DEFORESTATION AND REDD CONTRACTS

Lecture notes - Bård Harstad

University of Oslo
April 2016

1 Introduction

This note presents a new model of conservation and derives the contract preferred by a third party that benefits from conservation. We also show how the contract both influences, and should be influenced by, the countries' political regimes and state capacities.

Payments for environmental services (PES) are important in many situations, and the resource in our model could be fossil fuels or land use quite generally, but our analysis is motivated in particular by deforestation in the tropics and the emergence of contracts on reducing emissions from deforestation and forest degradation (REDD).

Deforestation in the tropics is an immensely important problem. The cumulative effect of deforestation amounts to about one quarter of anthropogenic greenhouse gas emissions, which generate global warming. The annual contribution from deforestation to CO_2 emissions is around ten percent, and the percentage is even higher for other greenhouse gases. Sadly, tropical forest loss has been *increasing* at an average rate of 2101 km^2 each year since 2000. In addition to the effect on global warming, deforestation leads to huge losses in biodiversity. The negative externalities of deforestation amount to \$2-4.5 trillion a year.

However, estimates suggest that deforestation could be halved at a cost of \$21-35 billion per year, or reduced by 20-30 percent at a price at \$10/tCO_2. Third parties are therefore interested in conservation. With the help of donor countries (in particular, Norway, Germany, and Japan), the World Bank and the United Nations are already offering financial incentives to reduce deforestation in a number of countries. Conservation contracts are favored by economists who view them as the natural Coasian solution, and they are also likely to be an important part of future climate change policies and treaties.

It is therefore essential to understand how conservation contracts should be designed. However, there is little theory that can guide real-world contract designers. A useful theory must take several facts into account. First, the causes of deforestation differ across regions: while local governments sell logging concessions in some countries, other countries fight illegal logging for timber or the burning of the forest for agriculture. Second, markets for timber and agricultural products are integrated and conservation in one region can lead to increased deforestation elsewhere: for conservation programs in the U.S. west, the leakage rate (i.e., the increased deforestation elsewhere per unit conserved in the west) was 43 percent at the regional level, 58 percent at the national level, and 84 percent at the continental level; for the 1987-2006 conservation program in Vietnam, the leakage rate was 23 percent, mostly due to increased logging in neighboring Cambodia and Laos. The presence of leakage is not surprising: conservation reduces the supply of forest products, and the price thus increases; a higher price increases deforestation and the cost of protection elsewhere.

As a third fact, the political regime seems to play an important, but puzzling, role. Decentralization of forest management has reduced deforestation in some regions, like the Himalayas. The reverse effect has been documented in other places, like Indonesia.

Despite all these differences, contracts tend to be similar across countries, and targeted mainly at the national governments. Norway, for example, recently declined to contract with the region Madre de Dios in Peru and it stated that it would only contract at the national level.

These facts and claims raise a number of important questions. How can we explain the inconsistent effect of the political regime on conservation? What is the optimal conservation contract, and how does it depend on state capacities or the driver of deforestation? Is it wise to contract with central governments only, or can local contracts be more effective? Can the existence of conservation contracts actually influence regime change, and when would that be beneficial and increase conservation?

The purpose of this note (Sections 2-3) is to provide a tractable model that can address the questions above. In the model, each country or district may benefit from extracting its resource, but the price of the harvest is reduced by the aggregate supply. To protect the remaining part of the resource, the monitoring effort must ensure that the expected penalty is at least as large as the harvest price motivating illegal logging. Thus, a district may want to limit the amount that is protected, and let some of it be harvested and offered to the market, since this reduces the price and thus the monitoring cost on the part that is to be protected.

The model can explain the inconsistent evidence regarding the effect of the political regime. Suppose districts are "strong" in that extraction is sales-driven and motivated mainly by the profit that can be earned by the districts. In this case, a district benefits if the neighbors conserve since that reduces their supply and the harvest price (and profit) increases. The positive (pecuniary) externality from conservation would be internalized by a central government, so centralizing authority will increase conservation. Alternatively, suppose logging is illegal or districts are "weak" in that they are unable to capture much of the profit, and they find it expensive to protect the resource. In that case, a district loses when neighbors conserve, since this increases the price and the pressure on the resource, and thus also the monitoring cost when the resource is protected. This negative externality implies that when authority is centralized, conservation declines. Consistent with our theory, countries in which decentralization reduced deforestation has been referred to as weak: in Nepal, "the Forest Department was poorly staffed and thus unable to implement and enforce the national policies, and deforestation increased in the 1960s and 1970s". In Indonesia, where decentralization increased deforestation, the state is stronger: "Deforestation in Indonesia is largely driven by the expansion of profitable and legally sanctioned oil palm and timber plantations and logging operations", according to some scholars.

A second purpose of this note (Section 4) is to derive the optimal conservation contract. If there is a single district, a simple contract (similar to a Pigou subsidy) implements the first best, regardless of the other parameters in the model. This finding, in isolation, supports today's use of contracts that are linear in the amount of avoided deforestation. With multiple districts, however, one district finds it optimal to extract more when the neighbors conserve or sign conservation contracts: the resulting higher price makes it profitable to extract if the district is strong, and expensive to protect if the district is weak. This leakage makes a contract less effective, and the optimal contract is weaker. Furthermore, contracting with one district generates externalities on the others' outside option. The donor cannot exploit this externality and the equilibrium contracts are too weak, leading to too much extraction, when districts are strong and the externality positive. When the districts are weak and the externality negative, however, there is too much conservation in equilibrium, since the donor takes advantage of the negative externality on the other districts for each conservation contract that is offered.

2 A Theory of Conservation

This section presents a model of conservation and resource extraction in which there are many districts and a common market for the harvest. The framework is general in that the resource can be any kind of resource (for example, oil or land), the harvest can be timber or agricultural products, and the districts can be countries or villages. To fix ideas, however, we refer to the resource as forest.

To motivate the framework we start by sequentially presenting two alternative models of conservation before we combine them. In both cases there is a regional market with $n \geq 1$ players, or districts, and x_i is the extraction level in district $i \in N = \{1, ..., n\}$. The x_i's are decided on simultaneously and the aggregate harvest, $x = \sum_{i \in N} x_i$, is sold on the common market. The larger is x, the smaller is the price. In the simplest possible setting, a linear demand curve can be derived from quadratic utility functions:

$$p = \overline{p} - ax, \tag{1}$$

where $\overline{p}$ and a are positive constants and p is the equilibrium price.

A sales-driven model. If districts are motivated by the profit generated by the sales, district i's payoff may be represented by $bpx_i - v_i x_i$, where b is the benefit of profit, p is the price for the harvest, and v_i is district i's marginal opportunity value when losing the forest. For example, v_i may represent the environmental benefits which the forest provides to i or the tax or lost transfer which i experiences from more extraction. In Section 4, we will let $v_i \equiv v + t_i$, where t_i is a tax or a transfer.

A protection-driven model. While the sales-driven model is standard, our protection-driven model is new. We now consider a setting in which districts do not extract to sell, but where they try to prevent illegal extraction. If protection is difficult, one must take into account that an illegal logger earns the price p by extracting a unit of the forest. This profit must be compared to the expected penalty, θ, which one faces when logging on that unit of the forest. The enforcement is preventive if and only if the expected penalty is larger than the benefit:

$$\theta \geq p. \tag{2}$$

We let districts set their expected penalties in advance in order to discourage extraction. In principle, the expected penalty can be increased by a larger fine or penalty, but there is a limit to how much the fine can be increased in economies with limited liabilities. To raise the expected penalty further, one must increase the monitoring probability, and this is costly.[1] We let $c > 0$ denote the cost of increasing monitoring enough to increase the expected penalty by one unit. Thus, if (2) holds, it will bind: there is no reason to monitor so much that (2) holds with strict inequality. Further, if (2) does not hold, then $\theta = 0$: if monitoring is not preventing logging, there is no reason to monitor at all. This implies that for each unit of the

[1] If π is the probability of being caught, while w is the largest possible penalty (for example, the wealth of an illegal logger), then monitoring requires $\pi \geq p/w$.

forest, either the district protects the unit and ensures that (2) binds, or the district does not protect at all, and that unit of the forest will be cut.

District i has a large forest or resource stock X_i, and it is allowed to monitor each unit with a different intensity. Since the optimal monitoring intensity for each unit ensures that the expected penalty is either p or 0, it follows that a part of the forest will be protected and conserved, perhaps as a national park, while the remaining part will not be sufficiently protected and thus will eventually be cut. Since x_i denotes the extraction level in district i, such that $X_i - x_i$ is the size of the forest that is conserved, then i's payoff is $-cp(X_i - x_i) - v_i x_i$, since $\theta = p$ for the part $(X_i - x_i)$ that is conserved.[2] The model thus suggests that conservation policies will be "place-based" (for example, restricted to geographically limited but protected national parks), as seems to be the case in Indonesia.

The combined model. More generally, district i may benefit by the part of the resource that is extracted and sold, x_i, at the same time that it finds it expensive to protect the remaining part, $X_i - x_i$. When the arguments above are combined, the utility of district i becomes:

$$u_i = bpx_i - cp(X_i - x_i) - v_i x_i. \tag{3}$$

Below we formally define districts as "strong" if they benefit a lot from the sale (b is large) while finding enforcement inexpensive (c is small). We will define districts as "weak" if, instead, b is small while c is large. This terminology is consistent with the literature on state capacity, discussed in the Introduction.

It will be convenient to assume that the aggregate resource stock is large enough to serve the entire market:

$$\overline{p} - aX < 0, \text{ where } X \equiv \sum_{i \in N} X_i.$$

Remark 1: Alternative interpretations and generalizations. There are several alternative interpretations of the combined model such as it is summarized in (3). First, even if all extraction is illegal, a district may have some concern for the welfare of the loggers, in particular if they are poor and/or citizens of the district. Parameter b may then represent this concern. Alternatively, parameter b may reflect the probability that the government in a district captures the profit from the illegal loggers, even in the areas where the forest is not

[2] To be precise, let S_i be i's forest stock of size X_i, and let θ_s be the expected penalty when logging unit $s \in S_i$. If the forest units are divisible then i's payoff is

$$u_i = -c \int_{S_i} \theta_s ds - v_i \int_{S_i} 1_s ds,$$

where $1_s = 1$ if $\theta_s < p$ but $1_s = 0$ if $\theta_s \geq p$. Since there will be a corner solution for the optimal θ_s, u_i can be written as $-cp(X_i - x_i) - v_i x_i$.

protected.

The model is simple and can easily be generalized in a number of ways. For example, one can allow the districts (or the donor, introduced in Section 4) to take into account some of the consumer surplus: this will merely make the analysis messier without altering the conclusions qualitatively. Since the sales are exported, in reality, it is reasonable that districts will *not* take consumer surplus into account.

Furthermore, note that we have linked the districts by assuming that the harvest is sold at a common downstream market, but we could equally well assume that districts hire labor or need inputs from a common upstream market. To see this, suppose that the price of the harvest is fixed at $\widehat{p}$, and consider the wage cost of the labor needed to extract. If the labor supply curve is linear in total supply, and loggers are mobile across districts, then we may write the wage as $\widehat{w} + ax$, where $\widehat{w}$ is a constant and $a > 0$ is the slope of the labor supply curve. Defining $\overline{p} \equiv \widehat{p} + \widehat{w}$, we can write this model as (1)-(3). It is thus equivalent to the model described above.

Remark 2: Nonpecuniary externalities. Without changing the analysis, we can easily allow for cross-externalities such that district i loses $\widetilde{v}_{-i}$ when the other districts extract. To see that our model already captures this case, suppose that i's true payoff is:

$$\widetilde{u}_i = bpx_i - cp\left(\widetilde{X}_i - x_i\right) - \widetilde{v}_i x_i - \widetilde{v}_{-i} \sum_{j \in N \setminus i} x_j.$$

We can then write the payoff as (3) if we simply define $u_i \equiv \widetilde{u}_i + \overline{p}\widetilde{v}_{-i}/a$ while $v_i \equiv \widetilde{v}_i - \widetilde{v}_{-i}$ and $X_i \equiv \widetilde{X}_i - \widetilde{v}_{-i}/ca$. Thus, a larger cross-externality can be captured by considering a reduction in v_i and X_i in the model described above.

3 Conservation and Political Regimes

Based on the combined model above, in which resource extraction can be sales-driven or protection-driven, this section discusses the equilibrium amount of extraction and conservation. In particular, we will focus on how equilibrium extraction depends on whether districts are weak or strong and the number of districts; discuss when one district benefits or loses if other districts conserve more; and investigate the effect of political centralization.

3.1 The Equilibrium

Let $X = \sum_{i \in N} X_i$ be the total size of the resource, while $\overline{v} = \sum_{i \in N} v_i/n$ is the average v_i. Given (1), it is straightforward to derive a district's downward-sloping best-response curve, coming from the first-order condition when deciding on x_i:

$$x_i = \frac{\overline{p} - ax_{-i}}{2a} + \frac{caX_i - v_i}{2a\,(b+c)}.$$

When we solve for the equilibrium x_i's, we immediately arrive at the first proposition.

Proposition 1. *In equilibrium, extraction is given by:*

$$x_i = \frac{(b+c)\,\overline{p} + acnX_i - ac\sum_{j\in N\setminus i} X_j - nv_i + \sum_{j\in N\setminus i} v_j}{a\,(b+c)\,(n+1)} \Rightarrow \tag{4}$$

$$x = \frac{n}{n+1}\frac{\overline{p}}{a} + \frac{acX - n\overline{v}}{a\,(b+c)\,(n+1)} \Rightarrow \tag{5}$$

$$p = \frac{\overline{p}}{n+1} - \frac{acX - n\overline{v}}{(b+c)\,(n+1)}. \tag{6}$$

We will consider only interior solutions such that the right-hand side of (4) is assumed to be positive but less than X_i for every i. If the right-hand side were instead negative (or larger than X_i), the equilibrium would be $x_i = 0$ (or $x_i = X_i$).

Quite intuitively, extraction is smaller if the districts' opportunity values are high; and p is then also high. However, aggregate extraction x is larger if demand is large ($\overline{p}/a$ large) or the protection cost c is large. If the benefit of sales, b, is large, then extraction increases in the typical sales-driven model (where c is large), but extraction is instead decreasing in b if the protection cost is large: the reason is that when protection is expensive, extraction is so large that the equilibrium price is low. If the weight on profit increases, equilibrium extraction will be reduced to raise the price. Thus, if districts get stronger in that b increases, they extract more if and only if they are also strong in that the protection cost is small.

$$\frac{\partial x}{\partial b} > 0 \text{ if and only if } c < \frac{n\overline{v}}{aX}.$$

Note from (4) that a district i extracts more if its own resource stock is large, since a larger x reduces p and thus the protection cost for the (large) remaining amount. However, if the other districts are large or have small opportunity costs, then these other districts will extract a lot and this reduces the price. When p is small, it is both less profitable to sell, and less expensive for i to protect its resource. For both reasons, district i conserves more when X_j is large or v_j is small, for $j \neq i$.

3.2 Pecuniary Externalities

Having solved for the equilibrium, we can be precise about when districts benefit from a high price. When we take the partial derivative of (3) with respect to the endogenous p, and substitute with (4), we get:

$$\frac{\partial u_i}{\partial p} = \frac{e_i}{a(n+1)} \text{ where } e_i \equiv b\overline{p} - c(aX - \overline{p}) - nv_i + \sum_{j \in N \setminus i} v_j. \tag{7}$$

With this, it is natural to define our labels in the following way.

Definition. *Districts are* strong *and extraction is* sales-driven *if districts benefit from a high price (i.e., $e_i > 0$). Districts are* weak *and extraction is* protection-driven *if districts benefit from a low price (i.e., $e_i < 0$).*

With this definition, districts are strong or, equivalently, extraction is sales-driven if $e_i > 0$, which holds not only when the benefit from profit (b) is large, but also when the market size ($\overline{p}/a$) is large compared to the total resource stock, and when protecting the resource has small costs (c) or low value (v_i). Note that we always have $e_i > 0$ in the standard Cournot model (where $c = 0$) when $x_i > 0$. In contrast, we say that districts are weak and extraction is protection-driven when districts benefit from a low price, since costly monitoring must increase accordingly. This requires that $e_i < 0$, which always holds in the model of illegal extraction (when $b = 0$ and $v_i = v_j$).

Since the price is endogenous and increases when the neighbors conserve, e_i can also be referred to as the intra-district (pecuniary) *externality* from conservation.

Proposition 2. *(i) District i benefits when another district conserves if and only if $e_i > 0$:*

$$\frac{\partial u_i}{\partial(-x_j)} = \frac{e_i}{n+1}.$$

(ii) At the equilibrium conservation levels, we also have:

$$u_i = \frac{1}{a(b+c)}\left[\left(\frac{e_i}{n+1}\right)^2 - acv_i X_i\right] \Rightarrow \tag{8}$$

$$sign\frac{\partial u_i}{\partial \overline{p}} = sign\frac{\partial u_i}{\partial v_j} = -sign\ \frac{\partial u_i}{\partial X_j} = sign\ e_i.. \tag{9}$$

Equation (9) shows that the sign of e_i is important when evaluating several changes. If the market size $\overline{p}$ increases, the price is higher; a high price is beneficial in a sales-driven model where $e_i > 0$, but not when districts are weak and find protection costly. If a district $j \neq i$

values conservation more, or if j's resource stock is smaller, then j is expected to extract less. District i's utility will then increase if and only if $e_i > 0$.

Remark 3: Nonpecuniary externalities. As mentioned in Remark 2, nonpecuniary externalities can be accounted for by simply redefining some parameters. If i loses $\widetilde{v}_{-i}$ when other districts extract, we still have $\partial u_i/\partial p = e_i/a\,(n+1)$ and the nonpecuniary externality is simply added to e_i, which can be written as:[3]

$$e_i = b\overline{p} - c\left(a\widetilde{X} - \overline{p}\right) - n\widetilde{v}_i + \sum_{j \in N \setminus i} \widetilde{v}_j + n\widetilde{v}_{-i}. \tag{10}$$

3.3 The Political Regime

The sign of e_i is also important for district i's *strategy.* If $e_i > 0$, district i prefers a high price, and thus i has an incentive to keep the price high by strategically conserving more. If $e_i < 0$, district i has an incentive to extract more to keep the price and thus the pressure low.

These strategic incentives are particularly important for a large district which influences the price more by a given change in x_i/X_i. It thus follows that while large districts conserve a *larger* fraction of their resource in a sales-driven model (in order to keep p high), they conserves a *smaller* fraction when extraction is protection-driven (in order to reduce p and thus the pressure on the resource). This can be seen by inserting (7) into (4) to get:

$$\frac{x_i}{X_i} = \frac{ac}{a\,(b+c)} + \frac{e_i/X_i}{a\,(b+c)\,(n+1)}.$$

Corollary 1. *A larger district i conserves a larger fraction of its resource if and only if $e_i > 0$:*

$$\frac{\partial x_i/X_i}{\partial X_i} = \frac{-e_i/X_i^2}{a\,(b+c)\,(n+1)}.$$

The effect of the number of districts, n, is equally ambiguous and interesting. In a sales-driven model, it is well known from Cournot games that if the number of sellers increases, then so does the aggregate quantity supplied, while the price declines. We should thus expect $\partial x/\partial n > 0$ in a sales-driven model. With protection-driven extraction, however, districts conserve *less* when they take into account the fact that the pressure on the resource weakens as a consequence. It is for this reason that large districts conserve less. By inserting (7) into

[3] Just as in Remark 2, if $\widetilde{X}$ is the actual forest size, $\widetilde{v}_i$ is i's actual cost of losing its forest while $\widetilde{v}_{-i}$ is i's direct loss when j cuts, then our analysis continues to hold if we define $v_i \equiv \widetilde{v}_i - \widetilde{v}_{-i}$ and $X_i \equiv \widetilde{X}_i - \widetilde{v}_{-i}/ca$. By substituting these parameters into (7), we get (10).

(5), we can see that $\partial x/\partial n < 0$ if and only if $\overline{e} < 0$:

$$x = \frac{cX}{b+c} + \frac{n\overline{e}}{a(b+c)(n+1)}, \text{ where } \overline{e} \equiv \frac{1}{n}\sum_{i\in N} e_i = (b+c)\overline{p} - acX - \overline{v}. \tag{11}$$

The number of districts is therefore important. If a set of districts centralizes authority, we will assume that the forest stocks are pooled and the extraction rates are set to maximize the sum of the merging districts' payoffs. Thus, the aggregate resource X remains unchanged, while the number of relevant governments n declines. To isolate this effect, we assume b and c stay unchanged after centralization.

Corollary 2. *Fix X and $\overline{v}$. Centralization (implying a smaller n) leads to more conservation if districts are strong ($\overline{e} > 0$) but less if districts are weak ($\overline{e} < 0$).*

The corollary holds whether it is only a couple of districts that centralize authority to a common central authority, or all the n districts that centralize power to a single government. If authority is centralized to a single central government, C, then $n = 1$ and (11) becomes:

$$x_C = \frac{cX}{b+c} + \frac{\overline{e}}{2a(b+c)} = \frac{\overline{p}(b+c) + acX - \overline{v}}{2a(b+c)}. \tag{12}$$

4 Conservation Contracts

The previous section derived equilibrium conservation as a function of the parameters in the model. In this section, we further assume that every district has a utility function that is linear and additive in the transfer $\tau_i \in \Re$. We have already suggested that district i's opportunity cost of extraction, v_i, may in part come from lost subsidies or a higher tax on extraction:

$$v_i = v + t_i, \tag{13}$$

where $t_i \in \Re$ can represent an extraction tax, so that the transfer to i would be $\tau_i = -t_i x_i$. Since we now let v be common for the districts, the externality (when $t_i = 0$) will be:

$$e \equiv (b+c)\overline{p} - acX - v.$$

Given (13), (4) shows that x_i is a function of $\mathbf{t} = (t_1, .., t_n)$:

$$x_i(\mathbf{t}) = \frac{e + ac(n+1)X_i - t_i n + \sum_{j\neq i} t_j}{a(b+c)(n+1)}. \tag{14}$$

Thus, u_i can be written as a function of $\mathbf{x}(\mathbf{t}) = (x_1(\mathbf{t}), ..., x_n(\mathbf{t}))$. From equations (1)-(3):

$$u_i^0(\mathbf{x}(\mathbf{t})) = bpx_i(\mathbf{t}) - cp(X_i - x_i(\mathbf{t})) - vx_i(\mathbf{t}),$$

where superscript 0 just indicates that the cost of t_i is not taken into account in the definition of u_i^0. With $\tau_i = -t_i x_i$, i's actual payoff is just as in (3):

$$u_i^0(\mathbf{x}(\mathbf{t})) + \tau_i = bpx_i(\mathbf{t}) - cp(X_i - x_i(\mathbf{t})) - (v + t_i)x_i(\mathbf{t}).$$

In this section we study contracts between the districts and a principal or a "donor" D. We assume that D benefits from conservation and that $u_D = -dx$, where $d > 0$ measures the damage D faces from the districts' extraction. Also D has a quasi-linear function for the payoff $u_D + \tau_D$, where τ_D is the transfer to D. By budget balance, $\tau_D = -\sum_{i \in N} \tau_i$. In the following, we will derive (1) the first-best (Pareto-optimal) allocation as well as the equilibrium contract between (2) D and a central government C and (3) D and $m \leq n$ districts.

4.1 The First Best

Since we have assumed transferable utilities and $n+1$ players, any Pareto optimal allocation $\mathbf{x} = (x_1, ..., x_n)$ must maximize $u_D(\mathbf{x}) + \sum_{i \in N} u_i^0(\mathbf{x})$. Pareto optimality cannot pin down the transfers or even the allocation of x_i's when x is given and v is the same for every district, but the Pareto-optimal x is unique.

Proposition 4. *(i) The first-best extraction level is given by:*

$$x_{FB} = \frac{cX}{b+c} - \frac{d-e}{2a(b+c)}. \tag{15}$$

(ii) The first-best x_{FB} is implemented by the decentralized equilibrium if and only if:

$$\frac{\sum_{i \in N} t_i}{n} = t_{FB} \equiv \left(\frac{n+1}{2n}\right) d + \left(\frac{n-1}{2n}\right) e. \tag{16}$$

Part (i) shows that the expression for x_{FB} equals the expression for x_C if simply $\overline{v}$ in (12) is replaced by $v + d$. Part (ii) of the proposition follows from combining (5), (13), and (15). It states that the first-best tax or (subsidy) rate t_{FB} is a weighted average of the two externalities e and d. To understand this, note that even when $d = 0$, $t_i > 0$ is optimal if and only if other districts benefit when i conserves more. This would be the case when districts are strong and extraction sales-driven. When districts are weak and extraction protection-driven, then $t_i < 0$

would be optimal instead.

When decision-making power is centralized to a single authority, then $n = 1$ and the Pigou tax is standard.

Corollary 4. *Under centralization, the first-best is implemented simply by $t_C = d$. Facing $t_C = d$, C will induce its districts to select x_{FB} by, for example $\bar{t} = t_{FB}$.*

The second part of the corollary is just pointing out that since C maximizes the sum of the districts' payoffs, it will tax extraction according to (16) if just d is replaced by t_C. We thus have a formula for how C can implement its desired policy for any given t_C:

$$\bar{t}\,(t_C) \equiv \left(\frac{n+1}{2n}\right) t_C + \left(\frac{n-1}{2n}\right) e.$$

4.2 Contracts under Centralization

While Proposition 3 describes the first best, we now derive the equilibrium contract if D can make a take-it-or-leave-it offer. We assume the extraction level is contractible so that the transfer from D can be a function of $\mathbf{x}$. Since there is a deterministic relationship between x and the contract, it suffices to consider linear conservation contracts of the type observed in reality (see the remark at the end of this section). If D contracts with C, this means:

$$\tau_C = \max\left\{0, \left(\overline{x}_C - x\right) t_C\right\},$$

where $\overline{x}_C$ is a baseline deforestation level. The contract, which consists of the pair $(t_C, \overline{x}_C)$, implies that C receives t_C dollars for every unit by which the actual extraction x is reduced relative to the baseline level $\overline{x}_C$. If $x \geq \overline{x}_C$, no payment is taking place. When $x < \overline{x}_C$, the transfer can be written as $\tau_C = t_C \overline{x}_C - t_C x$, with the last term being equivalent to a tax t_C, while the first term is equivalent to a lump-sum payment.

If $x < \overline{x}_C$, then C's payoff is $u_C^0\left(x_C\left(t_C\right)\right) + t_C\left(\overline{x}_C - x_C\left(t_C\right)\right)$, where $x_C\left(t_C\right)$ recognizes that x_C is a function of t_C. This function is given by (12), taking into account that $\overline{v} = v + t_C$. Note that x_C is then not a function of the baseline $\overline{x}_C$, which confirms that the $t_C \overline{x}_C$-part of the transfer is like a lump sum.

Since D's objective is to maximize

$$u_D - t_C \cdot \left(\overline{x}_C - x\right), \tag{17}$$

D would prefer to reduce the total transfer τ_C by reducing the baseline $\overline{x}_C$. However, D must ensure that the following incentive constraint for C is satisfied:

$$u_C^0 (x_C (t_C)) + t_C \cdot (\overline{x}_C - x_C (t_C)) \geq u_C^0 (\widehat{x}_C) \, \forall \widehat{x}_C > \overline{x}_C. \tag{IC$_\mathrm{C}$}$$

That is, C's payoff in equilibrium cannot be smaller than what C could achieve by optimizing as if there were no transfer. In equilibrium, $\overline{x}_C$ will be reduced by D until (IC$_\mathrm{C}$) binds with equality.

Proposition 5. *When D contracts with C, the contract $(t_C, \overline{x}_C)$ is:*

$$\begin{aligned} t_C &= d, \\ \overline{x}_C &= x_C (0) - \frac{d}{4a (b + c)}. \end{aligned} \tag{18}$$

Thus, the optimal rate $t_C = d$ is very simple and independent of the parameters in the model, whether the country is weak or strong, or whether extraction is sales-driven or protection-driven. To derive the result, just substitute (IC$_\mathrm{C}$) into (17), and note that D is induced to maximize the sum $u_C + u_D$.

Corollary 5. *When D contracts with C, the outcome is first best. C faces $t_C = d$ and induces its districts to select x_{FB} by setting the average tax equal to t_{FB}.*

The baseline $\overline{x}_C$ will be set such that (IC$_\mathrm{C}$) binds and C is exactly indifferent between choosing $x_C (t_C)$ and $x_C (0)$. The indifference means that the benchmark $\overline{x}_C$ will be strictly smaller than the business-as-usual level $x_C (0)$, as illustrated by (18), since otherwise C would strictly benefit from the contract. If $\overline{x}_C$ were not dictated by D, but instead had to equal some historical or business-as-usual level, then D would prefer some other $t_C \neq d$, and the first best would *not* be implemented. This result disproves the typical presumption that the reference level should equal the business-as-usual level.

4.3 Contracts under Decentralization

We now return to the model in which n districts act noncooperatively when deciding on the x_i's. As explained below, it suffices to consider actual conservation contracts of the form:

$$\tau_i = \max \{0, (\overline{x}_i - x_i) \, t_i\},$$

where $\overline{x}_i$ is the baseline for district i. Suppose D unilaterally designs the contract $(t_i, \overline{x}_i)$ for every $i \in M \subseteq N$, where $m = |M| \leq n$. Even if D would like to contract with all n districts, this may be unfeasible for exogenous (or political) reasons.

Just as under centralization, D must ensure that a district is no worse off in equilibrium where $x_i < \overline{x}_i$ than the district could be by ignoring the contract and picking any other extraction level $\widehat{x}_i > \overline{x}_i$:

$$u_i^0(\mathbf{x}(\mathbf{t})) + t_i \cdot (\overline{x}_i - x_i) \geq u_i^0(\widehat{x}_i, x_{-i}(\mathbf{t}))\, \forall \widehat{x}_i > \overline{x}_i. \qquad (\text{IC}_i)$$

The problem for D is to select the m pairs $(t_i, \overline{x}_i)$ in order to maximize $u_D - \sum_{i \in M} t_i \cdot (\overline{x}_i - x_i)$ subject to the m incentive constraints.

Proposition 6. *Suppose D contracts with $m \leq n$ districts. The optimal contract for D is:*

$$\begin{aligned} t_i &= \frac{2}{n+1} d, \\ \overline{x}_i &= x_i(\mathbf{0}) + \frac{4m - 3(n+1)}{4a(b+c)(n+1)} t. \end{aligned} \qquad (19)$$

Naturally, when there is only one district ($m = n = 1$), Proposition 6 coincides with Proposition 5. With $n > 1$ districts, however, contracting with district i means that x_i will decrease but every other x_j will increase: In fact, (14) shows that for every unit by which x_i is reduced, x is reduced by only $1/n$ units. This ratio makes it costly for D to reduce exctraction when n is large, and the optimal contract is weakened. Since the ratio is independent of the other parameters in the model, so is the rate t.

Similarly, x_i will increase when D contracts with several other districts. Thus, the larger is m, the larger must the baseline be for the contract to remain relevant. While contracts with i increases j's extraction level, the effect on the j's payoffs will depend on the externality e. If e is large, j benefits when D contracts with $i \neq j$; D is unable to cash in on this benefit, since the contract with i increases j's payoff also at the threat point when j ignores D's contract. It is thus intuitive that D will offer contracts to i that are too weak relative to the first best when the externality e is large.

If instead e is small and negative, then j loses when D contracts with i. However, D does not need to compensate j for this loss, since j loses also at the threat point and when j ignores a contract offered by D. Thus, when logging is mainly illegal, D offers contract that are too strong compare to the socially optimum.

Since the first-best t increases in e, while the equilibrium t does not, we get the following corollary.

Corollary 6. *At the equilibrium contracts, the conservation level is too large compared to the* *first best if and only if districts are weak (e/d is small):*

$$x < x_{FB} \Leftrightarrow \frac{\sum_{i \in N} t_i}{n} > t_{FB} \Leftrightarrow \frac{e}{d} < -\frac{(n+1)^2 - 4m}{n^2 - 1}.$$

Remark 4: Contract linearity and robustness. The donor can do no better with more general contracts (that are nonlinear or multilateral). Since there is a one-to-one mapping between $\mathbf{t}$ and $\mathbf{x}$, it is easy to see that (IC$_i$) remains unchanged if D can use more general contracts: If D wants to implement a particular $\mathbf{x}$, it must offer each $i \in M$ a transfer $\boldsymbol{\tau}_i(\mathbf{x})$ that makes i weakly better off compared to selecting any other x_i (leading to no transfer), given x_{-i}:

$$u_i^0(\mathbf{x}(\boldsymbol{\tau})) + \boldsymbol{\tau}_i(\mathbf{x}(\boldsymbol{\tau})) \geq u_i^0(\widehat{x}_i, x_{-i}(\boldsymbol{\tau})) \, \forall \widehat{x}_i > \overline{x}_i.$$

However, with bilateral contracts there are multiple equilibria that can be avoided by using multilateral contracts: If the other districts believed that i would ignore the contract (by extracting $x_i > \overline{x}_i$), the other districts would extract less and $x_i > \overline{x}_i$ would be strictly preferred by i. If this equilibrium is better for i, then i would prefer to announce publicly that it rejects the contract with D and that it will not accept any payments from D in the future. If such a pledge were credible, then D would need to satisfy i's participation constraints as well as i's incentive constraints.

5 Contracts with Local vs. Central Authorities

So far, we have assumed that the donor contracted with certain districts and governments, and we took their numbers and authority levels to be exogenously given. In some cases, the donor may be able to decide whether it wants to contract with a set of districts independently, or whether it instead wants to contract with their common central government.

As a start, consider a subset $L \subseteq M$ containing $l \equiv |L|$ districts. If these districts centralize authority, then l, m, and n all decrease by the same number, denoted by Δ. If L centralizes to a single government, then $\Delta = l - 1$, but we do not require this. We assume that such a regime change does not influence the forest areas over which D can contract. Hence, D contracts with $m - \Delta$ governments after the regime change, while the number of districts without a contract stays unchanged at $n - m$.

We say that L is "large" (relative to N) if:

$$\underline{\epsilon}_L \equiv 1 - \frac{l}{n+1} - \frac{l - \Delta}{n - \Delta + 1} < 0. \tag{20}$$

That is, for L to be large, it is *necessary* that L contains a majority of the decision-making districts *before* centralization $(l > (n+1)/2)$, and it is *sufficient* that L contains a majority *after* decentralization $(l - \Delta > (n - \Delta + 1)/2)$. If L is not large, we say that L is "small."

Our first observation concerns the effect on conservation.

Proposition 7. *If $L \subseteq M$ centralizes, x decreases if and only if e/d is large or M is large, i.e., if:*

$$\frac{e}{d} \geq 2\epsilon_M = 2\left(1 - \frac{m}{n+1} - \frac{m-\Delta}{n-\Delta+1}\right).$$

If M is large (say, $m = n$), then we know from the earlier intuition that centralization reduces x when e/d is large. If M is small, however, a large number $(n - m)$ of other districts will increase x when M reduces x, and thus the condition becomes harder to satisfy. Proposition 7 generalizes Proposition 3 (for the case in which there is no contract, $d = 0$).

We will now discuss when the donor would prefer contracting with a central government rather than with local governments. If a central government C is already active and regulating the local governments, then C can always undo D's offers to the districts; decentralized contracts would then not be an option for D. If the central government is absent or passive, however, then D may evaluate whether it should contract with the districts or instead propose a contract to the union of some districts. The latter option may require central authorities to be activated or created.

As we have already noted, when districts are strong and extraction sales-driven, then a district benefits if the others conserve more, and thus also if the others are offered conservation contracts. These positive externalities are internalized by a benevolent central government maximizing the sum of utilities. When positive externalities are appreciated by the contracting partner, D can extract more of the districts' surplus (by reducing the baseline). Thus, when e is large, D benefits when authority is centralized.

If instead the externality e is small, as when districts are weak and extraction is protection-driven, the argument is reversed. A district then experiences negative externalities when others conserve or sign conservation contracts with D. Negative externalities will be taken into account by central authorities, who will thus reject the contract unless it involves larger transfers. In this case, therefore, D benefits from decentralized contracts. This holds even when the first best requires centralization.

Proposition 8. *Decentralized contracts are preferred by D if and only if e/d is small or M is small:*

$$\frac{e}{d} \leq \epsilon_M = 1 - \frac{m}{n+1} - \frac{m-\Delta}{n-\Delta+1}.$$

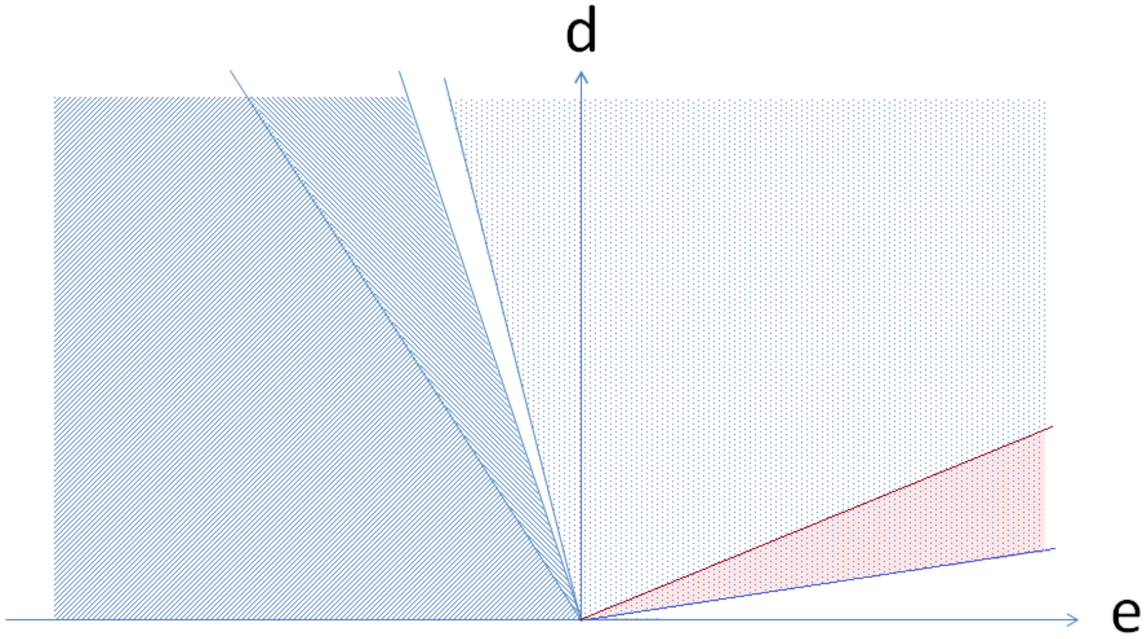

Figure 1: *Even if centralization leads to the first best, the donor prefers decentralized contracts when e and d are small (shaded area), while the districts prefer decentralization when e and d are large (dotted area). In the colored dotted area, the regime change raises extraction by more than the contracts reduce it. Lines are drawn for our two-district example.*

The donor is more likely to prefer decentralized contracts when M is small relative to N, since the other districts will, as a consequence, extract less when e is large. This result also implies that there is a unique number m that maximizes D's payoff (i.e., the second-order conditions w.r.t. m and Δ hold).

A comparison to Proposition 7 is interesting. When M is large, $2\underline{\epsilon}_M < \underline{\epsilon}_M < 0$. Thus, when $e/d \in (2\underline{\epsilon}_M, \underline{\epsilon}_M)$, D finds having decentralized contracts to be less expensive, even if centralization would have increased conservation. When M is instead small, $0 < \underline{\epsilon}_M < 2\underline{\epsilon}_M$. In this case, when $e/d \in (\underline{\epsilon}_M, 2\underline{\epsilon}_M)$, D finds having centralized contracts to be less expensive, even if decentralization would have increased conservation. When e/d is outside these intervals, D prefers the regime that maximizes conservation.

Corollary 7. *(i) If M is large, D always prefers decentralized contracts if this reduces x, but the converse is not true.*
(ii) If M is small, D always prefers centralized contracts if this reduces x, but the converse is not true.

Consider an example with two districts. Decentralized contracts are preferred by D in the shaded area, where $e/d < \underline{\epsilon}_M = -1/6 \approx -0.17$, even though decentralization reduces conservation when $e/d > 2\underline{\epsilon}_M = -1/3$ (where the shaded area has downward-sloping lines).

By extending the model, one can also derive the districts' preferences. One could then show that the districts prefer decentralization of authorities, before D offers the contracts, only when they are stronger and $e/d \in (\widehat{\epsilon}_L, \overline{\epsilon}_L) \approx (-0.16,\ 5.5)$, i.e., in the dotted area. Furthermore, one can show that when $e/d \in (8/3,\ 5.5)$, which corresponds to the colored and dotted area, the presence of D motivates the districts to decentralize and the accompanying increase in x outweighs the effect of the contracts.

6 Some Conclusions and Policy Lessons

This note presents a simple model of deforestation. It allows for many districts and recognize that since extracting some of the resource increases the harvest supply, it decreases the price and the monitoring costs for the part that is to be conserved. The externality from one district's conservation on others can be positive or negative, depending on state capacities and the size of the resource stock. The model can be used to study various types of resources and alternative motivations for extractions, but it is motivated in particular by deforestation in the tropics. The analysis generates several policy lessons.

First, decentralizing authority influences conservation. If districts are "strong" and extraction is sales-driven, then districts extract too much since they do not internalize the effect on other districts' profit. A transfer of authority from the local to the federal level will then lead to more conservation and less extraction. If districts are "weak" and extraction is driven by the high cost of protection, then districts might conserve too much since protection in one district can increase the pressure to extract in neighboring districts. In this case, centralizing authority will reduce conservation and increase extraction. These results may also help to explain the mixed empirical evidence: as discussed in the Introduction, decentralization has increased deforestation in Indonesia, while reducing it in other areas, such as the Himalayas.

Second, the optimal conservation contract depends on local institutions and the drivers of extraction. Under centralization to a single government, simple Pigou-like contracts are optimal and first best. With several independent districts, however, the equilibrium contract can lead to too much conservation when districts are weak and too little when they are strong. Finally, the model suggests that if logging is illegal, then it is less expensive to contract with local districts, but with sales-driven deforestation, then centralized contracts are better.

Note that the benchmark results we have derived rely on a number of limiting assumptions. In particular, ideally one should add a dynamic setting, the functional forms ought to be generalized, parameters might be privately known, and the outcome may also be stochastic.

7 Appendix: Proofs

Proof of Proposition 1.
Note that the first-order condition when maximizing (3) w.r.t. x_i and subject to (1) gives:

$$x_i = \frac{p}{a} + \frac{caX_i - v_i}{a(b+c)} \tag{21}$$
$$= \frac{\overline{p} - ax_{-i}}{2a} + \frac{caX_i - v_i}{2a(b+c)}, \tag{22}$$

if the right-hand side is in $[0, X_i]$. The second-order condition trivially holds. By summing over the x_i's as given by (21) and combining that sum with (1), we get (5) and (6), and by inserting (6) into (21), we get (4). *Q.E.D.*

Proof of Proposition 2.
(i) From (3) we immediately get (when $j \neq i$ and using the Envelope theorem):

$$\frac{\partial u_i}{\partial x_j} = \frac{\partial u_i}{\partial p}\frac{\partial p}{\partial x_j} = -a\left[(b+c)x_i - cX_i\right] = -\frac{(b+c)\overline{p} - acX - v_i n + \sum_{j\neq i} v_j}{n+1},$$

when we substitute in for (4). With (7), we can write $\partial u_i / \partial x_j = -e_i / (n+1)$.
(ii) When we combine (7) with (4) and (6), we get:

$$x_i = \frac{e_i}{a(b+c)(n+1)} + \frac{cX_i}{b+c} \text{ and } p = \frac{e_i}{(b+c)(n+1)} + \frac{v_i}{b+c}.$$

Thus, we can write (3) as:

$$u_i \equiv x_i\left((b+c)p - v_i\right) - pcX_i$$
$$= \left(\frac{e_i}{a(b+c)(n+1)} + \frac{cX_i}{b+c}\right)\frac{e_i}{n+1} - \left(\frac{e_i}{(b+c)(n+1)} + \frac{v_i}{b+c}\right)cX_i,$$

which can be written as (8). Given (7), differentiating (8) gives Corollary 2. *Q.E.D.*

Proof of Proposition 3.
(i) Since $u_D(\mathbf{x}) + \sum_{i\in N} u_i^0(\mathbf{x}) = bpx - cp(X - x) - vx - dx$, the f.o.c. when maximizing w.r.t. x can be written as (15). The second-order condition trivially holds.
(ii) With (13), we can write (11) as

$$x = \frac{ac}{a(b+c)}X + \frac{ne - \sum_i t_i}{a(b+c)(n+1)}. \tag{23}$$

This x equals x_{FB} if and only if (16) holds. *Q.E.D.*

Proof of Proposition 4.
For a given t_C, D prefers to reduce $\overline{x}_C$ as much as possible, so (IC_C) will bind. Solving (IC_C) for $(\overline{x}_C - x)t_C$ and inserting that term into (17), we note that D's objective is to maximize $-dx_C(t_C) + u_C^0(x_C(t_C)) - u_C^0(\widehat{x}_C) = u_D(x_C(t_C)) + u_C^0(x_C(t_C)) - u_C^0(\widehat{x}_C)$. D is thus maximizing the sum of payoffs (since $-u_C^0(\widehat{x}_C)$ is independent of t_C), implying the same outcome as in the first best: $x_C = x_{FB}$ and $t_C = d$.

To derive $\overline{x}_C$, note that we can rewrite a binding (IC_C) to:

$$t_C\overline{x}_C = u_C^0(\widehat{x}_C) - \left[u_C^0(x_C(t_C)) - t_C x_C\right], \tag{24}$$

where both $u_C^0(\widehat{x}_C)$ and the bracket follow from (8), and with (7) and (13), e_i is replaced by e when C ignores the contract while otherwise $e_i = e - t_C$. Thus, we can write (24) as:

$$t_C \overline{x}_C = \frac{1}{a(b+c)} \left[\frac{e^2}{4} - cavX - \frac{(e-t_C)^2}{4} + ca(v+t_C)X \right],$$

which can be rewritten as (18) when $t_C = d$. *Q.E.D.*

Proof of Proposition 5.
The proof starts by deriving $\max_{\widehat{x}_i} u_i^0(\widehat{x}_i, x_{-i}(\mathbf{t}))$. From (22) and (4), we find i's optimal response to $x_{-i}(\mathbf{t})$, if i decided to ignore the contract:

$$x_i^I = \frac{\overline{p} - ax_{-i}(\mathbf{t})}{2a} + \frac{caX_i - v}{2a(b+c)} = x_i + \frac{t_i}{2a(b+c)},$$

where x_i is given by (14). This results in a price

$$p^I = p - \frac{t_i}{2(b+c)},$$

where $p = \overline{p} - a\sum_i x_i(\mathbf{t})$. Thus,

$$\begin{aligned} u_i^0(\widehat{x}_i, x_{-i}(\mathbf{t})) &= \left[(b+c)p^I - v\right] x_i^I - p^I cX_i \\ &= \left[(b+c)\left(p - \frac{t_i}{2(b+c)}\right) - v\right]\left(x_i + \frac{t_i}{2a(b+c)}\right) - \left(p - \frac{t_i}{2(b+c)}\right) cX_i \\ &= u_i^0(\mathbf{x}(\mathbf{t})) + \left[(b+c)\left(p - \frac{t_i}{2(b+c)}\right) - v\right]\frac{t_i}{2a(b+c)} - \frac{t_i}{2}x_i + \frac{cX_i t_i}{2(b+c)} \\ &= u_i^0(\mathbf{x}(\mathbf{t})) + \frac{t_i^2}{4a(b+c)}, \end{aligned} \tag{25}$$

when we use (21). With this, (IC_i) boils down to

$$\tau_i \geq u_i^0(\widehat{x}_i, x_{-i}(\mathbf{t})) - u_i^0(\mathbf{x}(\mathbf{t})) = \frac{t_i^2}{4a(b+c)}. \tag{26}$$

(i) When only the IC's bind, D maximizes

$$\begin{aligned} & u_D + \sum_{i \in M} \left[u_i^0(\mathbf{x}(t)) - u_i^0(\widehat{x}_i, x_{-i}(\mathbf{t}))\right] \\ = & -d\left[\frac{ac}{a(b+c)}X + \frac{ne - \sum_i t_i}{a(b+c)(n+1)}\right] - \sum_{i \in M} \frac{t_i^2}{4a(b+c)}. \end{aligned}$$

For each t_i, $i \in M$, the first-order condition becomes

$$\frac{d}{a(b+c)(n+1)} - \frac{t_i}{2a(b+c)} = 0,$$

giving (i). The second-order condition trivially holds.

To find $\overline{x}_i^{IC}$, rewrite a binding (26) to:

$$\begin{aligned}
\tau_i &= t_i\left(\overline{x}_i - x_i\right) = \frac{t_i^2}{4a\left(b+c\right)} \Leftrightarrow \\
\overline{x}_i &= \frac{t_i}{4a\left(b+c\right)} + x_i\left(\mathbf{0}\right) - \frac{t_i n - \sum_{j\neq i} t_j}{a\left(b+c\right)\left(n+1\right)} = x_i\left(\mathbf{0}\right) + \frac{4m-3\left(n+1\right)}{4a\left(b+c\right)\left(n+1\right)}t \\
&= x_i^I - \frac{t_i}{4a\left(b+c\right)} = \frac{x_i + x_i^I}{2}.
\end{aligned}$$

(ii) Note that (PC_i) can be rewritten as:

$$t_i\overline{x}_i \geq u_i^0\left(\mathbf{x}\left(\mathbf{t}_{-i}\right)\right) - \left[u_i^0\left(\mathbf{x}\left(\mathbf{t}\right)\right) - t_i x_i\right]$$

where both $u_i^0\left(\mathbf{x}\left(\mathbf{t}_{-i}\right)\right)$ and the bracket follow from (8), so:

$$\begin{aligned}
t_i\overline{x}_i &\geq \frac{1}{a\left(b+c\right)}\left[\left(\frac{e+\sum_{j\neq i} t_j}{n+1}\right)^2 - cavX_i\right] \\
&\quad -\frac{1}{a\left(b+c\right)}\left[\left(\frac{e - nt_i + \sum_{j\neq i} t_j}{n+1}\right)^2 - ca\left(v+t_i\right)X_i\right] \\
&= \frac{t_i}{a\left(b+c\right)}\left[\frac{2n\left(e+\sum_{j\neq i} t_j\right) - n^2 t_i}{\left(n+1\right)^2} + caX_i\right]. \qquad (27)
\end{aligned}$$

Thus, D's problem becomes to maximize:

$$\begin{aligned}
& u_D + \sum_{i\in M}\left[u_i^0\left(\mathbf{x}\left(\mathbf{t}\right)\right) - u_i^0\left(\mathbf{x}\left(\mathbf{t}_{-i}\right)\right)\right] = -dx + \sum_{i\in M}\left[x_i t_i - t_i\overline{x}_i\right] \\
&= -dx + \sum_{i\in M} x_i t_i - \sum_{i\in M}\frac{t_i}{a\left(b+c\right)}\left[\frac{2n\left(e+\sum_{j\neq i} t_j\right) - n^2 t_i}{\left(n+1\right)^2} + caX_i\right].
\end{aligned}$$

Since x_i is given by (14) and x by (23), the f.o.c. w.r.t. t_i becomes:

$$\begin{aligned}
0 &= \frac{d}{a\left(b+c\right)\left(n+1\right)} + \frac{\left(b+c\right)\overline{p} + ac\left(n+1\right)X_i - acX - v - 2t_i n + \sum_{j\neq i} t_j}{a\left(b+c\right)\left(n+1\right)} \\
&\quad + \frac{1}{a\left(b+c\right)\left(n+1\right)}\sum_{j\in M\backslash i} t_j - \frac{1}{a\left(b+c\right)}\left[\frac{2n\left(e+\sum_{j\neq i} t_j\right) - 2n^2 t_i}{\left(n+1\right)^2} + caX_i\right] \\
&\quad - \sum_{j\in M\backslash i}\frac{t_j}{a\left(b+c\right)}\frac{2n}{\left(n+1\right)^2}.
\end{aligned}$$

Note that X_i disappears from the f.o.c., so we get the same $t_i = t_{PC}$ for every $i \in M$. The f.o.c. thus simplifies to:

$$\begin{aligned}
0 &= \left(n+1\right)d + \left(n+1\right)e - t_{PC}\left(2n-m+1\right)\left(n+1\right) \\
&\quad + \left(m-1\right)\left(n+1\right)t_{PC} - \left[2ne - 2n\left(n-m+1\right)t_{PC}\right] - 2n\left(m-1\right)t_{PC} \\
&= \left(n+1\right)d - \left(n-1\right)e - 2\left[\left(n-m+1\right) + n\left(m-1\right)\right]t_{PC},
\end{aligned}$$

which reveals that the second-order condition clearly holds. By solving for t, we get:

$$t_{PC} = \frac{(n+1)\,d-(n-1)\,e}{2\left[(n-m+1)+n\,(m-1)\right]} = \frac{(n+1)\,d-(n-1)\,e}{2+2m\,(n-1)}.$$

We can find $\overline{x}_i$ by inserting t_{PC} and $x_{i,0}$ from (14) into (27):

$$\overline{x}_i^{PC} = x_i\left(\mathbf{0}\right) + \frac{1}{a\,(b+c)}\left[\frac{n-1}{(n+1)^2}e + \frac{2n\,(m-1)-n^2}{(n+1)^2}t\right].$$

(iii) Note that (IC) is harder to satisfy than (PC) if $\overline{x}_i^{IC} > \overline{x}_i^{PC}$. A simple comparison completes the proof. *Q.E.D.*

Proof of Proposition 6.
(i) It is easy to see that (IC$_i$) remains unchanged if D can use more general contracts: If D wants to implement a particular vector $\mathbf{x}$, it must offer each $i \in M$ a transfer $\boldsymbol{\tau}_i\left(\mathbf{x}\right)$ that makes i weakly better off compared to selecting any other x_i (leading to a different transfer). To discourage such deviations, D should ensure that i receives no transfer if i deviates from the implemented plan. Thus, the incentive constraint is

$$u_i^0\left(\mathbf{x}\left(\boldsymbol{\tau}\right)\right) + \boldsymbol{\tau}_i\left(\mathbf{x}\left(\boldsymbol{\tau}\right)\right) \geq u_i^0\left(\widehat{x}_i, x_{-i}\left(\boldsymbol{\tau}\right)\right) \forall \widehat{x}_i > \overline{x}_i,$$

just as before.
(ii) Next, note that the participation constraint can always be weakened to make it weaker than the incentive constraint. To see this, write the participation constraint as:

$$u_i^0\left(\mathbf{x}\left(\boldsymbol{\tau}\right)\right) + \boldsymbol{\tau}_i\left(\mathbf{x}\left(\boldsymbol{\tau}\right)\right) \geq u_i^0\left(\mathbf{x}\left(\boldsymbol{\tau}_{-i}\right)\right),$$

and note that it is always possible to select $\boldsymbol{\tau}\left(\mathbf{x}\right)$ in such a way that $\mathbf{x}_{-i}\left(\boldsymbol{\tau}_{-i}\right) = x_{-i}\left(\boldsymbol{\tau}\right)$, that is, such that no $j \neq i$ will change x_j if i announces that i will not accept transfers from D. This is achieved, for example, if j receives transfers only when $x_j = x_j\left(\boldsymbol{\tau}\right)$. Of course, it may be that the transfer $\boldsymbol{\tau}_j$ must be larger when i rejects the contract and thus selects $x_i \neq x_i\left(\boldsymbol{\tau}\right)$, but this larger transfer will not have to be paid by D in equilibrium.
(iii) Thus, only the incentive constraint will bind when $\boldsymbol{\tau}\left(\mathbf{x}\right)$ can be a general function. Inserting the binding incentive constraints into D's objective function gives, as before, that D selects $\boldsymbol{\tau}$ or, equivalently, $\mathbf{x}$, to maximize:

$$u_D + \sum_{i \in M} u_i^0\left(\mathbf{x}\right) - u_i^0\left(\widetilde{x}_i\left(x_{-i}\right), x_{-i}\right),$$

where $\widetilde{x}_i\left(x_{-i}\right) = \arg\max_{x_i} u_i^0\left(x_i, x_{-i}\right)$. This is the same problem as in the proof of Proposition 5(i), and the outcome for x_i and $\boldsymbol{\tau}$ are thus also identical. *Q.E.D.*

Proof of Proposition 7.
From now on we frequently use $y \equiv e/d$. The following proof is more general than needed, since we allow for a regime change that changes $q \equiv n-m$ as well as m (to q' and m'), even though the text above does not consider changes in q. When inserting (19) into (11), we get:

$$x = \frac{c}{b+c}X + \frac{e}{a\,(b+c)} - \frac{e+2dm/\left(n+1\right)}{a\,(b+c)\,(n+1)}.$$

Thus, with decentralization, x increases if:

$$\begin{aligned}
0 &> \frac{e(n'-n)}{(n+1)(n'+1)} - 2d\frac{m'(n+1)^2 - m(n'+1)^2}{(n+1)^2(n'+1)^2} \Leftrightarrow \\
y &< 2\frac{m'(n+1)^2 - m(n'+1)^2}{(n+1)(n'+1)(n'-n)} = \frac{2}{(n'-n)}\left[m'\left(1-\frac{n'-n}{n'+1}\right) - m\left(1+\frac{n'-n}{n+1}\right)\right] \\
&= 2\left[\frac{m'-m}{n'-n} - \frac{m'}{n'+1} - \frac{m}{n+1}\right] = 2\left[1 - \frac{q'-q}{n'-n} - \frac{m'}{n'+1} - \frac{m}{n+1}\right].
\end{aligned}$$

Setting $q = q'$ and $\Delta = m' - m = n' - n$ completes the proof. *Q.E.D.*

Proof of Proposition 8.
With (23) and (26) we can write D's payoff as a function of m:

$$\begin{aligned}
-dx - \sum_{i \in M} \frac{t_i^2}{4a(b+c)} &= -d\left(\frac{ac}{a(b+c)}X + \frac{ne - m\left(\frac{2}{n+1}d\right)}{a(b+c)(n+1)}\right) - \frac{m\left(\frac{2}{n+1}d\right)^2}{4a(b+c)} \\
&= \frac{d^2}{a(b+c)}\frac{m}{(m+q+1)^2} - \frac{dcX}{b+c} - \frac{de}{a(b+c)}\frac{m+q}{m+q+1}.
\end{aligned}$$

By comparison, $m' > m$ increases D's payoff if

$$\begin{aligned}
\frac{m'}{(m'+q+1)^2} - y\frac{m'+q}{m'+q+1} &> \frac{m}{(m+q+1)^2} - y\frac{m+q}{m+q+1} \Leftrightarrow \\
y &< 1 - \frac{m'}{(m'+q+1)} - \frac{m}{(m+q+1)}. \quad \textit{Q.E.D.}
\end{aligned}$$