (Carraro and Siniscalco 1993, 1995, Barrett 1994) has instead focused on incentives to form coalitions by imposing external and internal stability criterions similar to those developed in earlier cartel theory. Subsequent research in the field (Carraro, 2005) has then stressed the importance of various institutional rules to ensure participation, stability, and solve the free-riding problem. There, institutional constraints are imposed at the outset and not derived from primitives. This stands in sharp contrast with the mechanism design approach that precisely derives optimal institutions from primitives - well-specified informational constraints and strategic behavior.[5] This approach is, by tradition, more normative and, by construction, does not care much on details of the negotiation process.

Another route away from the Coasean scenario which is complementary to ours consists in introducing commitment problems in dynamic games of complete information. In that vein, Beccherle and Tirole (2011), Battaglini and Harstad (2012) and Harstad (2012a, 2012b) analyze models where countries can limit global warming either by decreasing consumption or making some non-verifiable investments in abatement technologies. Countries may refrain from investing today as it would trigger less investment and more pollution from others tomorrow. Becherelle and Tirole (2011) show how today investments affect threat points in future negotiations and are thus chosen strategically in an incomplete contracting framework. Harstad (2012b) shows that short-term agreements can be worse than no contract at all from a welfare viewpoint. Harstad (2012a) derives optimal dynamic contracts when renegotiation allows to reach efficient outcomes.

At a more theoretical level, our paper also contributes on the mechanism design front. The mechanism design approach for public good provision[6] assumes that all participating agents have veto power, i.e., the fall-back option if anyone disagrees is no provision of the public good at all with zero payoffs. This assumption seems inadequate to tackle the specificities of environmental negotiations between sovereign countries, especially when various conjectures may be entertained on the commitment ability of coalitions and individual countries. This paper revamps the conflict between individual incentives, budget balance and participation constraints when those participation constraints explicitly take into account such conjectures. From a technical viewpoint, the characterization of such regime is complexified by the addition of type-dependent participation constraints to a mechanism design problem under budget bal-

[5] This stability program was developed in a complete information framework and often assumed away the possible heterogeneity between countries. On the difficulties in reaching agreements among heterogenous countries in a complete information setting, see Thoron (2008).

[6] See references above.

ance. We rely on and adapt techniques developed in Martimort and Stole (2011) to tackle those issues.

Organization of the paper. Section 2 presents the model. Section 3 first describes incentive feasible allocations. Focusing on the BAU scenario as the fall-back option, we delineate conditions for inefficient effort provision under asymmetric information. Finally, we analyze those inefficiencies and the properties of the nonlinear contribution schedule that implements second-best efforts. Section 4 investigates the impact of varying the commitment ability of the coalition to enforce punishments on non-ratifiers. Section 5 studies the enforcement problem. Finally, Section 6 analyzes the performances of simple policy instruments. Section 7 discusses the robustness of our findings and highlights a few alleys for further research. Proofs are in an Appendix.

2 The Model

Preferences and technology. Let consider a continuum of countries of unit mass which undertake activities that mitigate pollution emissions. By exerting a non-negative effort e_i, country i generates two kinds of benefits. The first benefits of size αe_i (where $\alpha \in [0,1)$) are purely *local* and accrue only to country i.[7] The second sort of benefits are instead *global*, worth $(1-\alpha)e_i$ and accrue to all countries worldwide. As α varies from zero to one, efforts go from having pure global to pure local consequences. Even though this modeling is consistent with αe_i being the pure local benefits of a clean environment,[8] a broader interpretation of this modeling is that the adoption of policies against global warming might have a more general positive economic impact at the local level (maybe by fostering growth through innovation in green technologies) but, as we will see below, these efforts are too low from a worldwide welfare point of view.

Countries are heterogeneous in terms of their marginal cost of exerting effort. For tractability, we adopt a quadratic formulation and assume that the disutility of effort writes as $C(e_i,\theta_i) = \frac{e_i^2}{2\theta_i}$, where θ_i is an efficiency parameter. Those costs should be understood in a broad sense, including not only technological but also opportunity costs (those possibly being associated to internal politics)[9] necessary to reach a given effort target. With that latter interpretation in mind, developed countries (at least some of them like the U.S.) may be considered as the least efficient ones while developing ones might actually face lesser internal constraints in adopting stringent regulations. Cost

[7]It will appear clearly in the sequel that the case $\alpha = 1$ is degenerate. There is no externality in that unlikely case and BAU is trivially optimal, a theoretical case that has no practical relevance.

[8]For instance, CO_2 is known as having a global impact whereas other greenhouse gazes like SO_2 or NO_x have also significant local impacts.

[9]Helm, Hepburn and Mash (2005) study the incentives of governments to implement lax carbon policies because of electoral concerns.

convexity captures the fact that emissions cannot be reduced too much without impairing the basic functioning of the economy by, for instance, imposing technological changes and adjustments that are increasingly harder to implement as efforts increase. Country i's payoff can be written as:

$$U_i = t_i + \alpha e_i + (1-\alpha)\mathcal{E} - \frac{e_i^2}{2\theta_i}.$$

$\mathcal{E}$ represents the "aggregate" effort taken worldwide.[10] The payment t_i stands for any financial compensation (taxes or subsidies) that this country may receive for undertaking the requested effort. The possibility of including monetary contributions into environmental treaties is indeed often explicit. For instance, Article 11 of the Kyoto Convention allows for the possibility of transfers from developed to developing countries under the aegis of an *International Green Fund*.[11]

Information. The efficiency parameters θ_i are independently drawn from the same cumulative distribution $F(\cdot)$ with support $\Theta = [\underline{\theta}, \bar{\theta}]$ (with $\underline{\theta} > 0$) and everywhere positive and atomless density $f(\theta) = F'(\theta)$. Let denote by $E_\theta(\cdot)$ the expectation operator with respect to θ.

The following monotonicity condition will ensure monotonicity of effort at the optimal mechanism under asymmetric information:

Assumption 1

$$\frac{d}{d\theta}\left(\frac{1-F(\theta)}{\theta f(\theta)}\right) \leq 0 \quad \forall \theta \in \Theta.^{12}$$

Country i has private information on its efficiency parameter θ_i while its effort in mitigating pollution is observable.[13] Countries cannot receive payments conditional on the realization of this efficiency parameter although efforts can be contractually specified and subsidized. In that respect, our model can also be applied when costs

[10]An alternative formulation of the objective would be $t_i + \alpha e_i + \beta\mathcal{E} - \frac{e_i^2}{2\theta_i}$ for some non-negative α and β. Normalizing by $\alpha + \beta$ and changing θ_i into $\theta_i(\alpha+\beta)$ gives us our posited formulation. The latter has the benefit of keeping the first best fixed as α changes. This simplifies comparative statics.

[11]Contributions may also be given a broader interpretation and be viewed as the benefits or costs that countries withdraw when climate negotiations are linked to negotiations on other issues such as R&D technology transfers, sovereign debt and trade agreements. (On this, see Barrett 2005.) Of course, those costs and benefits may entail deadweight losses that are abstracted away.

[12]Distributions (uniform, exponential, truncated normal...) satisfying the more common monotonicity of the hazard rate $\frac{d}{d\theta}\left(\frac{1-F(\theta)}{f(\theta)}\right) \leq 0$ (Bagnoli and Bergstrom, 2005) also satisfy the weaker Assumption 1.

[13]That efforts in curbing pollution emissions are publicly observable is actually a mild assumption. Indeed, much attention has recently been devoted by practitioners on this issue and they agree that a worldwide system of satellite observations to measure local emissions is technically feasible. Tirole (2008) forcefully recognizes this point.

are common knowledge but discriminatory mechanisms conditional on the countries' exact identity are banned.[14]

Finally, the following assumption requires that the externality is not too strong relative to the informational problem. We will show below that when the externality is large enough, an efficient allocation can still be implemented even under asymmetric information.

Assumption 2

$$\alpha > \alpha_1 = \frac{\underline{\theta}}{2E_{\tilde{\theta}}(\tilde{\theta}) - \underline{\theta}} \in (0, 1).$$

Assumption 2 certainly holds when the parameter α is close enough to one (the case of a weak externality) or when uncertainty on the productivity type θ is large enough so that $E_{\tilde{\theta}}(\tilde{\theta})$ is sufficiently above $\underline{\theta}$.

Mechanisms and incentive compatibility. Because efforts are publicly observable, a mechanism stipulates levels of compensation and effort for each country. However, the fact that countries are privately informed on their opportunity costs of exerting effort implies that payments and effort levels must be *incentive compatible*. We now turn to the description of such incentive compatibility allocations. By the Revelation Principle, there is no loss of generality in considering direct and truthful revelation mechanisms of the form $\{t(\hat{\theta}), e(\hat{\theta})\}_{\hat{\theta} \in \Theta}$. Those mechanisms determine compensations and effort levels as a function of a country's announcement $\hat{\theta}$ on its own type. In particular, those mechanisms replace any nonlinear contribution schedule $T(e)$ that would map observable effort levels into compensations. For technical reasons, we will assume that efforts and payments belong to a sufficiently large compact set; formally, $(e, t) \in [0, M] \times [-T, T]$ for M and T large enough.

This mechanism design approach relies implicitly on the use of a mediator (or and international external agency) who monitors and enforces, possibly under some observability constraints, the efforts made by treaty members.[15,16]

Following a truthful strategy, a type θ country exerts an effort $e(\theta)$. We rely on the Law of Large Numbers to identify the average global benefits of the countries'

[14]Such anonymous design was forcefully advocated by the Bush administration to justify its withdrawal from the 2001 Kyoto protocol when calling the treaty "unfair" for industrialized countries.

[15]This external party is often referred to in the informal literature. For instance Guesnerie (2008) has proposed mechanisms to trade pollution permits that also heavily rely on an *International Bank for Emissions Allowance Acquisition.*

[16]Of course, the solution to this mechanism design problem gives us an upper bound on aggregate welfare. More decentralized bargaining procedures may fail to reach the frontier of the set of incentive-feasible allocations under asymmetric information. See for instance Martimort and Moreira (2010) for a result along these lines in the context of public good provision.

efforts with its expected value, i.e., $(1-\alpha)\mathcal{E} \equiv (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta}))$. We may then define the equilibrium payoff $U(\theta)$ of a country with type θ as:

$$U(\theta) = t(\theta) + \alpha e(\theta) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta})) - \frac{e^2(\theta)}{2\theta}.$$

Incentive compatibility implies:

$$U(\theta) = \max_{\hat{\theta}\in\Theta} t(\hat{\theta}) + \alpha e(\hat{\theta}) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta})) - \frac{e^2(\hat{\theta})}{2\theta}. \tag{1}$$

In the sequel, we shall repeatedly use a more compact (dual) characterization of incentive compatibility by using the rent $U(\theta)$ instead of the payment $t(\theta)$ together with an effort level. An allocation is thus a pair $(U(\theta), e(\theta))$.

Budget balance. Assuming that no external source of funds is available, i.e., the mechanism must be self-financed, the following budget balance condition must also hold:

$$E_{\tilde{\theta}}(t(\tilde{\theta})) \leq 0.$$

It will be often useful to rewrite this constraint as:

$$E_{\tilde{\theta}}\left(e(\tilde{\theta}) - \frac{e^2(\tilde{\theta})}{2\tilde{\theta}}\right) \geq E_{\tilde{\theta}}\left(U(\tilde{\theta})\right). \tag{2}$$

The overall expected surplus generated by the countries' efforts should be at least equal to their overall expected payoff. Of course, this constraint is binding (no waste of resources) for optimal mechanisms under all circumstances below.

Participation constraints. Finally, the mechanism must satisfy a set of participation constraints to ensure that all countries join the agreement. Those participation constraints depend on the commitment ability of the coalition to enforce actions in case any country deviates and does not join in. In the sequel, we will bear particular attention to the *BAU outcome* that is achieved when the whole coalition breaks down as soon as any of the countries refuses to participate to the mechanism.[17]

The corresponding fall-back option is thus the (symmetric) Bayesian-Nash equilibrium where countries non-cooperatively choose their efforts. Let denote by $U_N(\theta)$ the payoff of a type θ country in such equilibrium. We have

$$U_N(\theta) = \max_{e} \alpha e - \frac{e^2}{2\theta} + (1-\alpha)E_{\tilde{\theta}}(e_N(\tilde{\theta}))$$

where the Bayesian-Nash level of effort $e_N(\tilde{\theta})$ is

$$e_N(\theta) = \arg\max_{e} \alpha e - \frac{e^2}{2\theta} + (1-\alpha)E_{\tilde{\theta}}(e_N(\tilde{\theta})) = \alpha\theta.^{18}$$

[17]Sections 4 and 5 develop alternative specifications of those participation constraints.

This immediately leads to the following expression of payoffs under BAU:

$$U_N(\theta) = \frac{\alpha^2}{2}\theta + (1-\alpha)\alpha E_{\tilde{\theta}}(\tilde{\theta}).$$

Since countries know their types when deciding whether to join the treaty or not, the corresponding *ex post* participation constraints are written as:

$$U(\theta) \geq U_N(\theta) \quad \forall \theta \in \Theta. \tag{3}$$

Complete information benchmark. Suppose that the countries' efficiency parameters are common knowledge. Type-dependent instruments can be used to fix efforts at their target levels and compensate countries for those efforts according to the exact cost they incur. Ex post participation constraints (3) are easily satisfied. Of course, worldwide welfare is maximized for the first-best level of effort

$$e^{FB}(\theta) = \theta \quad \forall \theta \in \Theta.$$

Because a given country does not internalize the impact of its own effort on other countries' welfare, efforts are too low under BAU.

3 Asymmetric Information and Second-Best Mechanisms

3.1 Incentive Compatibility

Consider now the case where the countries' efficiency parameters are private information. Incentive compatibility constraints should now be added to characterize feasible allocations. Next lemma describes incentive compatible allocations.

Lemma 1 *An allocation $(U(\theta), e(\theta))$ is incentive compatible if and only if:*

1. *$U(\theta)$ is absolutely continuous with at each point of differentiability (i.e., almost everywhere)*

$$\dot{U}(\theta) = \frac{e^2(\theta)}{2\theta^2}. \tag{4}$$

2. *$e(\theta)$ is non-decreasing.*

[18]Thanks to our separability assumption between returns from local and global benefits, non-deviating countries choose the same effort level whatever their beliefs on the deviant (and negligible) country as long as they revert to a non-cooperative behavior.

By mimicking a slightly less efficient type $\theta - d\theta$, a type θ country can exert the same effort level but at a lower marginal cost. The marginal gains from doing so is approximatively $\frac{e^2(\theta-d\theta)}{2\theta^2}d\theta \approx \frac{e^2(\theta)}{2\theta^2}d\theta$. To induce information revelation, the most efficient type must pocket an extra reward $U(\theta) - U(\theta - d\theta) \approx \dot{U}(\theta)d\theta$ that is precisely worth these marginal gains as shown in (4). From Lemma 1, it immediately follows that an incentive compatible mechanism must give greater payoffs to the most efficient countries. Presumably, these countries are also those which may get more by entering the agreement than by opting for their fall-back option.

It is standard to neglect the monotonicity condition on $e(\cdot)$ and obtain a relaxed optimization problem whose solution satisfies that extra condition when Assumption 1 holds. We will follow that approach in the remainder of the paper. Adopting *ex ante efficiency* as an optimization criterion, the so relaxed second-best optimization problem consists in finding an (absolutely continuous) profile $U(\cdot)$ that solves:

$$(\mathcal{P}^{SB}) : \quad \max_{U(\cdot),e(\cdot)} E_{\tilde{\theta}}(U(\tilde{\theta})) \quad \text{subject to (2), (3) and (4).}$$

3.2 Conditions for Efficiency

As a preliminary step, we investigate under which conditions efficiency might still be compatible with asymmetric information.

Proposition 1 *Under asymmetric information and when the fall-back option is BAU, the first-best allocation cannot be implemented if and only if Assumption 2 holds.*

To understand this result, one must figure out the impact of α on both participation and incentives. Consider first the participation problem. When the parameter α is small, positive externalities are significant and the cost of disagreement is high. This relaxes participation constraints and makes cooperation more attractive. However, on the incentives side, countries do not care much about the local impact of their effort and the incentives to free-ride by reducing efforts are large. Avoiding such free riding requires large compensations to stimulate provision. When α is small enough, the gains from cooperation are sufficiently large to compensate for the incentive cost. The first-best allocation can still be implemented.

When α is instead large enough, the global impact of each countries' individual effort is less significant. Countries choose efforts close to the first best even when they do not cooperate. By the same token, the gains from cooperation are also small. Although there is less free riding in effort provision, the gains from cooperation are too small to compensate for the incentive problem and allow efficiency.

3.3 Two Free-Riding Problems

We now characterize second-best allocations with BAU as the fall-back option. Inefficiencies depend on the tension between incentive compatibility, participation and budget balance. In this respect, we will distinguish two scenarios. In the first one, all countries except the less efficient ones strictly gain from joining the mechanism. Effort levels always remain above BAU. These *weak distortions* arise when the gains from cooperation are rather large. In the second scenario, i.e., for *strong distortions*, inefficiencies are more pronounced. Only the most efficient countries strictly prefer joining in. Less efficient ones keep on exerting their BAU effort level.

To describe more precisely those scenarios, a detour consisting in defining a few auxiliary variables is useful. Consider an effort schedule $\bar{e}(\theta,\zeta)$ and a critical type $\theta^*(\zeta)$ both parameterized by some parameter $\zeta \geq 1$

$$\bar{e}(\theta,\zeta) = \frac{\theta}{1+\frac{\zeta-1}{\zeta}\frac{1-F(\theta)}{\theta f(\theta)}} \tag{5}$$

and

$$\begin{cases} \frac{1-F(\theta^*(\zeta))}{\theta^*(\zeta) f(\theta^*(\zeta))} = \frac{1-\alpha}{\alpha}\frac{\zeta}{\zeta-1} & \text{if } \zeta \geq \zeta^*(\alpha) \ \textbf{(strong distortions)} \\ \theta^*(\zeta) = \underline{\theta} & \text{if } \zeta \in [1,\zeta^*(\alpha)) \ \textbf{(weak distortions)} \end{cases} \tag{6}$$

where

$$\zeta^*(\alpha) = \frac{1}{1-\frac{1-\alpha}{\alpha}\underline{\theta} f(\underline{\theta})}. \tag{7}$$

Anticipating on our findings below, $\bar{e}(\theta,\zeta)$ will actually be the second-best effort level when $\zeta = \hat{\zeta}$ is the Lagrange multiplier for *an aggregate feasibility constraint* obtained by consolidating incentive, participation and budget-balance constraints altogether. All types which are less efficient than the critical type $\theta^*(\zeta)$ (when interior) are just indifferent between exerting the BAU effort and the second-best effort level, i.e., $\bar{e}(\theta^*(\hat{\zeta}),\hat{\zeta}) = e_N(\theta^*(\hat{\zeta}))$. More generally, the parameter ζ measures how the strength of distortions.

With these notations in mind, we derive this aggregate feasibility constraint as:

$$\int_{\underline{\theta}}^{\theta^*(\zeta)} \left(e_N(\theta) - \frac{e_N^2(\theta)}{2\theta}\right) f(\theta)d\theta + \int_{\theta^*(\zeta)}^{\bar{\theta}} \left(\bar{e}(\theta,\zeta) - \frac{\bar{e}^2(\theta,\zeta)}{2\theta}\left(1+\frac{1-F(\theta)}{\theta f(\theta)}\right)\right) f(\theta)d\theta$$

$$= \int_{\underline{\theta}}^{\theta^*(\zeta)} U_N(\theta) f(\theta)d\theta + U_N(\theta^*(\zeta))(1-F(\theta^*(\zeta))). \tag{8}$$

Condition (8) simply expresses the fact that total welfare has to be fully redistributed among countries participating to the mechanisms while keeping incentive compatibility. Incentive compatibility explains the extra informational distortion (proportional to $\frac{1-F(\theta)}{\theta f(\theta)}$ on the left-hand side of (8)). Inducing effort profiles closer to the first best is now costly because it exacerbates free riding at the intensive margin; the most efficient countries having then incentives to pretend being less so. Inducing participation imposes that feasible rent profiles must remain above their BAU level. We show below that those constraints are actually binding on an interval $\Omega^c = [\underline{\theta}, \theta^*(\zeta)]$ (which might be reduced to a single point in the case of weak distortions). The BAU effort and rent profiles are then found respectively both on the left-hand side of condition (8) which evaluates total welfare and on the right-hand side which measures expected payoffs.

Observe that $\zeta^*(\alpha)$ is decreasing with α and that $1 - \frac{1-\alpha}{\alpha}\underline{\theta} f(\underline{\theta}) > 0$ (hence $\zeta^*(\alpha) > 1$ holds) when

$$\alpha > \alpha_2 = \frac{1}{1 + \frac{1}{\underline{\theta} f(\underline{\theta})}}. \tag{9}$$

Assumption 3 below (which is for instance satisfied by the uniform distribution to which we will refer later on) simplifies the analysis without losing any insight:

Assumption 3

$$\alpha_2 \leq \alpha_1 \Leftrightarrow E_{\tilde{\theta}}(\tilde{\theta}) \leq \underline{\theta} + \frac{1}{2f(\underline{\theta})}.$$

This assumption allows us to have a clear separation between parameters constellations where either strong or weak distortions arise.

Distortion regimes. We are now ready to describe the two distortion regimes, depending on the value of the multiplier $\hat{\zeta}$ which is obtained as the unique solution to the aggregate feasibility condition (8).[19]

Proposition 2 *Suppose that the fall-back option is BAU and that Assumption 3 holds. There exists $\hat{\alpha} \in (\alpha_1, 1)$ that defines two different profiles of payoffs at the optimal mechanism.*

1. ***Weak distortions.*** *For $\alpha \in [\alpha_1, \hat{\alpha}]$, $\hat{\zeta} \in (1, \zeta^*(\alpha)]$.*
2. ***Strong distortions.*** *For $\alpha \in (\hat{\alpha}, 1)$, $\hat{\zeta} > \zeta^*(\alpha)$.*

The intuition for those distortions is better understood when thinking of α as being close enough to α_1, i.e., small enough while Assumption 2 being still satisfied. In that case, the efficiency gains from coordinating effort levels are rather strong but yet not large enough to allow efficiency. Nevertheless, we expect rather small allocative distortions. More formally, the multiplier $\hat{\zeta}$ should be close to one so that effort is almost efficient. When α increases, the gains from coordination are lower and asymmetric information has more bite. Distortions are stronger and $\hat{\zeta}$ increases.

Rents profile. Depending on the scenario, the rents profile has different shapes which are described in the next proposition.

[19]The proof of uniqueness can be found in the Appendix.

Proposition 3 *Suppose that the fall-back option is BAU and that Assumptions 1, 2 and 3 hold together. The second-best profile of rents $\bar{U}(\theta)$ is such that the participation constraint (3) is binding*

1. *only at $\underline{\theta}$ when $\hat{\zeta} \leq \zeta^*(\alpha)$ (**weak distortions**);*

2. *on an interval $\Omega^c = [\underline{\theta}, \theta^*(\hat{\zeta})]$ with non-empty interior when $\hat{\zeta} > \zeta^*(\alpha)$ (**strong distortions**).*

Efforts profile. Turning now to the characterization of effort levels, we get:

Proposition 4 *Suppose that the fall-back option is BAU and that Assumptions 1, 2 and 3 hold together. The second-best profile of effort levels $\bar{e}(\theta)$ is continuous, increasing in θ, greater than the BAU level but downward distorted below the first best everywhere except at $\bar{\theta}$.*

1. *If $\underline{\theta} = \theta^*(\hat{\zeta})$ (**weak distortions**), then*

$$\bar{e}(\theta) = \bar{e}(\theta, \hat{\zeta}) > e_N(\theta) \quad \forall \theta \in \Theta; \tag{10}$$

2. *If $\underline{\theta} < \theta^*(\hat{\zeta})$ (**strong distortions**), then*

$$\bar{e}(\theta) = \begin{cases} \bar{e}(\theta, \hat{\zeta}) > e_N(\theta) & \text{if } \theta \in \Omega = (\theta^*(\hat{\zeta}), \bar{\theta}] \\ e_N(\theta) & \text{if } \theta \in \Omega^c = [\underline{\theta}, \theta^*(\hat{\zeta})]. \end{cases} \tag{11}$$

When Assumption 2 holds, we already know that efficiency cannot be achieved. One cannot find incentive compatible payments that implement efficient effort levels and that give all types strictly more than their BAU payoffs. The participation constraint (3) must be binding somewhere.

Under asymmetric information, the most efficient types (such that $\theta \in \Omega = (\theta^*(\hat{\zeta}), \bar{\theta}]$) have now some incentives to claim being less efficient and produce less effort than requested by the mechanism. Those efficient types want to free ride by exerting less effort even when ratifying the mechanism. By doing so, they still earn some rent above BAU.

To limit those incentives to free ride at the intensive margin, the optimal mechanism plays both on effort targets and compensations. First, effort is reduced below the first best for all types (except the most efficient one). This first distortion reduces how much can be saved by the most efficient types by mimicking slightly less efficient ones. Second, the mechanism also asks for a greater contribution from the least efficient types to make their allocation less attractive. This second distortion might push the least efficient types to opt out of the mechanism. It thus exacerbates free riding at the extensive margin. To avoid such possibility, the inefficient countries' contributions are limited so that participation constraints are binding on the lower tail of the types distribution. This is so either at a single point or on a whole interval depending on whether distortions are weak or strong.

Summarizing, there is a trade-off between the free-riding problems at the intensive and at the extensive margins. Asymmetric information introduces a conflict between the most efficient countries' incentives to truthful reveal and the least efficient types' incentives to participate.

Contributions. Observe that at any point of differentiability of the payment schedule, the incentive compatibility condition (1) also implies the following relationship between payments and efforts:

$$\dot{\bar{t}}(\theta) = \frac{\dot{\bar{e}}(\theta)}{\theta} \left(\bar{e}(\theta) - e_N(\theta)\right). \tag{12}$$

From Proposition 4, it follows that $\bar{t}(\cdot)$ is strictly increasing on $(\theta^*(\hat{\zeta}), \bar{\theta}]$ and constant on $[\underline{\theta}, \theta^*(\hat{\zeta})]$ if such interval has a non-empty interior. From the fact that the budget-balance constraint (2) is binding at the optimum, it also follows that

$$\bar{t}(\underline{\theta}) < 0 < \bar{t}(\bar{\theta}).$$

Inefficient countries always pay for joining the coalition even though they get the same payoff in and out. They are ready to pay exactly the benefit they receive from the greater effort exerted by those efficient types who produce above the BAU level. More precisely, for large inefficiencies (i.e., when $\hat{\zeta} > \zeta^*$), a country with a type in the interval $[\underline{\theta}, \theta^*(\hat{\zeta})]$ contributes a fixed amount which is *the expected (positive) externality* it enjoys from the agreement:

$$\bar{t}(\theta) = -(1-\alpha)\int_{\theta^*(\hat{\zeta})}^{\bar{\theta}} (\bar{e}(\theta) - e_N(\theta))f(\theta)d\theta < 0.$$

Indeed, when such inefficient country deviates and opts out of the coalition, the most efficient countries with types $\theta \in (\theta^*(\hat{\zeta}), \bar{\theta}]$ react by producing their BAU effort level which is strictly less than that requested by the mechanism. This punishment reduces the overall payoff of the deviating country by an amount which is equal to its contribution:

$$(1-\alpha)\int_{\theta^*(\hat{\zeta})}^{\bar{\theta}} (\bar{e}(\theta) - e_N(\theta))f(\theta)d\theta.$$

The optimal allocation can be implemented by means of a convex nonlinear contribution schedule. To show this, first observe that $\bar{e}(\theta)$ is an increasing function of θ when

Assumption 1 holds. Hence, we may define the inverse mapping $\bar{\theta}(e)$ on the relevant interval and a nonlinear payment schedule that implements the optimal allocation as:

$$T(e) = \bar{t}(\bar{\theta}(e)) = \int_{\underline{\theta}}^{\bar{\theta}(e)} \frac{\bar{e}^2(x)}{2x^2}dx - \alpha e + \frac{e^2}{2\bar{\theta}(e)} - (1-\alpha)E_{\tilde{\theta}}(\bar{e}(\tilde{\theta})).$$

Proposition 5 *$T(e)$ is flat for $e \leq e_N(\theta^*(\hat{\zeta}))$, strictly increasing and convex for $e > e_N(\theta^*(\hat{\zeta}))$.*

Observe that $T'(\bar{e}(\bar{\theta})) = 1-\alpha \geq T'(\bar{e}(\theta))$ for all θ. Indeed, the most efficient countries fully internalize the impact of their effort on global welfare since they receive a Pigovian (marginal) subsidy for doing so. Less efficient types are less rewarded at the margin and do not expand effort as much.

We will use this convexity later to derive simple(r) instruments which are able to efficiently approximate the optimal allocation.

4 Commitment Issues

We now investigate the properties of agreements under various scenarios on the commitment ability of the parties. Indeed, ratifying countries may not always be able to specify threats of retaliation on non-ratifiers. The two commitment scenarios that are considered below correspond to polar fall-back payoffs for a country that chooses not to ratify the mechanism. Those scenarios entail participation constraints in the mechanism design problem which are more or less stringent. That, in turn, affects the efficiency of the mechanism. The analysis unveils how the ability of treaty members to punish non-ratifiers is key to move away from the BAU outcome.

4.1 No Commitment

Suppose first that the mechanism cannot credibly impose any threat on non-ratifiers. Ratifying countries keep on playing the mechanism even after having contemplated a deviation from a country opting out. A non-ratifying country still chooses an effort level $e_N(\theta)$ while ratifiers keep on choosing the effort levels requested by the mechanism. The participation constraint becomes:

$$U(\theta) \geq \frac{\alpha^2}{2}\theta + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta})), \quad \forall \theta \in \Theta. \tag{13}$$

By refusing to abide to the agreement, a deviating country does not affect the aggregate effort but avoids paying any contribution on its own. Of course, this scenario leads to an extreme form of free riding at the extensive margin.[20]

[20]Those strong incentives to free-ride arise because each country is infinitely small in the world as a whole. This is itself a strong assumption that could be relaxed by considering the case of a limited number of countries (or few blocks of countries).

Proposition 6 *When any commitment to inefficient threats is not possible, the only feasible allocation is BAU.*

It is therefore impossible to achieve any positive results beyond BAU when the mechanism is not contingent on the participation of all countries. To improve on BAU, a treaty must stipulate obligations/commitments of the ratifying members which depend on the behavior of all countries. Interestingly, the Kyoto protocol included such contingent restrictions as it required the ratification by countries representing 55% of worldwide emissions to bring the treaty into force.

4.2 Worst Punishments

Let us consider now the opposite polar case where coalition members can collectively punish non-ratifiers. This is of course an extreme and unrealistic assumption that implies commitment to inefficient threats, more precisely zero effort by non-deviating countries to minimize the deviation payoff for such non-ratifier. Even though choosing an effort level $e_N(\theta)$ remains optimal for such country, the *worst punishment* yields a payoff from not joining in which is now given by:

$$U_W(\theta) = \frac{\alpha^2}{2}\theta.$$

Inducing participation requires:

$$U(\theta) \geq U_W(\theta) \quad \forall \theta \in \Theta. \tag{14}$$

Proposition 7 *The first-best allocation can always be implemented when the fall-back option is the Worst-Punishment outcome.*

Because the fall-back option entails zero effort by non-deviating countries, the gains from cooperation increase. It allows to implement the first best even when incentive constraints matter.[21]

5 Limits on Enforcement

Section 3.3 has featured the different shapes that an optimal mechanism may take with a sole focus on incentive constraints as an impediment to efficiency. Environmental treaties may on top also face enforcement problems.

In this respect, the optimal mechanism characterized in Section 3.3 has some surprising features, especially when the participation constraint is binding on a non-empty interval $\Omega^c = [\underline{\theta}, \theta^*(\hat{\zeta})]$ (the case of **strong distortions**). Indeed, types in that

[21] This result is reminiscent of other works in Bayesian environments with a finite number of players (Makowski and Mezzetti 1994 among others).

interval exert their BAU effort whether they join the mechanism or not. This makes the mechanism particularly vulnerable to an enforcement problem if contributions are paid once the countries' efforts are already sunk. Once those indifferent types have already chosen their effort, they could just choose not to contribute and free ride on the most efficient ones. This perverse possibility brought by such timing is indeed particularly relevant in the case of the 1997 Kyoto protocol where the 38 most developed countries (the so-called Annex I) committed themselves to a certain level of emissions before any system of contributions were established.

We model this enforcement problem by viewing the countries' decision to comply with the mechanism as a moral hazard variable: A given country abides to the mechanism if it finds it optimal to obey the course of actions it requests.[22] Otherwise, other countries retaliate.[23] To capture the idea that international mechanisms are only enforced imperfectly, we assume that, whenever a country does not comply, it is only punished with some exogenous probability $\delta < 1$ by compliant countries. At this punishment stage, those countries return to their BAU effort levels while the deviating country itself reduces its own effort down to its BAU level. With probability $1-\delta$, the deviating country is not punished, keeps the effort requested by the mechanism but does not pay. Therefore, a country with type θ abides to the mechanism whenever the following *enforcement constraint* holds:[24]

$$U(\theta) \geq (1-\delta)\left(-\frac{e^2(\theta)}{2\theta} + \alpha e(\theta) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta}))\right) + \delta U_N(\theta). \tag{15}$$

This enforcement constraint (15) can also be written as:

$$t(\theta) \geq \frac{\delta}{1-\delta}(U_N(\theta) - U(\theta)). \tag{16}$$

Contributions cannot be too large without impairing the play of the mechanism. On the other hand, the enforcement constraint certainly holds for the most efficient countries which are subsidized by the mechanism, receive positive transfers and get more than their BAU payoff.

[22]To motivate this approach, observe that, in an international context, any mechanism between sovereign countries may lack of the perfect enforceability technology that is available when private parties contract under the aegis of a Court of Law.

[23]Laffont and Martimort (2002, Chapter 9) present a related model of enforcement in a static principal-agent relationship whereas Levin (2003) and Athey, Bagwell and Sanchirico (2004) study enforcement issues in other specific dynamic contexts.

[24]Although our analysis does not rely on a full-fledged dynamic modeling, this enforcement constraint admits an interpretation in terms of repeated games. Everything happens as if parties were committed to a stationary mechanism that covers an infinite number of periods with a discount factor δ. Types are stationary and drawn once for all. (See Baron and Besanko, 1984, for such information structure.) The mechanism defines a repeated game with per-period payoff $U(\theta)$ and a discount factor δ. Whenever a country does not contribute, non-deviating ones play trigger strategies and BAU follows in the continuation.

Under limited enforcement, the optimization problem becomes:

$$(\mathcal{P}^E): \quad \max_{U(\cdot)\in W(\Theta), e(\cdot)} E_{\tilde{\theta}}(U(\tilde{\theta})) \quad \text{subject to (2), (4) and (15).}$$

In a first pass to assess the new inefficiencies involved, we may first investigate conditions under which the first-best levels of effort are no longer implementable.

Proposition 8 *The first-best allocation cannot be implemented under limited enforcement when*

$$\alpha > \alpha_1(\delta) = \alpha_1 - \frac{2(1-\delta)(E_{\tilde{\theta}}(\tilde{\theta}) - \underline{\theta})}{\delta(2E_{\tilde{\theta}}(\tilde{\theta}) - \underline{\theta})}. \tag{17}$$

Because the enforcement constraint (15) is stronger than (3), it becomes harder to implement the efficient level of effort and $\alpha_1(\delta) \leq \alpha_1$.

From a technical viewpoint, the enforcement constraint (15) is complex for two reasons. First, it is a *mixed constraint* of the kind $g(U(\theta), e(\theta), E_{\tilde{\theta}}(e(\tilde{\theta})), \theta) \geq 0$, involving both the state variable $U(\theta)$, the control $e(\theta)$ and its average value $E_{\tilde{\theta}}(e(\tilde{\theta}))$. Second, $g(\cdot)$ so defined is neither quasi-concave nor does it satisfy standard constraints qualification conditions. Henceforth, standard sufficiency theorems from optimal control cannot be applied to characterize an optimal mechanism.[25] To avoid those technical difficulties without loss in terms of economic insights, we will now replace (15) with the more stringent *state-dependent constraint* :

$$U(\theta) \geq U_N(\theta) + (1-\delta)(1-\alpha)(E_{\tilde{\theta}}(e(\tilde{\theta})) - E_{\tilde{\theta}}(e_N(\tilde{\theta}))) \quad \forall \theta. \tag{18}$$

This condition is more stringent because the right-hand side of (15) is maximized for $e(\theta) = e_N(\theta)$ and thus always less than that of (18). However, next lemma shows that, for *strong distortions* where (15) binds on an interval with non-empty interior, replacing (15) with (18) entail no loss for mechanisms that implement efforts above BAU.

Lemma 2 *Optimal mechanisms with strong distortions where (15) is binding on an interval Ω^c with non-empty interior and such that $e(\theta) \geq e_N(\theta)$ for all θ satisfy (18).*

Equipped with Lemma 2, we define a new mechanism design problem where the "type-dependent" constraint (18) replaces the more complex "mixed-constraint" (15):

$$(\mathcal{P}^E): \quad \max_{U(\cdot)\in W(\Theta), e(\cdot)} E_{\tilde{\theta}}(U(\tilde{\theta})) \quad \text{subject to (2), (4) and (18) .}$$

Solving this problem offers a characterization of regimes with strong distortions.

[25]The fact that $g(\cdot)$ is not quasi-concave in (U, e) precludes the use of sufficiency results of the Mangasarian type (Seierstad and Sydsaeter, 1987, Chapter 6, p. 358). Moreover, as we will show in Lemma 2, whenever $g(U(\theta), e(\theta), E_{\tilde{\theta}}(e(\tilde{\theta})), \theta) = 0$ on an interval with non-empty interior, we have $e(\theta) = e_N(\theta)$ and $\frac{\partial g}{\partial e}(U(\theta), e_N(\theta), E_{\tilde{\theta}}(e(\tilde{\theta})), \theta) = 0$ at such θ which means that constraint qualification fails and Arrow-type sufficiency results (Seierstad and Sydsaeter, 1987, Chapter 6, p. 368) cannot be used either.

Proposition 9 *Assume that (17) holds. Under limited enforcement, an optimal mechanism with strong distortions is such that there exists $\hat{\zeta} > 1$ such that (15) is binding on an interval $\Omega^c = [\underline{\theta}, \theta^*(\hat{\zeta})]$ with $\theta^*(\hat{\zeta}) > \underline{\theta}$ solving:*

$$\frac{1 - F(\theta^*(\hat{\zeta}))}{\theta^*(\hat{\zeta}) f(\theta^*(\hat{\zeta}))} = \frac{1-\alpha}{\alpha} \left(\frac{\hat{\zeta}}{\hat{\zeta} - 1} - 1 + \delta \right). \tag{19}$$

The effort profile is then:

$$\bar{e}(\theta) = \begin{cases} \left(1 - \frac{\hat{\zeta}-1}{\hat{\zeta}}(1-\delta)(1-\alpha)\right) \frac{\theta}{1 + \frac{\hat{\zeta}-1}{\hat{\zeta}} \frac{1-F(\theta)}{\theta f(\theta)}} > e_N(\theta) & \text{if } \theta \in \Omega = (\theta^*(\hat{\zeta}), \bar{\theta}] \\ e_N(\theta) & \text{if } \theta \in \Omega^c = [\underline{\theta}, \theta^*(\hat{\zeta})]. \end{cases} \tag{20}$$

Comparing (20) with (10) shows that reducing the effort level of the most efficient countries towards the BAU level relaxes the enforcement constraint (18).[26] Comparing (19) and (6), we observe also that $\theta^*(\hat{\zeta})$ is greater when Assumption 1 holds. In other words, the area where the enforcement constraint binds is larger than with the weaker participation constraint. Distortions are more pronounced under limited enforcement.

6 Approximate Implementation

The convexity of the nonlinear contribution schedule $T(e)$ found in Proposition 5 suggests that this schedule could be conveniently approximated by a pair of simple linear schemes.[27] To replicate the flat part of $T(e)$ and approximate the optimal mechanism for lower levels of effort, the first option within this menu has countries paying up-front a fixed amount $\underline{T}$ and still exerting their BAU effort. Only the least efficient countries choose that scheme. The second linear option entails both a greater up-front contribution $\overline{T} > \underline{T}$ but also a Pigovian subsidy $1 - \alpha$ per unit of effort so that the first-best effort is exerted by the more efficient types opting for that scheme. This option is meant to capture the properties of the optimal mechanism for the highest levels of effort.[28] Finally, budget balance holds when the fixed contributions from both groups cover the needed subsidies.

Let us denote by θ^* the cut-off type who is just indifferent between those two options. By incentive compatibility and single-crossing, types below θ^* choose their BAU effort while those above choose the efficient effort. This leads us to the following indifference condition for θ^*:

$$\alpha e^{FB}(\theta^*) - \frac{(e^{FB}(\theta^*))^2}{2\theta^*} - \overline{T} + (1-\alpha) \left(\int_{\underline{\theta}}^{\theta^*} e_N(\tilde{\theta}) d\tilde{\theta} + \int_{\theta^*}^{\bar{\theta}} e^{FB}(\tilde{\theta}) d\tilde{\theta} \right)$$

[26]Of course, the values of the multiplier $\hat{\zeta}$ differ in the two scenarios.

[27]This insight is well-known in the procurement/regulation literature (Rogerson 2003, Wilson 1993).

[28]Observe that all countries taking such linear scheme equalize their opportunity costs of effort so that re-trading among them won't be a valuable option.

$$= \alpha e_N(\theta^*) - \frac{e_N^2(\theta^*)}{2\theta^*} - \underline{T} + (1-\alpha)\left(\int_{\underline{\theta}}^{\theta^*} e_N(\tilde{\theta})d\tilde{\theta} + \int_{\theta^*}^{\bar{\theta}} e^{FB}(\tilde{\theta})d\tilde{\theta}\right).$$

Simplifying, we obtain:

$$\overline{T} = \underline{T} + (1-\alpha^2)\frac{\theta^*}{2}. \tag{21}$$

To ensure participation of the least efficient countries, their upfront contribution must just balance the externality gain created by the extra effort of countries with types above θ^*. This extra effort being $e^{FB}(\theta) - e_N(\theta) = (1-\alpha)\theta$, the expected externality on types below θ^* becomes $(1-\alpha)^2 \int_{\theta^*}^{\overline{\theta}} \theta f(\theta)d\theta$. This gives the following expression for $\underline{T}$:

$$\underline{T} = (1-\alpha)^2 \int_{\theta^*}^{\overline{\theta}} \theta f(\theta)d\theta. \tag{22}$$

Finally, the menu must be budget balanced, where the expenses are the subsidies per unit of effort given to the most efficient agents and the resources are the lump-sum contributions paid by both groups, namely:

$$F(\theta^*)\underline{T} + (1 - F(\theta^*))\overline{T} = (1-\alpha)\int_{\theta^*}^{\overline{\theta}} \theta f(\theta)d\theta. \tag{23}$$

Using the expressions of $\overline{T}$ and $\underline{T}$ drawn from (21) and (22) and inserting into (23), θ^* is implicitly defined as a solution to the following equation (for $\alpha < 1$):

$$\mathcal{J}(\theta^*) = \frac{\theta^*}{2}(1 - F(\theta^*))(1+\alpha) - \alpha \int_{\theta^*}^{\overline{\theta}} \theta f(\theta)d\theta = 0. \tag{24}$$

Remark first that $\theta^* = \overline{\theta}$ is a solution and that $\mathcal{J}'(\overline{\theta}) < 0$. Moreover, Assumption 1 implies that $\mathcal{J}(\cdot)$ is quasi-concave and there are thus at most two solutions to (24). More precisely, note that $\mathcal{J}(\underline{\theta}) > 0$ if and only if $\alpha \leq \alpha_1$. Therefore, for $\alpha \leq \alpha_1$, $\theta^* = \underline{\theta}$, the first best is always implemented with a single linear contract of slope $1-\alpha$ and we recover our previous findings. On the contrary, for $\alpha > \alpha_1$, we then have $\theta^* \in (\underline{\theta}, \overline{\theta})$, and the type space is split into two connected subsets taking different contracts.

Simulations. One may now wonder how significant is the welfare loss from using the simple two-item menu above instead of the optimal nonlinear mechanism. As the following numerical simulations show, the loss is surprisingly small and therefore the two-item menu turns to be a good approximation of the optimal mechanism.

Let us characterize the optimal contract and its two-item approximation for a uniform distribution on $\Theta = [1, 2]$. For this particular specification, we find $\alpha_1 = \alpha_2 = 0.5$. Moreover, tedious computations show that $\hat{\alpha} = 0.726$. Following the insights of Proposition 2, we will take $\alpha = 0.65$ and $\alpha = 0.85$ to respectively illustrate the cases of *weak* and *strong distortions*.

• For *weak distortions*, i.e., $\alpha = 0.65$, we know that $\theta^*(\hat{\zeta}) = \underline{\theta} = 1$. Moreover, computations lead to $\hat{\zeta} = 1.397$ so that the optimal effort is everywhere given by

$$\bar{e}(\hat{\zeta}, \theta) = \frac{\theta^2}{0.792\theta + 0.416}.$$

From this, we compute that the aggregate welfare under the optimal mechanism is roughly equal to 0.367. In this example, the first-best welfare would be equal to 0.75. Observe that the second-best outcome is relatively far away from the first best, half of the overall surplus being lost due to informational constraints.

If a two-item menu is instead offered, (24) yields $\theta^* = 1.300$, i.e., the thirty percent least efficient countries pay the lower amount $\underline{T}$. Equations (21) and (22) yield then

$$\underline{T} = 0.190 \text{ and } \bar{T} = 0.565.$$

Finally, the aggregate welfare achieved with such menu is roughly worth 0.328. Comparing with the optimal mechanism, the relative welfare loss from using the simple menu instead of the optimal mechanism is 10.7 percent.This is admittedly small, especially compared to the size of surplus lost from informational constraints even with the optimal mechanism. Of course, that mild loss must be put beside the significantly simpler design of the two-item menu compared with the optimal mechanism.

• For *strong distortions*, i.e., $\alpha = 0.85$, we know that $\theta^*(\hat{\zeta}) > 1$. Computations lead to $\hat{\zeta} = 1.779$ and $\theta^*(\hat{\zeta}) = 1.425$. The optimal effort is everywhere given by

$$\bar{e}(\hat{\zeta}, \theta) = \begin{cases} \frac{\theta^2}{0.557\theta + 0.886} & \text{if } \theta \in (1.425, 2] \\ 0.85\theta & \text{if } \theta \in [1, 1.425]. \end{cases}$$

This corresponds to a value of the aggregate welfare under the optimal mechanism which is now roughly equal to 0.380.

If a two-item is instead offered, (24) yields $\theta^* = 1.700$, i.e., the thirty percent most efficient countries pay the higher contribution $\overline{T}$ and receive the Pigovian subsidies per unit of efforts. Equations (21) and (22) yield then

$$\underline{T} = 0.012 \text{ and } \bar{T} = 0.247.$$

It is worth noticing that the contribution asked from the least efficient countries is rather small in that case.

The aggregate welfare achieved with such menu is approximatively equal to 0.373. Now, the relative welfare loss from using the menu instead of the optimal mechanism is less than 2 percent; a surprisingly small loss indeed.

Even though our simple menu above does not perfectly fit any existing real-world mechanism, it lends itself into a nice and realistic interpretation. Suppose that developing countries face lower marginal opportunity costs of reducing pollution because they just do not produce as much as developed countries. Those countries self-select on a scheme with a subsidy. They exert first-best efforts, get subsidized for that, but contribute to fund this program by giving back large fixed contributions. *A contrario,* the more developed countries face higher opportunity costs and do not expand effort beyond BAU. *Per capita,* those countries contribute less to the global funding of the system but, as our numerical examples illustrate, the fraction of countries that self-select by choosing a fixed payment may be significant.

Finally, our mechanism bears some strong resemblance with another proposal, the so-called *Global Public Good Purchase* pushed forward by Bradford (2008). In Bradford's (complete information) mechanism, countries make a set of voluntary contributions to an International Agency; this agency buys then any reduction below the BAU allowances. In our mechanism, countries choose between only two possible levels of their initial contributions that are pocketed by an agency. Some countries choose a larger contribution but also receive a subsidy for any effort made in reducing pollution. Instead, others do not receive any subsidy and keep on exerting their BAU effort.

7 Final Remarks

In practice, climate-change policies are implemented by means of markets for pollution permits (or quotas). A key feature of such mechanism is to allow further rounds of decentralized trade if some countries (reps. firms within those countries) want to trade quotas beyond their initial allocation. In the framework of our model, one may wonder what could be the impact of allowing resale of "effort" quotas. The answer is immediate. Opening markets for trading effort quotas would just drive all participating countries to equalize their opportunity costs to the prevailing market price. This feature stands in sharp contrast with the strict convexity of the optimal mechanism which implies that those countries which exert more effort than in the BAU scenario do so at different rates. In other words, allowing decentralized trade would undermine the screening properties of the mechanism. *A contrario*, the approximate mechanism sketched in Section 6 is robust to such trades, at least as far as the most efficient countries are concerned. Indeed, those countries all get the same Pigovian subsidy and would not gain from further trading quotas.

The main thrust of our analysis is also robust to the introduction of some redistributive concerns although some effects may be magnified. *Ex ante* efficiency is only one possibility (among a whole continuum) for choosing a normative criterion to assess the performances of climate-change policies. Adopting the definition of *interim efficient allocations* given by Holmström and Myerson (1983), we could as well consider a welfare criterion attributing type-dependent non-negative social weights to each possible type. To understand how the optimal mechanism would be modified with such redistributive concerns, suppose for instance, that ethic considerations lead to give to low-income countries (presumably those with the lowest opportunity costs of exerting depolluting efforts) a slightly greater weight in the objective. Effort for those most efficient types should not be so distorted away from the first best. Those efficient countries end up significantly above their BAU payoffs. For the least efficient types instead, effort distortions are exacerbated and the effort profile may severely drop off as costs increase. In terms of the payoffs profile, while the most efficient countries end up much above the BAU level, more countries might just be also indifferent between joining the agreement or not. Of course, such features are also reflected into the approximate menu that could be used in practice. The incentive option is taken by fewer efficient countries but, for those countries, the lump-sum contributions also diminish.

We deliberately chose to study a very parsimonious model to highlight the trade-off between the various forms of free-riding in the most illuminative way. More detailed modelings of the production processes in each country and of the intertemporal impact of investments would lead to more complex analysis but the very same economic insights are much likely to pertain. As long as the BAU outcome leads to excessively low effort levels compared to the socially optima, a mechanism with two options (the first with incentive properties and the second being only a fixed contribution) would certainly perform pretty well.

Finally, one may wonder whether our model which, for tractability reasons, adopted the short-cut of viewing the world as made of a continuum of countries could say anything on the case of big actors (China, U.S., India...) whose strategic behavior might significantly impact aggregate emissions. One way of addressing this "size" issue would be to introduce atoms with positive masses in the types distribution. Although, we shall leave for further research the complete analysis of such cases, it is worth pointing out a few directions in which our results might be modified. First, the presence of an atom in the middle of the types distribution certainly violates Assumption 1. Bunching in the optimal allocation will arise, with types nearby the atom being all given the same effort target. The nonlinear contribution schedule that implements such allocation will then exhibit a kink at the bunch, while preserving enough convexity. This convexity is enough to again justify an approximate implementation by a menu of two options. The point is that the presence of a non-atomistic player modifies the identity of the marginal type which is indifferent between the flat and the incentive option.

Equipped with the mechanism design methodology developed in this paper, we believe that a number of other important questions could be addressed in future research. A first important extension should consider the design of dynamic mechanisms. In particular, one may want to assess the performance of menus of linear contracts in those dynamic environments. A second extension would be to go more deeply into the analysis of the relationship between local politics and international agreements. The analysis of such two-tier mechanism design problem will be particularly fruitful to understand institutional design behind the climate-change problem.[29] At last and taking a broader perspective, our methodology and the workhorse model we have proposed could certainly be also useful to analyze how sovereign countries deal with other multilateral externalities problems such as fiscal fraud, fight against global terrorism or global health problems.

References

Aldy, J., and R. Stavins (ed.), 2007, *Architectures for Agreement: Addressing Global Climate Change in the Post-Kyoto World*, Cambridge University Press.

Athey, S., K. Bagwell and C. Sanchirico, 2004, "Collusion and Price Rigidity," *Review of Economic Studies*, 71: 317-349.

Bagnoli, M. and T. Bergstrom, 2005, "Log-Concave Probability and its Applications," *Economic Theory*, 26: 445-469.

Baliga, S. and E. Maskin, 2003, "Mechanism Design for the Environment," in J. Vincent and K.-G. Mäler eds., *Handbook of Environmental Economics*, Elsevier.

Baron, D. and D. Besanko, 1984, "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 84: 267-302.

Barrett, S., 1994, "Self-Enforcing International Environmental Agreements," *Oxford Economic Papers*, 46: 878-894.

Barrett, S., 2003, *Environment and Statecraft: The Strategy of Environmental Treaty-Making*, 25: 11-34.

Battaglini, M. and B. Harstad, (2012), "Participation and Duration of Climate Contracts," mimeo Northwestern University and Olso University.

Beccherle, J. and J. Tirole, 2011, *Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements*, *Journal of Public Economics*, 95: 1339-1348.

Bradford, D., 2008, "Improving on Kyoto: Greenhouse Gas Control as the Purchase of a Global Public Good," in R. Guesnerie and H. Tulkens eds. *The Design of Climate Policy*, MIT Press.

[29]Those models could use the framework developed by Laffont and Martimort (2005) for solving transnational public good problems.

Carraro, C., 2005, "Institution Design for Managing Global Commons," in G. Demange and M. Wooders eds., *Group Formation in Economics*, Cambridge University Press.

Carraro, C. and D. Siniscalco, 1993, "Strategies for International Protection of the Environment," *Journal of Public Economics*, 52: 309-328.

Carraro, C. and D. Siniscalco, 1995, "International Coordination of Environmental Policies and Stability of Global Environmental Agreements," in L. Bovenberg and S. Cnossen eds., *Public Economics and the Environment in an Imperfect World*, Kluwer Academics.

Chander, P. and H. Tulkens, 1995, "A Core-Theoretic Solution for the Design of Cooperative Agreements and Transfrontier Pollution," *International Tax and Public Finance*, 2: 279-294.

Chander, P. and H. Tulkens, 1997, "The Core of an Economy with Multilateral Environment Exernalities," *International Journal of Game Theory*, 26: 379-401.

Chander, P. and H. Tulkens, 2008, "Cooperation, Stability, Self-Enforcement in Agreements," in R. Guesnerie and H. Tulkens eds., *The Design of Climate Policy*, MIT Press.

Galbraith, G and R. Vinter, 2004, "Regularity of Optimal Controls for State-Constrained Problems," *Journal of Global Optimization*, 28: 305-317.

Guesnerie, R., 2008, "Design of Post-Kyoto Climate Schemes: Selected Questions in Analytical Perspective," in R. Guesnerie and H. Tulkens eds. *The Design of Climate Policy*, MIT Press.

Harstad, B. 2012a, "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations," *Review of Economic Studies*, forthcoming.

Harstad, B. 2012b, "The Dynamics of Climate Agreements", mimeo.

Helm, C. and F. Wirl, 2011, "International Environmental Agreements: Incentive Contracts with Multilateral Externalities," mimeo University of Vienna.

Helm, D., C. Helpburn and R. Mash, 2005, "Credible Carbon Policy," in D. Helm, ed. *Climate-Change Policy*, Oxford University Press.

Holmström, B., and R. Myerson, 1983, "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, 51: 1799-1819.

Laffont, J.J. and D. Martimort, 2005, "The Design of Transnational Public Good Mechanisms for Developing Countries," *Journal of Public Economics*, 89: 159-196.

Laffont, J.J. and E. Maskin, 1982, "The Theory of Incentives: An Overview," in *Advances in Economic Theory*, ed. W. Hildenbrand. Cambridge University Press.

Levin, J. , 2003, "Relational Incentive Contracts," *American Economic Review*, 93: 837-855.

Makowski, L. and C. Mezzetti, 1995, "Bayesian and Weakly Robust First-Best Mechanisms: Characterizations," *Journal of Economic Theory*, 64: 500-519.

Martimort, D. and H. Moreira, 2010, "Common Agency and Public Good Provision under Asymmetric Information," *Theoretical Economics*, 5: 159-213.

Martimort, D. and L. Stole, 2011, "Public Contracting in Delegated Agency Games," mimeo Paris School of Economics.

Milgrom, P. and I. Segal, 2002, "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70: 583-601.

Neeman, Z., 1999, "Property Rights and Efficiency of Voluntary Bargaining under Asymmetric Information," *Review of Economic Studies*, 66: 679-691.

Rob, R., 1989, "Pollution Claim Settlements under Private Information," *Journal of Economic Theory*, 47: 307-333.

Rogerson, W., 2003, "Simple Menus of Contracts in Cost-Based Procurement and Regulation," *American Economic Review*, 93: 919-926.

Seierstad, A. and K. Sydsaeter, 1987, *Optimal Control Theory with Economic Applications,* North-Holland.

Stern, N., 2006, *The Economics of Climate Change: The Stern Review,* Cambridge University Press.

Thoron, S., 2008, "Heterogeneity in Negotiations of International Agreements," in R. Guesnerie and H. Tulkens eds., *The Design of Climate Policy*, MIT Press.

Tirole, J., 2008, "Some Economics of Global Warming," *Rivista di Politica Economica,* 98: 9-42.

Wilson, R., 1993, *Nonlinear Pricing*, Oxford University Press.

Appendix

Proof of Lemma 1. Define $f(t,e,\theta,E) = t+\alpha e+(1-\alpha)E-\frac{e^2}{2\theta}$. Observe that $f(t,e,E,\theta)$ is differentiable and absolutely continuous in θ since $\theta \geq \underline{\theta} > 0$ for any (t,e,E). Moreover, $|f_\theta(t,e,E,\theta)| = \frac{e^2}{2\theta^2}$ is bounded by some integrable function $\frac{M^2}{2\theta^2}$ when $e \in [0,M]$. From Theorem 2 and Corollary 1 in Milgrom and Segal (2002), it follows immediately that $U(\theta)$ is absolutely continuous and thus almost everywhere differentiable with:

$$U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \frac{e^2(x)}{2x^2}dx. \tag{A1}$$

Condition (4) follows at any point of differentiability.

Incentive compatibility implies for any pair $(\theta, \hat{\theta})$:

$$t(\theta) + \alpha e(\theta) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta})) - \frac{e^2(\hat{\theta})}{2\theta} \geq t(\hat{\theta}) + \alpha e(\hat{\theta}) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta})) - \frac{e^2(\hat{\theta})}{2\theta},$$

Reversing the role of θ and $\hat{\theta}$ and summing both sides of the inequalities so obtained, using the fact that $-\frac{e^2}{2\theta}$ satisfies increasing differences, and simplifying yields immediately $e(\theta) \geq e(\hat{\theta})$ for $\theta \geq \hat{\theta}$. $e(\cdot)$ is non-decreasing and thus a.e. differentiable.

Reciprocally, since $U(\cdot)$ is absolutely continuous and satisfies everywhere (A1), we have:

$$U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \frac{e^2(x)}{2x^2}dx = t(\theta) + \alpha e(\theta) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta})) - \frac{e^2(\theta)}{2\theta}.$$

From this, incentive compatibility immediately follows since:

$$t(\theta) + \alpha e(\theta) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta})) - \frac{e^2(\theta)}{2\theta} - \left(t(\hat{\theta}) + \alpha e(\hat{\theta}) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta})) - \frac{e^2(\hat{\theta})}{2\theta}\right)$$

$$= \int_{\hat{\theta}}^{\theta} \frac{e^2(x) - e^2(\hat{\theta})}{2x^2}dx \geq 0$$

when $e(\cdot)$ is non-decreasing. ■

Proof of Propositions 1 and 7. An important step of the analysis consists in consolidating the incentive compatibility constraint (4) and the feasibility condition (2). In this respect, let define a *critical type* θ^* as:

$$\theta^* = \max \arg \min_{\theta \in \Theta} U(\theta) - U_l(\theta)$$

where $l = N, W$. Of course, such critical type depends on the choice of the mechanism since it affects the profile of implementable rent $U(\theta)$. From continuity of $U(\theta) - U_l(\theta)$ and compactness of Θ, such θ^* necessarily exists for any implementable profile $U(\theta)$.

Note that satisfying the participation constraint (3) at θ^* is enough to have it satisfied for all θ. Hence, a necessary and sufficient condition for (3) to hold is that

$$U(\theta^*) \geq U_l(\theta^*). \tag{A2}$$

Using again (A1) yields

$$U(\theta) = U(\theta^*) + \int_{\theta^*}^{\theta} \frac{e^2(x)}{2x^2} dx. \tag{A3}$$

Integrating by parts on each interval $[\underline{\theta}, \theta^*]$ and $[\theta^*, \bar{\theta}]$, we finally obtain the following expression of the average payoff of countries:

$$E_{\tilde{\theta}}(U(\tilde{\theta})) = U(\theta^*) + E_{\tilde{\theta}} \left(\frac{(1_{\tilde{\theta} \geq \theta^*} - F(\tilde{\theta}))e^2(\tilde{\theta})}{2\tilde{\theta}^2 f(\tilde{\theta})} \right)$$

where $1_{\tilde{\theta} \geq \theta^*} = \begin{cases} 1 & \text{if } \tilde{\theta} \geq \theta^* \\ 0 & \text{otherwise.} \end{cases}$

Finally, the feasibility condition can be rewritten as

$$E_{\tilde{\theta}} \left(e(\tilde{\theta}) - \frac{e^2(\tilde{\theta})}{2\tilde{\theta}} \right) \geq U(\theta^*) + E_{\tilde{\theta}} \left(\frac{(1_{\tilde{\theta} \geq \theta^*} - F(\tilde{\theta}))e^2(\tilde{\theta})}{2\tilde{\theta}^2 f(\tilde{\theta})} \right). \tag{A4}$$

Notice that any rent profile for a mechanism that implements the first-best effort level $e^{FB}(\theta)$ is such that $\underline{\theta}$ is the critical type since $U(\theta) - U_l(\theta)$ (for $l = N, W$) is increasing ($\dot{U}(\theta) - \dot{U}_l(\theta) = \frac{1-\alpha^2}{2} > 0$ when $\alpha < 1$). Hence, a necessary and sufficient condition for the participation constraint (3) to hold everywhere is that it holds at $\underline{\theta}$. That remark being made, the feasibility constraint and the critical type's participation constraint are altogether satisfied when:

$$E_{\tilde{\theta}} \left(e^{FB}(\tilde{\theta}) - \frac{(e^{FB}(\tilde{\theta}))^2}{2\tilde{\theta}} \right) \geq U_l(\underline{\theta}) + E_{\tilde{\theta}} \left(\frac{(1 - F(\tilde{\theta}))(e^{FB}(\tilde{\theta}))^2}{2\tilde{\theta}^2 f(\tilde{\theta})} \right).$$

This amounts to check

$$E_{\tilde{\theta}} \left(e^{FB}(\tilde{\theta}) - \frac{(e^{FB}(\tilde{\theta}))^2}{2\tilde{\theta}} \left(1 + \frac{1 - F(\tilde{\theta})}{\tilde{\theta} f(\tilde{\theta})} \right) \right) = \frac{1}{2} \int_{\underline{\theta}}^{\bar{\theta}} (\theta f(\theta) - 1 + F(\theta)) d\theta \geq U_l(\underline{\theta})$$

$$\Leftrightarrow \begin{cases} \frac{\underline{\theta}}{2} \geq \frac{\alpha^2}{2}\underline{\theta} + (1-\alpha)\alpha E_{\tilde{\theta}}(\tilde{\theta}) & \text{if } l = N \\ \frac{\underline{\theta}}{2} \geq \frac{\alpha^2}{2}\underline{\theta} & \text{if } l = W. \end{cases} \tag{A5}$$

Hence, when $l = N$, we get an impossibility if Assumption 2 holds. Instead, when $l = W$, (A5) holds and one can find budget-balanced transfers that ensure that the first best is implemented. ■

Proofs of Propositions 3 and 4. We first characterize the optimal mechanism when Assumption 2 holds. The proof of Propositions 3 and 4 is a direct consequence of this characterization.

Neglecting the monotonicity condition on $e(\cdot)$ that will be checked ex post; we first rewrite the so relaxed optimization problem under asymmetric information as:

$$(\mathcal{P}^{SB}): \quad \max_{U(\cdot)\in W(\Theta), e(\cdot)} E_{\tilde{\theta}}(U(\tilde{\theta})) \quad \text{subject to (2), (3) and (4)}$$

where $W(\Theta)$ is the set of absolutely continuous arcs on Θ.

$(\mathcal{P}^{SB})$ is a generalized Bolza problem with an isoperimetric constraint (2) and a state-dependent constraint (3). We denote by ζ the non-negative multiplier of the former constraint. This allows us to write the Lagrangian for this problem as:

$$L(\theta, U, e, \zeta) = f(\theta)\left(U + \zeta\left(e - \frac{e^2}{2\theta} - U\right)\right).$$

Let then define the Hamiltonian as

$$H(\theta, U, e, \zeta, q) = L(\theta, U, e, \zeta) + q\frac{e^2}{2\theta^2}.$$

This Hamiltonian is linear in U and strictly concave in e when

$$q \leq \xi\theta f(\theta). \tag{A6}$$

This latter condition is checked below for the optimal profile.

Following Galbraith and Winter (2004), the necessary optimality conditions that are satisfied by a normal extremum $(\bar{U}(\theta), \bar{e}(\theta))$ can be written as follows.

Proposition A.1 ***Necessary conditions*** *(Galbraith and Winter, 2004). There exists an absolutely continuous function $p(\theta)$, a function $q(\theta)$, and a non-negative measure $\mu(d\theta)$ which are all defined on Θ such that:*

$$-\dot{p}(\theta) = \frac{\partial H}{\partial U}(\theta, \bar{U}(\theta), \bar{e}(\theta), \zeta, q(\theta)), \tag{A7}$$

$$\bar{e}(\theta) \in \arg\max_{e\geq 0} H(\theta, \bar{U}(\theta), e, \zeta, q(\theta)), \tag{A8}$$

$$q(\theta) = p(\theta) - \int_{\underline{\theta}}^{\theta^-} \mu(d\theta), \quad \forall \theta \in (\underline{\theta}, \bar{\theta}], \tag{A9}$$

$$supp\{\mu\} \subset \{\theta \; s.t. \; \bar{U}(\theta) = U_N(\theta)\} = \Omega^c, \tag{A10}$$

$$p(\underline{\theta}) = -p(\bar{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \mu(d\theta) = 0. \tag{A11}$$

Sufficient conditions. *Those necessary conditions are also sufficient (Martimort and Stole, 2011, Appendix B).*

Condition (A7) describes how the costate variable $p(\cdot)$ evolves whereas (A8) is the optimality condition for the control. Some explanations for the other conditions are in order. From (A9), the left-side limit of $q(\cdot)$ at any θ is the costate variable deflated by a term related to the measure w.r.t. μ of the open interval $[\underline{\theta}, \theta)$.[30] This costate variable measures the distortions induced by asymmetric information. From (A10), the support of the measure μ is contained in the subset of types for which the participation constraint (3) is binding. Together, with (A8), it implies that distortions due to asymmetric information are less significant on intervals where the participation constraint is binding. Sufficiency is obtained by adapting the same Arrow-type argument as in Martimort and Stole (2011, Appendix B). Conditions (A7) to (A11) are also sufficient for $(\bar{U}(\theta), \bar{e}(\theta))$ to be an optimum.

Let us rewrite some of these optimality conditions. First, observe that (A7) can be transformed as

$$-\dot{p}(\theta) = f(\theta)(1-\zeta). \tag{A12}$$

From (A11), we get

$$p(\bar{\theta}) = \int_{\underline{\theta}}^{\bar{\theta}} \mu(d\theta). \tag{A13}$$

We may rewrite (A12) as

$$p(\theta) = p(\bar{\theta}) + (1-\zeta)(1-F(\theta)). \tag{A14}$$

Second, (A8) yields the first-order condition

$$\zeta f(\theta)\left(1 - \frac{\bar{e}(\theta)}{\theta}\right) = -q(\theta)\frac{\bar{e}(\theta)}{\theta^2}. \tag{A15}$$

In the sequel, we consider two possibilities for the subset of types where the participation constraint (A2) is binding. In **Case 1 (strong distortions)** below, this participation constraint is binding on an interval $\Omega^c = [\underline{\theta}, \theta^*]$ with non-zero measure. **Case 2 (weak distortions)** deals with the case where $\Omega^c = \{\underline{\theta}\}$.

Case 1. $\Omega^c = [\underline{\theta}, \theta^*]$**, with** $\theta^* > \underline{\theta}$**.**

Analysis of the set of types Ω^c where the participation constraint (3) is binding. [31] Several facts immediately follow from the optimality conditions.

[30]Such formulation is made necessary to take into account the fact that μ may be singular at θ.

[31]From the sufficiency conditions in Proposition A.1, finding a vector (p, q, e) that induces such allocation and satisfies the necessary conditions (A7) to (A11) validates this *"guess and try"* approach.

- Since $\mu = 0$ on $\Omega = (\theta^*, \bar{\theta}]$, (A13) implies that

$$p(\bar{\theta}) = \int_{\underline{\theta}}^{\theta^*} \mu(dx). \tag{A16}$$

- Consider now $\Omega = (\theta^*, \bar{\theta}]$ (with non-zero measure) where (A2) is slack, i.e., $\bar{U}(\theta) > U_N(\theta)$. On the interior of such interval, $\mu = 0$ and (A9) implies that

$$q(\theta) = p(\theta) - \int_{\underline{\theta}}^{\theta^*} \mu(dx). \tag{A17}$$

Using (A14), (A16) and (A17) yields

$$q(\theta) = (1 - \zeta)(1 - F(\theta)). \tag{A18}$$

Finally inserting (A18) into (A15) yields the expression optimal effort level $\bar{e}(\theta, \zeta)$ given by (5) (where we make the dependence on ζ explicit for further references).

- Consider now an interval $\Omega^c = [\underline{\theta}, \theta^*]$ with non-zero measure where (A2) is binding, i.e., $\bar{U}(\theta) = U_N(\theta)$. Differentiating with respect to θ in the interior of $\Omega^c = [\underline{\theta}, \theta^*]$ yields

$$\dot{\bar{U}}(\theta) = \dot{U}_N(\theta) \Leftrightarrow \bar{e}(\theta) = e_N(\theta).$$

Therefore, (A15) becomes now:

$$q(\theta) = -\left(\frac{1-\alpha}{\alpha}\right)\zeta\theta f(\theta) \quad \forall \theta \in (\underline{\theta}, \theta^*). \tag{A19}$$

From (A9), (A14) and (A19), we deduce that

$$\int_{\underline{\theta}}^{\theta^-} \mu(d\theta) = p(\bar{\theta}) + (1-\zeta)(1 - F(\theta)) + \left(\frac{1-\alpha}{\alpha}\right)\zeta\theta f(\theta) \quad \forall \theta \in (\underline{\theta}, \theta^*)$$

or, using (A16)

$$-\int_{\theta^-}^{\theta^*} \mu(d\theta) = (1-\zeta)(1 - F(\theta)) + \left(\frac{1-\alpha}{\alpha}\right)\zeta\theta f(\theta) \quad \forall \theta \in (\underline{\theta}, \theta^*). \tag{A20}$$

Let us look for a positive measure μ that is absolutely continuous with respect to the Lebesgue measure on $(\underline{\theta}, \theta^*]$ and so writes as $\mu(d\theta) = g(\theta)d\theta$ for some measurable and non-negative function $g(\cdot)$ on this interval.

Before studying further the properties of $g(\cdot)$, we prove the following Lemma:

Lemma A.1 *Assume that Assumption 1 holds. Take* $k \leq \frac{1}{\underline{\theta} f(\underline{\theta})}$ *and define uniquely* $\theta^* \in [\underline{\theta}, \bar{\theta}]$ *as the solution to*

$$k = \frac{1 - F(\theta^*)}{\theta^* f(\theta^*)} > 0. \tag{A21}$$

Then, we have

$$\frac{d}{d\theta}\left(1-F(\theta)-k\theta f(\theta)\right) \le 0 \quad \forall \theta \in [\underline{\theta}, \theta^*]. \tag{A22}$$

Proof. Observe that Assumption 1 can be rewritten as

$$0 \ge \frac{d}{d\theta}\left(\frac{1-F(\theta)}{\theta f(\theta)}\right) = -\frac{1}{\theta} - \frac{(1-F(\theta))}{\theta^2 f^2(\theta)}\frac{d}{d\theta}\left(\theta f(\theta)\right) \Leftrightarrow -(1-F(\theta))\frac{d}{d\theta}\left(\theta f(\theta)\right) \le \theta f^2(\theta).$$

From this, it follows that

$$\frac{d}{d\theta}\left(1-F(\theta)-k\theta f(\theta)\right) = -f(\theta) - k\frac{d}{d\theta}\left(\theta f(\theta)\right) \le f(\theta)\left(-1 + k\frac{\theta f(\theta)}{1-F(\theta)}\right).$$

Using the definition of k from (A21) and again Assumption 1, we get:

$$k \le \frac{1-F(\theta)}{\theta f(\theta)} \quad \forall \theta \le \theta^*$$

Therefore, we get

$$\frac{d}{d\theta}\left(1-F(\theta)-k\theta f(\theta)\right) = -f(\theta) - k\frac{d}{d\theta}\left(\theta f(\theta)\right) \le 0 \quad \forall \theta \le \theta^*$$

which yields (A22). ■

Consider now $k = \frac{\zeta(1-\alpha)}{(\zeta-1)\alpha}$ and observe that $k \le \frac{1}{\underline{\theta} f(\underline{\theta})}$ when $\zeta > \zeta^*$ where ζ^* is defined in (7).

Differentiating (A20) with respect to θ yields

$$g(\theta) = (1-\zeta)\left(-f(\theta) - k\frac{d}{d\theta}\left(\theta f(\theta)\right)\right) \quad \forall \theta \in (\underline{\theta}, \theta^*). \tag{A23}$$

From Lemma A.1, applied to such k, $g(\cdot)$ is indeed non-negative on $[\underline{\theta}, \theta^*]$ if $\zeta > 1$. More precisely, when $\zeta > 1$, we get:

$$g(\theta) = (1-\zeta)\frac{d}{d\theta}\left(1-F(\theta)-k\theta f(\theta)\right) \ge 0 \quad \forall \theta \in (\underline{\theta}, \theta^*). \tag{A24}$$

By construction, μ has no mass point at θ^*. This implies that $\bar{e}(\theta^*, \zeta) = e_N(\theta^*)$ and $\theta^*(\zeta)$ is thus defined by (6) when interior.

Note also that putting altogether (A16) and (A24) implies that

$$p(\bar{\theta}) = \mu(\{\underline{\theta}\}) + (1-\zeta)\int_{\underline{\theta}}^{\theta^*} \frac{d}{d\theta}\left(1-F(\theta)-k\theta f(\theta)\right) d\theta$$

where $\mu(\{\underline{\theta}\})$ is the mass that the measure μ charges at $\underline{\theta}$. Using (A21), this latter equation can be rewritten as:

$$p(\bar{\theta}) = \mu(\{\underline{\theta}\}) - (1-\zeta) - \frac{1-\alpha}{\alpha}\zeta\underline{\theta}f(\underline{\theta}). \tag{A25}$$

But from (A11) and (A12), we get

$$p(\underline{\theta}) = p(\bar{\theta}) + 1 - \zeta = 0. \tag{A26}$$

Inserting into (A25) yields

$$\mu(\{\underline{\theta}\}) = \frac{1-\alpha}{\alpha}\zeta\underline{\theta}f(\underline{\theta}) > 0 \tag{A27}$$

which shows that μ has a mass point at $\underline{\theta}$.

Concavity of $H(\theta, U, e, \zeta, q)$ in e. Observe that, for $\theta \in \Omega^c$, $q(\theta)$ as defined by (A19) is negative and thus (A6) holds where $q = q(\theta)$. For $\theta \in \Omega$, we deduce from (A18) that $q(\theta) < 0$ and thus (A6) again holds.

Continuity of $\bar{e}(\cdot)$ at θ^*. This continuity immediately follows from the fact that μ has no charge at θ^*. This implies "smooth-pasting" of the rent profile with:

$$U(\theta^*) = U_N(\theta^*) \text{ and } \dot{U}(\theta^*) = \dot{U}_N(\theta^*).$$

Monotonicity of $\bar{e}(\cdot)$. It immediately follows from the fact that $\bar{e}(\cdot)$ is everywhere continuous and, trivially increasing on Ω^c but also on Ω from Assumption 1.

Case 2. $\Omega^c = \{\underline{\theta}\}$**.** Observe that $k = \frac{\zeta(1-\alpha)}{(\zeta-1)\alpha} > \frac{1}{\underline{\theta}f(\underline{\theta})}$ when $\zeta \leq \zeta^*$. In that case, the participation constraint (A2) is binding at $\underline{\theta}$ only. From (A27), the measure μ has a charge at $\underline{\theta}$ only. When $\zeta \geq 1$, we have

$$\mu(\{\underline{\theta}\}) = \frac{1-\alpha}{\alpha}\zeta\underline{\theta}f(\underline{\theta}) \geq (\zeta - 1)\frac{\zeta^*}{\zeta^* - 1} \geq 0. \tag{A28}$$

The optimal effort is still given by (5) on the whole interval $[\underline{\theta}, \bar{\theta}]$.

Proof that $\hat{\zeta} > 1$. Observe that, when binding, (2) can be rewritten as:

$$\int_{\underline{\theta}}^{\theta^*(\zeta)} \left(e_N(\theta) - \frac{e_N^2(\theta)}{2\theta}\right) f(\theta)d\theta + \int_{\theta^*(\zeta)}^{\bar{\theta}} \left(\bar{e}(\theta,\zeta) - \frac{\bar{e}^2(\theta,\zeta)}{2\theta}\right) f(\theta)d\theta$$

$$= \int_{\underline{\theta}}^{\theta^*(\zeta)} U_N(\theta)f(\theta)d\theta + \int_{\theta^*(\zeta)}^{\bar{\theta}} \left(U_N(\theta^*(\zeta)) + \int_{\theta^*(\zeta)}^{\theta} \frac{\bar{e}^2(\xi,\zeta)}{2\xi^2}d\xi\right) f(\theta)d\theta \tag{A29}$$

where we make explicit the dependence of $\bar{e}(\cdot)$ and θ^* on ζ as specified in (5) and (6) to express the left-hand side and where we use (A3) to rewrite the right-hand side.[32]

Let denote respectively by $L(\zeta)$ and $R(\zeta)$ the left-hand and right-hand sides of (A29). The following observations are readily made.

1. $L(\zeta) - R(\zeta)$ *is strictly increasing.* First, observe that

$$\frac{\partial \bar{e}}{\partial \zeta}(\theta, \zeta) = -\frac{\frac{1-F(\theta)}{f(\theta)}}{\left(\zeta + (\zeta - 1)\frac{1-F(\theta)}{\theta f(\theta)}\right)^2} < 0. \tag{A30}$$

   Using the fact that $\bar{e}(\theta, \zeta)$ is continuous at $\theta = \theta^*(\zeta)$, i.e., $\bar{e}(\theta^*(\zeta), \zeta) = e_N(\theta^*(\zeta))$, we have:

$$L'(\zeta) = \int_{\theta^*(\zeta)}^{\bar{\theta}} \frac{\partial \bar{e}}{\partial \zeta}(\theta, \zeta)\left(1 - \frac{\bar{e}(\theta, \zeta)}{\theta}\right) f(\theta) d\theta = (\zeta - 1) \int_{\theta^*(\zeta)}^{\bar{\theta}} \frac{\partial \bar{e}}{\partial \zeta}(\theta, \zeta) \frac{\frac{1-F(\theta)}{\theta f(\theta)}}{\zeta + (\zeta - 1)\frac{1-F(\theta)}{\theta f(\theta)}} f(\theta) d\theta. \tag{A31}$$

   Using the fact that $U_N(\theta, \zeta)$ is continuous at $\theta = \theta^*(\zeta)$, we have

$$R'(\zeta) = \dot{\theta}^*(\zeta) \int_{\theta^*(\zeta)}^{\bar{\theta}} \left(\dot{U}_N(\theta^*(\zeta)) - \frac{\bar{e}^2(\theta^*(\zeta), \zeta)}{2(\theta^*(\zeta))^2}\right) f(\theta) d\theta + \int_{\theta^*(\zeta)}^{\bar{\theta}} \int_{\theta^*(\zeta)}^{\theta} \frac{\partial \bar{e}}{\partial \zeta}(\xi, \zeta) \frac{\bar{e}(\xi, \zeta)}{\xi^2} f(\theta) d\xi d\theta.$$

   Using that $\dot{U}_N(\theta^*(\zeta)) = \frac{e_N^2(\theta^*(\zeta))}{2(\theta^*(\zeta))^2}$, and continuity of $\bar{e}(\cdot, \zeta)$ at $\theta = \theta^*(\zeta)$, i.e., $\bar{e}(\theta^*(\zeta), \zeta) = e_N(\theta^*(\zeta))$, we get

$$R'(\zeta) = \int_{\theta^*(\zeta)}^{\bar{\theta}} \left(\int_{\theta^*(\zeta)}^{\theta} \frac{\partial \bar{e}}{\partial \zeta}(\xi, \zeta) \frac{\bar{e}(\xi, \zeta)}{\xi^2} d\xi\right) f(\theta) d\theta.$$

   Integrating by parts yields

$$R'(\zeta) = \int_{\theta^*(\zeta)}^{\bar{\theta}} (1 - F(\theta)) \frac{\partial \bar{e}}{\partial \zeta}(\theta, \zeta) \frac{\bar{e}(\theta, \zeta)}{\theta^2} d\theta = \int_{\theta^*(\zeta)}^{\bar{\theta}} \frac{\partial \bar{e}}{\partial \zeta}(\theta, \zeta) \frac{\zeta \frac{1-F(\theta)}{\theta f(\theta)}}{\zeta + (\zeta - 1)\frac{1-F(\theta)}{\theta f(\theta)}} f(\theta) d\theta. \tag{A32}$$

   Using (A31) and (A32) we finally get

$$L'(\zeta) - R'(\zeta) = -\int_{\theta^*(\zeta)}^{\bar{\theta}} \frac{\partial \bar{e}}{\partial \zeta}(\theta, \zeta) \frac{\frac{1-F(\theta)}{\theta f(\theta)}}{\zeta + (\zeta - 1)\frac{1-F(\theta)}{\theta f(\theta)}} f(\theta) d\theta > 0.$$

2. Notice that when $\zeta = 1$, $\theta^*(\zeta) = \underline{\theta}$ and $L(1) < R(1)$ indeed amounts to (2).

3. We have

[32]Observe that this formula encompasses both **Case 1** which applies for $\zeta \geq \zeta^*$ and **Case 2** which applies for $\zeta \in [1, \zeta^*]$.

Lemma A.2

$$\lim_{\zeta \to +\infty} L(\zeta) - R(\zeta) > 0. \tag{A33}$$

Proof. Consider the following problem:

$$\mathcal{V}^M = \max_{e(\cdot),\theta^*} \int_{\underline{\theta}}^{\theta^*} \left(e_N(\theta) - \frac{e_N^2(\theta)}{2\theta} \right) f(\theta)d\theta + \int_{\theta^*}^{\bar{\theta}} \left(e(\theta) - \frac{e^2(\theta)}{2\theta} \left(1 + \frac{1 - F(\theta)}{\theta f(\theta)} \right) \right) f(\theta)d\theta$$

$$- \int_{\underline{\theta}}^{\theta^*} U_N(\theta) f(\theta)d\theta - U_N(\theta^*)(1 - F(\theta^*)). \tag{A34}$$

First, observe that $\mathcal{V}^M \geq 0$. Indeed, taking $e(\theta) = e_N(\theta)$ and $\theta^* = \bar{\theta}$ obviously yields 0 for the maximand.

The above maximum is achieved for $(\bar{e}_\infty(\theta), \theta^*_\infty)$ where

$$\bar{e}_\infty(\theta) = \frac{\theta}{1 + \frac{1-F(\theta)}{\theta f(\theta)}} \tag{A35}$$

and

$$\begin{cases} \frac{1-F(\theta^*_\infty)}{\theta^*_\infty f(\theta^*_\infty)} = \frac{1-\alpha}{\alpha} & \text{if } \frac{1-\alpha}{\alpha} < \frac{1}{\underline{\theta} f(\underline{\theta})} \\ \theta^*_\infty = \underline{\theta} & \text{if } \frac{1-\alpha}{\alpha} \geq \frac{1}{\underline{\theta} f(\underline{\theta})}. \end{cases} \tag{A36}$$

Condition 1 ensures that $\theta^*_\infty \in (\underline{\theta}, \bar{\theta})$ always exists whenever (9) holds. That $\mathcal{V}^M > 0$ immediately follows from observing that $\mathcal{V}^M$ is not achieved for $e_N(\theta)$ and $\theta^* = \bar{\theta}$. Finally, this strict inequality amounts to (A33). ■

From Items [1.], [2.] and [3.] above, there exists $\hat{\zeta} > 1$ such that

$$L(\hat{\zeta}) = R(\hat{\zeta}).$$

Integrating by parts and manipulating finally yields (8). ■

Proof of Proposition 2. Because Assumption 3 holds, we have $1 > \frac{1-\alpha}{\alpha}\underline{\theta} f(\underline{\theta}) > 0$ and thus $\zeta^*(\alpha) > 1$ for any $\alpha \geq \alpha_1 \geq \alpha_2$. A first implication is that, for $\zeta \leq \zeta^*(\alpha)$, we get $\theta^*(\zeta) = \underline{\theta}$. Because $L(\cdot) - R(\cdot)$ is strictly increasing as shown above, we have $\hat{\zeta} \leq \zeta^*(\alpha)$ if and only if

$$L(\zeta^*(\alpha)) \geq R(\zeta^*(\alpha)) \Leftrightarrow J(\alpha) \geq U_N(\underline{\theta}, \alpha) \tag{A37}$$

where

$$J(\alpha) = \int_{\underline{\theta}}^{\bar{\theta}} \left(\bar{e}(\theta, \zeta^*(\alpha)) - \frac{\bar{e}^2(\theta, \zeta^*(\alpha))}{2\theta} \left(1 + \frac{1 - F(\theta)}{\theta f(\theta)} \right) \right) f(\theta)d\theta$$

and where, for future reference, we make explicit the dependence of $U_N(\cdot)$ on α.

We compute:

$$J'(\alpha) = \frac{\partial \zeta^*}{\partial \alpha}(\alpha) \int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial \bar{e}}{\partial \zeta}(\theta, \zeta^*(\alpha)) \left(1 - \frac{\bar{e}(\theta, \zeta^*(\alpha))}{\theta} \left(1 + \frac{1 - F(\theta)}{\theta f(\theta)}\right)\right) f(\theta) d\theta$$

$$= \int_{\underline{\theta}}^{\bar{\theta}} \frac{\underline{\theta} f(\underline{\theta})(1 - F(\theta))^2}{\theta f(\theta)} \frac{((1-\alpha)\underline{\theta} f(\underline{\theta}) - \alpha)}{\left(\alpha + (1-\alpha)\frac{(1-F(\theta))\underline{\theta} f(\underline{\theta})}{\theta f(\theta)}\right)^3} d\theta.$$

We have $J'(\alpha) \leq 0$ for any $\alpha \geq \alpha_2$ (with equality only at $\alpha = \alpha_2$).

Moreover, for $\alpha = 1$, we have $\zeta^*(1) = 1$ and $\bar{e}(\theta, \zeta^*(1)) = e^{FB}(\theta)$. Therefore, we get:

$$J(1) = \int_{\underline{\theta}}^{\bar{\theta}} \left(e^{FB}(\theta) - \frac{(e^{FB}(\theta))^2}{2\theta} \left(1 + \frac{1 - F(\theta)}{\theta f(\theta)}\right)\right) f(\theta) d\theta = \frac{\underline{\theta}}{2} = U_N(\underline{\theta}, 1). \tag{A38}$$

We also find:

$$J'(1) = -\underline{\theta} f(\underline{\theta}) \int_{\underline{\theta}}^{\bar{\theta}} \frac{(1 - F(\theta))^2}{\theta f(\theta)} d\theta.$$

From Assumption 1, we immediately derive the inequality

$$\frac{(1 - F(\theta))^2}{\theta f(\theta)} \leq \frac{1 - F(\theta)}{\underline{\theta} f(\underline{\theta})}$$

with an equality only at $\theta = \underline{\theta}$. Therefore, we get:

$$-J'(1) < \int_{\underline{\theta}}^{\bar{\theta}} (1 - F(\theta)) d\theta = E_\theta(\theta) - \underline{\theta} = -U'_N(\underline{\theta}, \alpha)|_{\alpha=1}. \tag{A39}$$

It follows from $J(\cdot)$ and $U_N(\underline{\theta}, \cdot)$ continuity, that there exists $\alpha_3 < 1$ such that

$$J(\alpha) < U_N(\underline{\theta}, \alpha) \quad \forall \alpha \in (\alpha_3, 1). \tag{A40}$$

Moreover, Assumption 3 implies that $\zeta^*(\alpha_1) > 1$. Therefore, we get $\bar{e}_\infty(\theta) \leq \bar{e}(\theta, \zeta^*(\alpha_1)) \leq \bar{e}(\theta, 1) = e^{FB}(\theta)$ (with an equality only at $\bar{\theta}$). Since $\bar{e}_\infty(\theta)$ is a pointwise maximizer of the concave function $e - \frac{e^2}{2\theta}\left(1 + \frac{1-F(\theta)}{\theta f(\theta)}\right)$, we have:

$$J(\alpha_1) > \int_{\underline{\theta}}^{\bar{\theta}} \left(e^{FB}(\theta) - \frac{(e^{FB}(\theta))^2}{2\theta} \left(1 + \frac{1 - F(\theta)}{\theta f(\theta)}\right)\right) f(\theta) d\theta = \frac{\underline{\theta}}{2} = J(1) = U_N(\underline{\theta}, \alpha_1) \tag{A41}$$

where the last equality follows from observing that $U_N(\underline{\theta}, \alpha_1) = U_N(\underline{\theta}, 1)$ and that (A38) amounts to $J(1) = U_N(\underline{\theta}, \alpha)$ for $\alpha = \alpha_1$. We deduce from this and the fact that $J(\cdot)$ and $U_N(\underline{\theta}, \cdot)$ are continuous that necessarily $\alpha_3 \in (\alpha_1, 1)$.

From (A37) and (A40), we also get:

$$\hat{\zeta} > \zeta^*(\alpha) \quad \forall \alpha \in (\alpha_3, 1).$$

From (A37) and (A41), we deduce that there exists $\alpha_4 \in (\alpha_1, \alpha_3]$ such that

$$J(\alpha) \geq U_N(\underline{\theta}, \alpha) \quad \forall \alpha \in [\alpha_1, \alpha_4]. \tag{A42}$$

Finally, we get

$$\hat{\zeta} \leq \zeta^*(\alpha) \quad \forall \alpha \in [\alpha_1, \alpha_4].$$

We now prove that $\alpha_3 = \alpha_4$. Let denote by $\hat{\alpha}$ this common value. Observe that:

$$\frac{d}{d\alpha}\left(\frac{J'(\alpha)}{(1-\alpha)\underline{\theta}f(\underline{\theta})-\alpha}\right) = -3\int_{\underline{\theta}}^{\bar{\theta}} \frac{\underline{\theta}f(\underline{\theta})(1-F(\theta))^2}{\theta f(\theta)} \frac{\left(1-\frac{(1-F(\theta))\underline{\theta}f(\underline{\theta})}{\theta f(\theta)}\right)}{\left(\alpha+(1-\alpha)\frac{(1-F(\theta))\underline{\theta}f(\underline{\theta})}{\theta f(\theta)}\right)^4} d\theta < 0$$

where this inequality follows from the fact that the numerator in the integrand is non-negative when Assumption 1 holds. Similarly, we compute:

$$\frac{d}{d\alpha}\left(\frac{U_N'(\underline{\theta}, \alpha)}{(1-\alpha)\underline{\theta}f(\underline{\theta})-\alpha}\right) = \frac{\underline{\theta}f(\underline{\theta})(\underline{\theta}-E_{\tilde{\theta}}(\tilde{\theta}))+E_{\tilde{\theta}}(\tilde{\theta})}{((1-\alpha)\underline{\theta}f(\underline{\theta})-\alpha)^2} > 0$$

where the last inequality follows from the fact that Assumptions 2 and 3 altogether imply

$$\underline{\theta}f(\underline{\theta})(\underline{\theta}-E_{\tilde{\theta}}(\tilde{\theta}))+E_{\tilde{\theta}}(\tilde{\theta}) \geq E_{\tilde{\theta}}(\tilde{\theta}) - \frac{\underline{\theta}}{2} > 0.$$

Define now $\varpi(\alpha) = \frac{J'(\alpha)-U_N'(\underline{\theta},\alpha)}{(1-\alpha)\underline{\theta}f(\underline{\theta})-\alpha}$. This continuous function is decreasing over $(\alpha_1, 1)$ with $\varpi(\alpha_1) > 0 > \varpi(1)$ where the first of these inequalities follows from $J'(\alpha_1) < 0 < U_N'(\underline{\theta}, \alpha_1)$ and the second from (A39). Because $(1-\alpha)\underline{\theta}f(\underline{\theta}) - \alpha < 0$ for $\alpha \geq \alpha_1 > \alpha_2$, we deduce that $J'(\alpha) - U_N'(\underline{\theta}, \alpha)$ is non-positive on $[\alpha_1, \tilde{\alpha}]$ and non-negative on $[\tilde{\alpha}, 1]$ for some $\tilde{\alpha} \in (\alpha_1, 1)$. From (A38) and (A41), it follows that $J(\alpha) - U_N(\underline{\theta}, \alpha)$ is decreasing and then increasing on $[\alpha_1, 1]$ with a unique $\hat{\alpha}$ on the decreasing part such that:

$$J(\hat{\alpha}) = U_N(\underline{\theta}, \hat{\alpha}).$$

■

Proof of Proposition 5. From (12), we immediately get:

$$T'(\bar{e}(\theta)) = \frac{\bar{e}(\theta)}{\theta} - \alpha = \begin{cases} \frac{1}{1+\frac{\hat{\zeta}-1}{\hat{\zeta}}\frac{1-F(\theta)}{\theta f(\theta)}} - \alpha & \text{if } \bar{e}(\theta) > e_N(\theta) \Leftrightarrow \theta > \theta^*(\hat{\zeta}) \\ 0 & \text{if } \bar{e}(\theta) = e_N(\theta) \Leftrightarrow \theta \leq \theta^*(\hat{\zeta}) \end{cases}$$

where the first equality follows from (5). Note that $T'(e)$ is continuous at $\bar{e}(\theta^*(\hat{\zeta}))$ (such that $\bar{e}(\theta^*(\hat{\zeta}) = e_N(\theta^*(\hat{\zeta}))$ if it is interior. Differentiating once more, we get:

$$\dot{\bar{e}}(\theta)T''(\bar{e}(\theta)) = \begin{cases} -\frac{\frac{\hat{\zeta}-1}{\hat{\zeta}}\frac{d}{d\theta}\left(\frac{1-F(\theta)}{\theta f(\theta)}\right)}{\left(1+\frac{\hat{\zeta}-1}{\hat{\zeta}}\frac{1-F(\theta)}{\theta f(\theta)}\right)^2} > 0 & \text{if } \bar{e}(\theta) > e_N(\theta) \\ 0 & \text{if } \bar{e}(\theta) = e_N(\theta). \end{cases}$$

Hence, $T(e)$ is convex and strictly so if and only if $e > e_N(\theta^*(\hat{\zeta}))$. It is flat when $e \leq e_N(\theta^*(\hat{\zeta}))$. ■

Proof of Proposition 6. Observe that the budget balance condition (2) altogether with the participation constraints (13) yield the following simpler inequality:

$$\int_{\underline{\theta}}^{\bar{\theta}} \left(\alpha e(\theta) - \frac{e^2(\theta)}{2\theta}\right) f(\theta)d\theta \geq \frac{\alpha^2}{2}\int_{\underline{\theta}}^{\bar{\theta}} \theta f(\theta)d\theta. \tag{A43}$$

The pointwise maximum of the left-hand side is $e_N(\theta) = \alpha\theta$ and then the left- and right-hand sides of (A43) are both equal. Therefore, the optimal mechanism robust to any individual deviation consists in proposing the BNE outcome which is, by definition, also incentive compatible. ■

Proof of Proposition 8. The first best $e^{FB}(\theta) = \theta$ is implementable when (15) holds for all θ, i.e., when there exists a profile $U^{FB}(\theta)$ such that $\dot{U}^{FB}(\theta) = \frac{(e^{FB}(\theta))^2}{2\theta^2} = \frac{1}{2}$ and:

$$U^{FB}(\theta) \geq V^{FB}(\theta) = (1-\delta)\left(-\frac{(e^{FB}(\theta))^2}{2\theta} + \alpha e^{FB}(\theta) + (1-\alpha)E_{\tilde{\theta}}(e^{FB}(\tilde{\theta}))\right) + \delta U_N(\theta) \quad \forall \theta. \tag{A44}$$

Observe that, with the first-best profile of effort, $\dot{U}^{FB}(\theta) = \frac{1}{2} > \dot{V}^{FB}(\theta) = \delta\frac{\alpha^2}{2} - (1-\delta)(1-\alpha)$. Hence, (A44) holds for all θ if it holds at $\underline{\theta}$.

Mimicking the analysis in the Proof of Proposition 1, the first-best effort level is thus implementable when:

$$\int_{\underline{\theta}}^{\bar{\theta}} \left(e^{FB}(\theta) - \frac{(e^{FB}(\theta))^2}{2\theta}\left(1 + \frac{1-F(\theta)}{\theta f(\theta)}\right)\right) f(\theta)d\theta \geq V^{FB}(\underline{\theta}).$$

Simplifying yields the condition:

$$\frac{\underline{\theta}}{2} \geq \delta U_N(\underline{\theta}) + (1-\delta)\left(-\left(\frac{1}{2} - \alpha\right)\underline{\theta} + (1-\alpha)E_{\tilde{\theta}}(\tilde{\theta})\right). \tag{A45}$$

Simplifying further, (A45) does not hold when (17) holds. ■

Proof of Lemma 2. Let Ω^c be an interval with non-empty interior where (15) binds. On such interval, we have

$$U(\theta) = (1-\delta)\left(-\frac{e^2(\theta)}{2\theta} + \alpha e(\theta) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta}))\right) + \delta U_N(\theta). \tag{A46}$$

Differentiating w.r.t θ and taking into account (1), the following condition holds a.e.:

$$\delta\left(\frac{e^2(\theta)}{2\theta^2} - \frac{e_N^2(\theta)}{2\theta^2}\right) = (1-\delta)\left(\alpha - \frac{e(\theta)}{\theta}\right)\dot{e}(\theta) \tag{A47}$$

which admits the solution $e(\theta) = e_N(\theta)$. This solution is the unique one that is non-decreasing as requested by the monotonicity conditions for incentive compatibility.[33]

Whenever (15) is binding on an interval Ω^c for the profile $(U(\theta), e(\theta))$, the simpler constraint (18) is also binding. Indeed, inserting the condition $e(\theta) = e_N(\theta)$ which holds on Ω^c into (A46), we get that (18) holds as an equality.

Observe that (18) always implies (15). Indeed, using the definition of $e_N(\theta)$, (18) implies, that (15) holds since:

$$U(\theta) \geq (1-\delta)\left(-\frac{e^2(\theta)}{2\theta} + \alpha e(\theta) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta}))\right) + \delta U_N(\theta).$$

Denote

$$V(\theta) = (1-\delta)\left(-\frac{e^2(\theta)}{2\theta} + \alpha e(\theta) + (1-\alpha)E_{\tilde{\theta}}(e(\tilde{\theta}))\right) + \delta U_N(\theta).$$

Observe also that $e(\theta) \geq e_N(\theta)$ and $e(\cdot)$ monotonically increasing imply:

$$\dot{U}(\theta) = \frac{e^2(\theta)}{2\theta^2} \geq \dot{U}_N(\theta) = \frac{\alpha^2}{2} \geq \dot{V}(\theta) = (1-\delta)\left(\dot{e}(\theta)\left(\alpha - \frac{e(\theta)}{\theta}\right) + \frac{e^2(\theta)}{2\theta^2}\right) + \delta\frac{\alpha^2}{2}. \quad \text{(A48)}$$

Denote by θ_0 the highest bound of Ω^c. By definition and continuity,

$$U(\theta_0) = U_N(\theta_0) = V(\theta_0).$$

(A48) then implies that the property

$$U(\theta) \geq U_N(\theta) \geq V(\theta)$$

holds for all $\theta \geq \theta_0$ which yields the result. ■

Proof of Proposition 9. First, observe that, we may rewrite (18) as

$$U(\theta) \geq U_N(\theta) + (1-\delta)(1-\alpha)\gamma \quad \forall \theta \tag{A49}$$

where

$$\gamma = E_{\tilde{\theta}}(e(\tilde{\theta}) - e_N(\tilde{\theta})) \tag{A50}$$

Neglecting as usual the monotonicity condition that $e(\cdot)$ is non-decreasing that will be checked ex post; we define a mechanism design problem as:

$$(\mathcal{P}^E_\gamma): \quad \max_{U(\cdot)\in W(\Theta), e(\cdot)} E_{\tilde{\theta}}(U(\tilde{\theta})) \quad \text{subject to (2), (4), (A49) and (A50)}.$$

$(\mathcal{P}^E_\gamma)$ is again a generalized Bolza problem with two isoperimetric constraints (2) and (A50) and a state-dependent constraint (A49). Our first step is to solve for such problem. The solution then defines a value function $V^E(\gamma)$. In a second step, optimizing in γ yields then the optimal value $\hat{\gamma}$.

[33]The other solution to the differential equation (A47) is indeed such that $\dot{e}(\theta) = -\frac{\delta}{2(1-\delta)}\left(\frac{e(\theta)}{\theta} + \alpha\right) < 0$ which violates the monotonicity condition.

Denoting by ζ the non-negative multiplier of (2) and by κ the multiplier of (A50), we write the Lagrangian for $(\mathcal{P}_\gamma^E)$ as:

$$L_\gamma(\theta, U, e, \zeta, \kappa) = f(\theta)\left(U + \zeta\left(e - \frac{e^2}{2\theta} - U\right)\right) + \kappa(\gamma - f(\theta)(e - e_N(\theta))).$$

Let then define the Hamiltonian as

$$H_\gamma(\theta, U, e, \zeta, \kappa, q) = L_\gamma(\theta, U, e, \zeta, \kappa) + q\frac{e^2}{2\theta^2}.$$

This Hamiltonian is linear in U and strictly concave in e when again (A6) holds. This latter condition is again checked below for the optimal profile.

Necessary and sufficient conditions. We proceed as in the previous appendices to write the conditions that a normal extremum $(\bar{U}(\theta,\gamma), \bar{e}(\theta,\gamma))$ must satisfy. Optimality implies that there exists an absolutely continuous function $p(\theta)$, a function $q(\theta)$, and a non-negative measure $\mu(d\theta)$ which are all defined on Θ such that:

$$-\dot{p}(\theta) = \frac{\partial H_\gamma}{\partial U}(\theta, \bar{U}(\theta,\gamma), \bar{e}(\theta,\gamma), \zeta, \kappa, q(\theta)), \tag{A51}$$

$$\bar{e}(\theta,\gamma) \in \arg\max_{e \geq 0} H_\gamma(\theta, \bar{U}(\theta,\gamma), e, \zeta, \kappa, q(\theta)), \tag{A52}$$

$$q(\theta) = p(\theta) - \int_{\underline{\theta}}^{\theta^-} \mu(d\theta), \quad \forall\theta \in (\underline{\theta}, \bar{\theta}], \tag{A53}$$

$$supp\{\mu\} \subset \{\theta \text{ s.t. } \bar{U}(\theta,\gamma) = U_N(\theta) + (1-\delta)(1-\alpha)\gamma\} = \Omega_\gamma^c, \tag{A54}$$

$$p(\underline{\theta}) = -p(\bar{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} \mu(d\theta) = 0. \tag{A55}$$

Let us rewrite some of these optimality conditions. First, observe that (A51) can be transformed again into (A12) and then (A14). Using (A55), we again get (A13).

Second, (A52) yields the first-order condition

$$f(\theta)\left(\zeta - \kappa - \zeta\frac{\bar{e}(\theta,\gamma)}{\theta}\right) = -q(\theta)\frac{\bar{e}(\theta,\gamma)}{\theta^2}. \tag{A56}$$

As before, we distinguish between two scenarios for the subset of types Ω_γ^c where the enforcement constraint (A49) is binding.

Case 1. Strong distortions. $\Omega_\gamma^c = [\underline{\theta}, \theta(\gamma,\zeta,\kappa)]$ **with** $\underline{\theta} < \theta(\gamma,\zeta,\kappa)$**.** Several facts immediately follow.

- Equation (A13) implies again that $p(\bar{\theta})$ solves (A16).

- Consider now the interval $\Omega_\gamma = (\theta(\gamma,\zeta,\kappa),\bar{\theta}]$ where (A49) is slack, i.e., $\bar{U}(\theta,\gamma) > U_N(\theta) + (1-\delta)(1-\alpha)\gamma$. On the interior of such interval, $\mu = 0$ and (A53) implies that again $q(\theta)$ is given by (A17).

  Using (A14), (A16) and (A17) yields again that $q(\theta)$ solves (A18) on Ω_γ. Finally inserting (A18) into (A56) yields the following expression of the optimal effort level $\bar{e}(\theta,\gamma,\zeta,\kappa)$ (where we make the dependence on ζ and κ explicit for further references):

$$\bar{e}(\theta,\gamma,\zeta,\kappa) = \left(1-\frac{\kappa}{\zeta}\right)\frac{\theta}{1+\frac{\zeta-1}{\zeta}\frac{1-F(\theta)}{\theta f(\theta)}}. \tag{A57}$$

  Define $\theta(\gamma,\zeta,\kappa)$ such that $\bar{e}(\theta(\gamma,\zeta,\kappa),\gamma,\zeta,\kappa) = e_N(\theta(\gamma,\zeta,\kappa))$, i.e.,

$$\frac{1-F(\theta(\gamma,\zeta,\kappa))}{\theta(\gamma,\zeta,\kappa)f(\theta(\gamma,\zeta,\kappa))} = \frac{\zeta(1-\alpha)}{(\zeta-1)\alpha} - \frac{\kappa}{(\zeta-1)\alpha}. \tag{A58}$$

  Assume for the time being that the right-hand side of (A58) is non-negative (this will be the case for the optimal value $\hat{\gamma}$ found below) and set $\theta(\gamma,\zeta,\kappa) = \underline{\theta}$ whenever this right-hand side is greater than $\frac{1}{\underline{\theta} f(\underline{\theta})}$.

- Consider now the interval $\Omega_\gamma^c = [\underline{\theta},\theta(\gamma,\zeta,\kappa)]$ with non-zero measure where (A49) is binding, i.e., $\bar{U}(\theta,\gamma) = U_N(\theta) + (1-\delta)(1-\alpha)\gamma$. Differentiating with respect to θ in the interior of Ω^c yields

$$\dot{\bar{U}}(\theta,\gamma) = \dot{U}_N(\theta) \Leftrightarrow \bar{e}_\gamma(\theta) = e_N(\theta).$$

  Therefore, (A56) becomes now:

$$q(\theta) = -\theta f(\theta)\left(\frac{1-\alpha}{\alpha}\zeta - \frac{\kappa}{\alpha}\right) \quad \forall\theta \in (\underline{\theta},\theta(\gamma,\zeta,\kappa)). \tag{A59}$$

  From (A53), (A14), (A16) and (A59) we deduce that

$$-\int_{\theta^-}^{\theta(\gamma,\zeta,\kappa)} \mu(d\theta) = (1-\zeta)(1-F(\theta)) + \theta f(\theta)\left(\frac{1-\alpha}{\alpha}\zeta - \frac{\kappa}{\alpha}\right) \quad \forall\theta \in (\underline{\theta},\theta(\gamma,\zeta,\kappa)). \tag{A60}$$

  Let us look for a positive measure μ that is absolutely continuous with respect to the Lebesgue measure on $(\underline{\theta},\theta^*]$ and so writes as $\mu(d\theta) = g(\theta)d\theta$ for some measurable and non-negative function $g(\cdot)$ on this interval.

  Define $k' = \frac{\zeta(1-\alpha)}{(\zeta-1)\alpha} - \frac{\kappa}{(\zeta-1)\alpha}$ (and consider the case where $k' \geq 0$ from our assumption made after (A58)). Differentiating (A20) with respect to θ yields

$$g(\theta) = (1-\zeta)\left(-f(\theta) - k'\frac{d}{d\theta}(\theta f(\theta))\right) \quad \forall\theta \in (\underline{\theta},\theta(\gamma,\zeta,\kappa)). \tag{A61}$$

  From Lemma A.1 applied to such k', $g(\cdot)$ is indeed non-negative on $[\underline{\theta},\theta(\gamma,\zeta,\kappa)]$ if $\zeta > 1$. Note that by construction, μ has no mass point at $\theta(\gamma,\zeta,\kappa)$. This implies that $\bar{e}(\cdot)$ is continuous at $\theta(\gamma,\zeta,\kappa)$.

Concavity of $H(\theta, U, e, \zeta, q)$ in e. Observe that, for $\theta \in \Omega^c$, $q(\theta)$ as defined by (A59) is negative and thus (A6) holds where $q = q(\theta)$. For $\theta \in \Omega$, we deduce from (A18) that $q(\theta) < 0$. and thus (A6) again holds.

Monotonicity of $\bar{e}(\cdot)$. It immediately follows from the fact that $\bar{e}(\cdot)$ is everywhere continuous and, trivially increasing on Ω^c but also on Ω from Assumption 1.

Computing κ and ζ. Observe that the rent profile (making the dependence in (ζ, κ) explicit) is defined as

$$\bar{U}(\theta, \gamma, \zeta, \kappa) = \begin{cases} U_N(\theta(\gamma, \zeta, \kappa)) + (1-\delta)(1-\alpha)\gamma + \int_{\theta(\gamma,\zeta,\kappa)}^{\theta} \frac{\bar{e}^2(x)}{2x^2} dx & \text{if } \theta \geq \theta(\gamma, \zeta, \kappa) \\ U_N(\theta) + (1-\delta)(1-\alpha)\gamma & \text{if } \theta \leq \theta(\gamma, \zeta, \kappa). \end{cases}$$

Therefore, we may rewrite (2) as

$$\int_{\underline{\theta}}^{\theta(\gamma,\zeta,\kappa)} \left(e_N(\theta) - \frac{e_N^2(\theta)}{2\theta} \right) f(\theta) d\theta + \int_{\theta(\gamma,\zeta,\kappa)}^{\bar{\theta}} \left(\bar{e}(\theta, \gamma, \zeta, \kappa) - \frac{\bar{e}^2(\theta, \gamma, \zeta, \kappa)}{2\theta} \right) f(\theta) d\theta$$

$$= \int_{\underline{\theta}}^{\theta(\gamma,\zeta,\kappa)} U_N(\theta) f(\theta) d\theta + \int_{\underline{\theta}}^{\theta(\gamma,\zeta,\kappa)} \left(U_N(\theta(\gamma, \zeta, \kappa)) + \int_{\theta(\gamma,\zeta,\kappa)}^{\theta} \frac{\bar{e}^2(x, \gamma, \zeta, \kappa)}{2x^2} dx \right) f(\theta) d\theta$$
$$+ (1-\delta)(1-\alpha)\gamma.$$

Or, integrating by parts,

$$\int_{\underline{\theta}}^{\theta(\gamma,\zeta,\kappa)} \left(e_N(\theta) - \frac{e_N^2(\theta)}{2\theta} \right) f(\theta) d\theta + \int_{\theta(\gamma,\zeta,\kappa)}^{\bar{\theta}} \left(\bar{e}(\theta, \gamma, \zeta, \kappa) - \frac{\bar{e}^2(\theta, \gamma, \zeta, \kappa)}{2\theta} \left(1 + \frac{1-F(\theta)}{\theta f(\theta)} \right) \right) f(\theta) d\theta$$

$$= \int_{\underline{\theta}}^{\theta(\gamma,\zeta,\kappa)} U_N(\theta) f(\theta) d\theta + U_N(\theta(\gamma, \zeta, \kappa))(1 - F(\theta(\gamma, \zeta, \kappa))) + (1-\delta)(1-\alpha)\gamma. \quad \text{(A62)}$$

The multipliers κ and ζ are thus solutions to the system defined by (A62) and

$$\gamma = \int_{\theta(\gamma,\zeta,\kappa)}^{\bar{\theta}} (\bar{e}(\theta, \gamma, \zeta, \kappa) - e_N(\theta)) f(\theta) d\theta. \quad \text{(A63)}$$

Case 2. Weak distortions. $\Omega_\gamma^c = \{\underline{\theta}\}$. Observe that $k' \geq \frac{1}{\underline{\theta} f(\underline{\theta})}$ when $\zeta \leq \zeta^*(\gamma, \zeta, \kappa)$ where

$$\frac{1}{\underline{\theta} f(\underline{\theta})} = \frac{\zeta^*(\gamma, \zeta, \kappa)(1-\alpha)}{(\zeta^*(\gamma, \zeta, \kappa) - 1)\alpha} - \frac{\kappa}{(\zeta^*(\gamma, \zeta, \kappa) - 1)\alpha}. \quad \text{(A64)}$$

The enforcement constraint (A49) is then binding at $\underline{\theta}$ only and the measure μ has a charge at $\underline{\theta}$ only. The optimal effort is still given by (A57) but on the whole interval $[\underline{\theta}, \bar{\theta}]$.

Optimal value $\hat{\gamma}$. To compute the optimal value of γ, observe that raising γ by $d\gamma$ raises the whole profile of rents by $(1-\delta)(1-\alpha)d\gamma$ which has a cost $(\zeta - 1)(1-\delta)(1-\alpha)d\gamma$

while at the same time, the benefit of such marginal increase is by definition $\kappa d\gamma$. At the optimum, $\hat{\gamma}$ is found so that:

$$\kappa = (\zeta - 1)(1-\delta)(1-\alpha) \tag{A65}$$

Optimal values $\hat{\zeta}$ and $\hat{\kappa}$. The value $\hat{\zeta}$ is obtained when (A65) is inserted into the system (A62)-(A63). From this value, we then get $\hat{\kappa} = (\hat{\zeta}-1)(1-\delta)(1-\alpha)$. Inserting (A65) into (A57) and (A58) respectively then yields the expression of the optimal effort $\bar{e}(\theta, \hat{\zeta}) = \bar{e}(\theta, \hat{\gamma}, \hat{\zeta}, \hat{\kappa})$ given by (20) and the expression of the optimal cut-off $\theta^*(\hat{\zeta}) = \theta(\gamma, \hat{\zeta}, \hat{\kappa})$ given by (19).

Define then ζ^* such that

$$\frac{1-\alpha}{\alpha}\left(\frac{\zeta^*}{\zeta^*-1} - 1 + \delta\right) = \frac{1}{\underline{\theta} f(\underline{\theta})}.$$

Observe that, for $\hat{\zeta} \leq \zeta^*$, we have $\theta^*(\hat{\zeta}) = \underline{\theta}$ and **Case 2 (weak distortions)** arises. For $\hat{\zeta} > \zeta^*$, we have $\theta^*(\hat{\zeta}) > \underline{\theta}$ and **Case 1 (strong distortions)** arises.

With those notations at hands, $\hat{\zeta}$ solves the following equation in ζ:

$$\int_{\underline{\theta}}^{\theta^*(\zeta)} \left(e_N(\theta) - \frac{e_N^2(\theta)}{2\theta}\right) f(\theta) d\theta$$
$$+ \int_{\theta^*(\zeta)}^{\bar{\theta}} \left(\bar{e}(\theta,\zeta) - \frac{(\bar{e}(\theta,\zeta))^2}{2\theta}\left(1 + \frac{1-F(\theta)}{\theta f(\theta)}\right)\right) f(\theta) d\theta + \int_{\theta^*(\zeta)}^{\bar{\theta}} (1-\delta)(1-\alpha)(e_N(\theta) - \bar{e}(\theta,\zeta)) f(\theta) d\theta$$
$$= \int_{\underline{\theta}}^{\theta^*(\zeta)} U_N(\theta) f(\theta) d\theta + U_N(\theta^*(\zeta))(1 - F(\theta^*(\zeta))). \tag{A66}$$

Mimicking steps in the Proof of Propositions 3 and 4, let again denote respectively by $L(\zeta)$ and $R(\zeta)$ the left-hand and right-hand sides of (A66).

Proof that $\hat{\zeta} > 1$. When $\zeta = 1$, we have $\theta^*(\zeta) = \underline{\theta}$, $\bar{e}(\theta, \zeta) = e^{FB}(\theta)$ and $L(1) < R(1)$ indeed amounts to (17). Proceeding as in the Proof of Propositions 3 and 4, we show that $L(\zeta) - R(\zeta)$ is strictly increasing, and proceeding as in Lemma A.2, we show that $\lim_{\zeta \to +\infty} L(\zeta) - R(\zeta) > 0$. Hence, (A66) admits a unique solution $\hat{\zeta} > 1$. ■