DEFORESTATION AND CONSERVATION CONTRACTS

Bård Harstad

Lecture notes
March 2019
University of Oslo

1 Motivation on Deforestation and REDD

These lecture notes will introduce a simple model that can be used to discuss deforestation and other resource extraction problems, whether the resource extraction activity is legal or illegal. The model can be used to explain some empirical puzzles, and thereafter it will be used to discuss how conservation contracts, such as REDD agreements, should be designed.

Sections 2 and 3 are relatively easy, while Section 4 is more demanding. Section 4.3 is most demanding and can be skipped. The appendix contains details and technical proofs for the particularly interested student.

Deforestation in the tropics is an immensely important problem. The cumulative effect of deforestation amounts to about one quarter of the anthropogenic greenhouse gas emissions that generate global warming (?). The annual contribution from deforestation to CO_2 emissions is around 10 percent (?), and the percentage is even higher for other greenhouse gases. In addition to the effect on global warming, deforestation leads to huge losses in biodiversity. Nevertheless, tropical forest loss has been *increasing* at an average rate of $2,101$ km^2 yearly since 2000.[1]

A substantial fraction of deforestation is illegal. Although we do not know the exact numbers—thanks to the very nature of illegality—estimates suggest that between thirty and eighty percent of tropical deforestation is illegal, depending on the country in question. For a set of countries with tropical forests, the estimated fractions of logging that is illegal, as well

[1] ?. ? offer more precise estimates of deforestation between 2000 and 2005. The overall message that tropical deforestation has been increasing remains robust.

as these countries' forest cover and deforestation rates, are reported in the below table.[2]

Country\Year	*Forest Cover in 2000 (1000 ha)*	*Deforestation rate in 2000-2010*	*Illegal logging in 2013*
Brazil	545943	5%	> 50%
Cameroon	22116	10%	65%
Ghana	6094	19%	70%
Indonesia	99409	5%	60%
Laos	16433	6%	80%
Malaysia	21591	5%	35%
Papua New Guinea	30133	5%	70%
Republic of the Congo	22556	1%	70%

Illegal resource extraction is substantial also for other types of resources, such as gold and coal.[3]

It is costly for countries to protect their resources and prevent illegal extraction. Brazil, the country with the largest tropical forest cover, has in the recent years spent more than $100m (USD) on monitoring and controlling illegal forest activities. The expenditures have also increased over the last fifteen years, according to Figure 1.[4]

One problem with the effort to reduce deforestation and conserve is so-called leakage. Markets for timber and agricultural products are integrated, and reduced logging at one location raises the regional price of timber or agricultural products, and thus it 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 U.S. 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.[5]

[2] The numbers on illegal logging are borrowed from **?**: 12 and they measure the percentages of total timber production that are estimated to be illegal. The numbers on forest cover are from Mongabay http://rainforests.mongabay.com/deforestation, accessed 16 Oct. 2016.10.16, and the deforestation rates are changes in the total (net) Forest Cover, 2000-2010, relative to the Forest Cover in 2000. Other estimates on illegal deforestation are of similar magnitude, also for the fraction of illegal conversion of land to agriculture (rather than timber): see **?**, **?**, and **?**.

[3] On gold, **?**: 242 document that: "some 90% of Indonesia's 65,000 –75,000 small-scale gold miners are operating illegally, as well as over 80% of the 200,000 gold panners operating in the Phillipines;" "Over 90% of Brazilian gold panners are operating illegally;" and: "Many of Zimbabwe's 300,000 gold panners along Mazowe, Angwa and Insiza Rivers are unregistered (39%), as are some of Zambia's gemstone miners (15%)." On coal, see **?**.

[4] The numbers are from **?**, who have looked at 2055 budgets from 116 Brazilian budget programs and focused on those actions whose objective and description directly aim at forest conservation. They classified the expenditures as institutional cost and operational cost. The operational costs are further classified into enabling, incentive, and disincentive costs on the basis of the goal of the instrument. Figure 1 only reports the disincentive costs, roughly defined as follows: "Disincentive-based instruments included the establishment and management of protected areas, monitoring and control of deforestation, forest degradation and forest fires, as well as the regulation of economic activities that cause high social and environmental impacts on forest areas, such as mining" (**?**: 213).

[5] The numbers for the U.S. are from **?** and **?**; **?** provided the study of Vietnam. Other estimates complement these numbers: according to **?**, 75% of EU's, 70% of that of Australia and New Zealand, and 46% of that of United States' reduced timber harvest are replaced by increased logging in the tropics. **?** summarize the

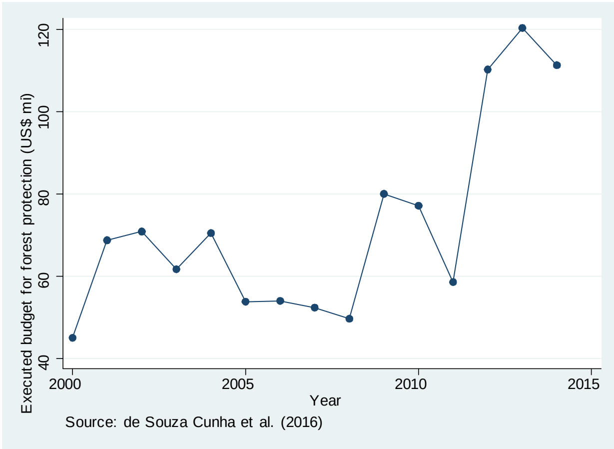

Figure 1: *Spending on disincentive-based instruments to protect forests in Brazil*

In these notes, we develop a model of resource extraction that takes into consideration the above facts. Although the model itself is general and can be applied to many types of exhaustible resources (such as land or fossil fuels), it fits well to the case of tropical deforestation. In the model, logging can be illegal as well as controlled by the governments. To protect a parcel of the forest, the government must monitor so much that the expected penalty is larger than the profit from illegal logging. The total enforcement cost is thus larger when there is a large profit of harvesting (timber or agricultural products), as will be the case when there is little logging elsewhere. Thus, a district may want to leave substantial parts of the forest unprotected, only to reduce the pressure and thus the enforcement cost on the part that is to be protected.

The part of the forest that is unprotected will be logged. This logging can be legal or illegal as far as our model is concerned, since the district's weight on the associated profit can be any number, small or large. If this weight is large while the enforcement cost is small, the game between the districts is similar to a standard Cournot game: if one district extracts less, the (timber) price increases and the other districts are better off. This pecuniary externality implies that if real decision powers were centralized to a federal government, extraction would be deliberately reduced in order to increase the profit for everyone.

This insight is reversed if the enforcement cost is large, or if the districts are unable to benefit much from the profit of logging. In these cases, reducing extraction in one district raises the price and thus the enforcement cost for the others. The larger cost makes the other districts worse off. A central authority would take this negative externality into account so, in this situation, centralization would lead to more logging.

Empirically, the effect of decentralization on deforestation can indeed go either way, de-

findings on forest conservation leakage: the estimates vary widely between 5% and 95%, but typical estimates are around 40%.

pending on the country in question. For Indonesia, ?: 1751 find that "as the number of jurisdictions within a provincial wood market increases, deforestation rises and prices fall." In Nepal, in contrast, deforestation accelerated under national management after 1957, and deforestation decreased by 14 percent after 1993 in response to decentralization of forest management, according to ?. Consistent with the latter finding, ?: 4146 find that "forests in the Indian central Himalayas have been conserved at least as well and possibly better under decentralized management and at much lower cost."[6] The difference between the countries is puzzling, but it is consistent with our theory, as we explain in Section 5.

Our second contribution is to use the model to analyze the design and the effects of conservation contracts. As mentioned above, tropical deforestation is harmful also for the North. The global negative externalities of deforestation amount to \$2–4.5 trillion a year, according to *The Economist*. In addition, 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 of \$10/tCO_2.[7] 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. Also for other types of resources, such as fossil fuel reserves, a climate coalition's optimal policy may be to pay nonparticipants to conserve particular reserves (?).[8]

In our model, a donor can offer payments to districts in return for reduced levels of resource extraction. If the donor contracts with a single central government, the outcome will be first best. When resource extraction is decentralized, however, there are contractual externalities. If the donor contracts with one district, the other districts benefit by extracting more, as long as they profit from extraction and find enforcement inexpensive. This benefit induces the donor to offer less—there will be too much extraction in equilibrium—and the donor would prefer to contract with a central authority instead, if that were feasible.[9]

In contrast, if the enforcement cost is large and districts profit less from extraction, a district's outside option worsens when the donor contracts with a neighbor. In this case, the donor finds it less expensive to contract with the districts individually, and these contracts

[6] ?, ?, and ? present similar evidence to that of ?.

[7] See ? and ?, respectively.

[8] Payments for environmental services (PES) can be important in many situations, even though 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). See ? for PES more generally, or ? and ? for an explanation of the difference between alternative concepts such as RED, REDD, and REDD+.

[9] In line with this argument, Norway recently declined to contract with the region Madre de Dios in Peru and stated that it would only contract at the national level.

lead to too little extraction, compared to the first best. The negative contractual externality also implies that the districts might become worse off when the donor offers conservation contracts, if the environment is characterized by large enforcement costs and illegal logging.

2 A Model of Deforestation

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

The novel part of the theory is the way in which we model enforcement. There are $n \geq 1$ districts and X_i is the size of the forest or resource stock in district $i \in N = \{1, ..., n\}$. Parameter $v > 0$ measures district i's value of each unit of X_i that is conserved. Each parcel or unit of this stock can be illegally cut, so i must decide how much to monitor and protect the various units. With free entry of illegal loggers, the price p they can obtain for cutting unit j will be compared with the expected penalty, θ_j, which they face when logging illegally that unit of the forest. The expected penalty is preventive if and only if it is larger than the benefit from logging: $\theta_j \geq p$. The price p will be a decreasing function of aggregate extraction.

We let districts set the expected penalties in advance in order to discourage extraction. This approach contrasts with the approach in much of the literature on inspection games (**?**), where the decisions to monitor and violate the law are taken simultaneously, but our assumption is more in line with the real world, in which penalties and monitoring follow from legislation that is publicly committed to at the outset. In principle, the expected penalty can be increased by a larger fine or penalty, but there is an upper boundary for how much the fine can be increased in economies with limited liability. To raise the expected penalty further, one must increase the monitoring probability, which is costly.[10]

We let $c \geq 0$ denote the cost of increasing monitoring enough to raise the expected penalty by one. Since enforcement is costly and will succeed if and only if $\theta_j \geq p$, for every unit j it is optimal with either $\theta_j = p$ or $\theta_j = 0$. Thus, district i sets the θ_j's for the different units of the forest so as to maximize:

$$\int_{X_i} \left(v \cdot \mathbf{1}_{\theta_j \geq p} - c \cdot \theta_j\right) dj = (v - cp)\left(X_i - x_i\right), \quad (1)$$

[10] If π is the probability of being caught, while ω is the largest possible penalty (for example, the wealth of an illegal logger), then monitoring is effective if and only if $\pi \geq p/\omega$.

where $X_i - x_i = \int_{X_i} \mathbf{1}_{\theta_j \geq p} dj$ is the amount that is conserved, and $\mathbf{1}_{\theta_j \geq p} = 1$ if $\theta_j \geq p$ and $\mathbf{1}_{\theta_j \geq p} = 0$ otherwise. It follows that a part of the forest will be protected and conserved, perhaps as a national park, while the remaining part (x_i) will not be protected and therefore it will eventually be cut. The model thus predicts that conservation policies will be "place-based" (for example, restricted to geographically limited but protected national parks), as seems to be the case in many countries, such as Indonesia, where "national and provincial governments zone areas of forest land to be logged" (**?**: 1328).

Given (1), district i's problem boils down to choosing $x_i \in [0, X_i]$. Since p is a decreasing function of aggregate extraction, $x \equiv \sum_{i \in N} x_i$, district i's payoff can be written as:

$$u_i(x_i, x_{-i}) = bp(x)\, x_i + (v - cp(x))\,(X_i - x_i)\,, \tag{2}$$

where $x_{-i} \equiv \sum_{j \in N \setminus i} x_j$ and parameter $b \geq 0$ measures the *weight* district i places on the profit $p(x)\, x_i$ of the x_i units that are extracted in the district. By varying the parameters b and c, the model nests several special cases that have intuitive interpretations. In the simplest model of illegal extraction, one would think that $b = 0$. However, if the government places some weight on the welfare or profit of the illegal loggers, who might be poor citizens, then $b > 0$ may measure this weight. Or, if the loggers are large corporations, b can measure the *probability* that the profit is detected and captured at the border, for example. Alternatively, all extraction x_i may be *legal* and controlled by the districts. In this case, b is likely to be large, since a district can spend its revenues just as it pleases. If b is large while the enforcement cost is small, extraction is purely sales-driven, just like in a standard Cournot game. In this situation, we may say that the *property rights are strong*. In contrast, we may say that the *property rights are weak* if the enforcement cost c is large, while b is small. Note that this situation may arise whether extraction is illegal (so that a district's benefit b from the profit is zero or small), or if extraction is legal, if just the ability to benefit from the profit is small relative to the enforcement cost.[11] We will be more precise about these concepts below.

In general, these parameters are likely to vary with the details of the political system, which may pin down the fraction (b) of the public revenues that a decision maker can capture and the cost (c) of ensuring that local public agencies are not corrupt. In addition, geography may play an important role in determining the cost of protecting a resource. **?** have shown

[11] As an intermediate possibility, b can be interpreted as the *fraction* of total extraction that is legal. To see this, suppose that if the government in district i decides to extract x_i^s units for sale in order to raise revenues, such extraction may require infrastructure and roads, which in turn may also proportionally raise the amount of illegal extraction to αx_i^s, where $\alpha > 0$ measures the amount of illegal extraction when the government extracts and, for example, builds roads. Such a complementarity is documented by **?**. Total extraction is then $x_i = (1 + \alpha)\, x_i^s$ even though the fraction of the total profit, captured by the government in district i, is only $b \equiv 1/(1 + \alpha)$. The larger the fraction of illegal extraction, the smaller b is.

that natural resources located close to international borders are more likely to motivate war between countries, essentially implying that the location of the resource influences the cost of protecting them. At the same time, the fact that countries go to war over oil suggests that it is very important to become the owner and hold the property rights over this type of resource. One may not have the same willingness to fight if the resource, in any case, were burdened with illegal extraction. Thus, minerals and oil might be characterized by a smaller enforcement cost and a larger benefit to the owner that extracts it. This argument suggests that the geographical concentration of the resource is also important. Forests are naturally spread out, and may thus be harder to monitor than geographically concentrated gold mines.

To simplify, we start by considering the case of a linear demand curve:

$$p(x) = \overline{p} - ax, \tag{3}$$

where $\overline{p}$ and a are positive constants. The Appendix allows for nonlinear demand and proves that our main results continue to hold, qualitatively.

Remark 1: Generalizations. Our model is simple and can easily be extended in several ways. For example, we allow for district-specific v_i's in the Appendix. One can also allow the districts to take into account some of the consumer surplus: this generalization will merely make the analysis messier without altering the conclusions qualitatively. Since tropical timber and agricultural products are to a large extent exported, it is reasonable that districts will *not* take consumer surplus into account in reality.

Instead of letting parameter $v \geq 0$ measure the value of the forest, it can alternatively represent a district's marginal cost of *extracting* the resource. In this case, it is more natural to write the utility function as:

$$\widehat{u}_i(x_i, x_{-i}) = bp(x)x_i - cp(x)(X_i - x_i) - vx_i.$$

This utility function is equivalent to (2) in our analysis, since we can define $u_i(x_i, x_{-i}) \equiv \widehat{u}_i(x_i, x_{-i}) + vX_i$, and since the last term, vX_i, is a constant.

Furthermore, note that we link the districts by assuming that the extracted resource 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 extracted resource 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 (2)–(3). It is thus equivalent to the model described above.

Finally, a static model represents the real world well if the purpose of extraction is to produce (f.ex. agricultural) products forever after on the land, since then p is driven by the accumulated x, and not the per-period quantity. For timber or fossil fuels, one may argue that the time profile will be more important.

Remark 2: Nonpecuniary externalities. This note emphasizes that the districts influence each other through the market. However, the model can easily be reformulated to also allow for nonpecuniary externalities, meaning that district i loses $\widetilde{v}_{-i} > 0$ when the other districts extract. To see that our model already permits such externalities, suppose that i's payoff is:

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

where $\widetilde{X}_i$ is i's stock and $\widetilde{v}_i$ is the marginal value of i's stock for i. This utility function can be rewritten as (2) if we simply define $v_i \equiv \widetilde{v}_i - \widetilde{v}_{-i}$, $X_i \equiv \widetilde{X}_i - \widetilde{v}_{-i}/ca$, and $u_i \equiv \widetilde{u}_i - (X_i - \overline{p}/a + \widetilde{v}_i/ac)\widetilde{v}_{-i}$, where the last term is a constant. Therefore, our analysis would be unchanged if we allowed for such nonpecuniary externalities: any interested reader can account for a larger externality $\widetilde{v}_{-i}$ by reducing v_i and X_i in the results below.[12] Although we now simplify by assuming that the v_i's are homogenous, heterogeneous v_i's are permitted in the Appendix.

3 The Equilibrium of the Model

This section discusses the equilibrium amount of extraction and conservation and investigates the effect of political centralization. These results are interesting in themselves, they might explain empirical irregularities, and they are necessary to describe before we analyze conservation contracts in the next section.

[12] A nonpecuniary externality can simply be added to the pecuniary externality such as it is defined by (6), below. Thus, our measure of the total externality e, as it is defined in Section 4.2, will then increase in both the pecuniary and the nonpecuniary externality, and it can be written as:

$$e \equiv (b + c)\overline{p} - ac\widetilde{X}_i - \widetilde{v}_i + (n + 1)\widetilde{v}_{-i}$$

The fact that we can reformulate $\widetilde{u}_i$ as u_i hinges on the assumption that the demand function is linear. With nonlinear demand and externalities, the pecuniary and the nonpecuniary externalities may interact through p in arbitrary and complex ways. It is beyond the scope of this paper to investigate these effects in detail.

3.1 Equilibrium Conservation

Each district $i \in N$ decides on x_i, taking as given the other districts' extraction level, $\mathbf{x}_{-i} = \{x_j\}_{j \in N \setminus i}$ and $x_{-i} = \sum_{j \in N \setminus i} x_j$. It is easy to see that extraction levels are strategic substitutes, so that i prefers to extract less if other districts are expected to extract more. This property holds for all parameters of the model: If property rights are weak in that c is large and b is small, a large x_j lowers $p(x)$ and thus the protection cost when i decides on how much to conserve. If property rights are strong in that c is small while b is large, a large x_j lowers the price and therefore the marginal profit i gets from extraction.

We refer to the equilibrium values of x_i and x as x_i^0 and x^0, respectively. To ensure that the solutions for the x_i^0's are interior, it is convenient to assume that all stocks are large and that $X \equiv \sum_{i \in N} X_i > \overline{p}/a \Leftrightarrow p(X) < 0$. If X_i were small, we would typically get a corner solution where district i extracted zero or everything. Such corner solutions are not worth our attention here, since they are unlikely to be robust or survive under more general functional forms.

Proposition 1. *If c or X_i increases, or v decreases, then x_i increases, x increases, and p decreases. Furthermore, x_j decreases in X_i, $j \neq i$:*

$$\begin{aligned} x_i^0 &= \frac{b\overline{p} - v}{ab(n+1)} + c\frac{v + ab[(n+1)X_i - X]}{ab(b+c)(n+1)}, \text{ and} \\ x^0 &= \frac{nb\overline{p} - nv}{ab(n+1)} + c\frac{nv + abX}{ab(b+c)(n+1)}, \text{ if} \\ X_i &\geq \max\left\{\frac{(b+c)\overline{p} - caX - v}{ab(n+1)}, -\frac{(b+c)\overline{p} - caX - v}{ac(n+1)}\right\}, \forall i \in N. \end{aligned} \tag{4}$$

Quite intuitively, a district extracts more if the enforcement cost c is large. Furthermore, a district i extracts more if its own resource stock is large, since a larger x_i reduces $p(x)$ and thus the protection cost for the (large) remaining amount. Similarly, because $j \in N \setminus i$ extracts less when X_j is large, the level of x_i decreases in X_j as x_i and x_j are strategic substitutes: when $p(x)$ is small, it is both less profitable for i to extract, and less expensive for i to protect its resource. For both reasons, district i conserves more when X_j is large, for $j \neq i$.

Proposition 1 also shows that aggregate extraction is larger if demand is large (as measured by $\overline{p}/a$). In this case, understanding parameter b's ambiguous effect is straightforward. In the expressions for x_i^0 and x^0, the first terms on the right-hand sides refer to the equilibrium Cournot levels (as if $c = 0$): these terms increase in the benefit of profit, b. The second terms show the additional extraction level due to costly enforcement. The larger c is, the larger these terms are, and extracting more than the Cournot level becomes optimal even though this reduces revenues. When b increases, revenues become more important and thus the second term is smaller. In sum, x increases in b if and only if protection is inexpensive:

$$\frac{\partial x}{\partial b} > 0 \text{ if and only if } c < \frac{n\overline{v}}{aX}. \tag{5}$$

3.2 The Effect of (De)Centralization

In this subsection we study the effect of (de)centralization on conservation. If a set of districts centralizes authority, we will assume that the forest stocks are pooled and that the extraction rates are set to maximize the sum of the merging districts' payoffs. Thus, the aggregate resource stock X remains unchanged, while the number of relevant governments n declines. To isolate this effect, we assume b and c remain the same after centralization. Any changes in b and c could be added to the effects we isolate.

To understand the effect of decentralization, note that there are nontrivial (pecuniary) externalities in this economy. If $j \in N$ extracts less, the price increases and this increase affects every districts' payoffs. The externality can be positive or negative. Since $i \in N \backslash j$ maximizes (2), the envelope theorem gives:

$$\frac{\partial u_i \left(x_i, x_{-i}\right)}{\partial \left(-x_j\right)} = a \left[\left(b+c\right) x_i - cX_i\right]. \tag{6}$$

If x_i is large, it is beneficial for district i that the price be high, and then i benefits when j extracts less. When i maximizes u_i by deciding on x_i, x_i is given by (12) and, combined with (6), we get the equilibrium level of externality:

$$\frac{\partial u_i \left(x_i^0, x_{-i}^0\right)}{\partial \left(-x_{-i}\right)} = \frac{e}{n+1}, \text{ where } e \equiv \left(b+c\right)\overline{p} - acX - v.$$

This equation shows that the equilibrium externality, as measured by e, increases in the market size $\overline{p}$ but decreases in the resource value v and in the aggregate stock, X. This is intuitive.[13] Further, if the benefit of profit, b, is large, it is valuable for i that the price be high, and then i benefits when j extracts less. If, instead, the enforcement cost c is large, it is more important to reduce the need to monitor and thus the pressure on the resource. In this case, the externality e is small and possibly negative.

[13] The intuition is as follows. If $\overline{p}$ is large or v is small, a district extracts more, and it becomes more important that the price is high. In this situation, a district benefits if the others extract less. If X_i increases, district i protects more, and it is more likely that district i is harmed by the larger enforcement expenditures when the others extract less. If the neighbor's stock is large, the neighbors extract more and district i finds it optimal to extract less. With more to protect, it is more likely that district i is harmed when j extracts less if the other stocks are large.

If *property rights are strong*, it is reasonable that the owner of a resource finds enforcement inexpensive and benefits from extraction. In our model, this corresponds to a small c, a large b, and thus a large externality e. If instead *property rights are weak*, enforcement is costly and a resource-owner might benefit less from extraction. This corresponds to a large c, a small b, and a small e.

The property rights, or the externality, e, will be a sufficient statistic for many of our results below. For example, the level of e will dictate how extraction levels are influenced by heterogeneity and the number of districts.

Proposition 2.
(i) Small districts extract smaller fractions of their resources if and only if property rights are weak:

$$\frac{x_i^0}{X_i} - \frac{x_j^0}{X_j} = \left(\frac{1}{X_i} - \frac{1}{X_j}\right) \frac{e}{a\,(b+c)\,(n+1)}.$$

(ii) If authority is decentralized, more is extracted if and only if property rights are strong:

$$\frac{\partial x^0}{\partial n} = \frac{e}{a\,(b+c)\,(n+1)^2}.$$

The Appendix proves that the general claims hold also when the demand function is nonlinear.

Part (i) of Proposition 2 suggests that the sign of e is important for a district's *strategy*. If $e > 0$, district i prefers a high price, and thus i has an incentive to keep the price high by strategically extracting less. If $e < 0$, district i has an incentive to extract more to reduce the price, and thus the pressure from illegal loggers. These strategic incentives are particularly important for a large district that influences the price more by a given change in x_i/X_i. The theory thus predicts that while large districts extract a *smaller* fraction of their resources when property rights are strong, they extract a *larger* fraction when property rights are weak.

Part (ii) follows as the natural next step in this line of reasoning. If multiple districts merge and centralize authority, the merged unit will be larger and it ought to increase conservation if and only if the externality is positive. The result holds whether it is only a couple of districts that centralize power to a common central authority, or whether all the n districts centralize power to a single government. Intuitively, the members of the merged unit will internalize the externalities on each other and thus extract less if and only if extraction is harmful to the partners. With strong property rights, it is well known from Cournot games that if the number of sellers increases, so does the aggregate quantity supplied, while the price declines. Proposition 2(ii) confirms this intuition. With weak property rights, in contrast, districts extract *more* when they take into account the fact that the pressure on the resource weakens as a consequence. In this case, the result is reversed, and centralization increases the amount of extraction.

4 Contracts on Conservation (REDD)

In this section we study contracts between the districts and a principal or a "donor" D. We assume that the donor's payoff is $U_D = u_D(x) - \tau$, where $\tau \geq 0$ is transfers and $u_D(x) = -dx$, so $d > 0$ measures the donor's marginal damage from aggregate extraction. The donor's payoff can equivalently be a function of the remaining stock $(X - x)$, and, as shown in the appendix, most of our results hold qualitatively if the damage function is nonlinear.[14]

Just like the donor does, we assume that every district $i \in N$ has a total payoff that is linear and additive in the transfer τ_i that i receives, so $U_i = u_i(x_i, x_{-i}) + \tau_i$, where u_i is given by (2), and $\tau \equiv \sum_{i \in N} \tau_i$, by budget balance.

Before the districts simultaneously choose the x_i's, the donor can unilaterally offer transfers (in a take-it-or-leave-it fashion) that are contingent on the entire vector of extraction levels. That is, the donor commits to transfer functions $\tau_i(\mathbf{x})$ for every $i \in N$. Since everyone can calculate the response of the other districts, it suffices for the donor to suggest a carefully chosen vector $\mathbf{x}_M^* = \{x_i^*\}_{i \in M}$ and to pay a district if and only if $x_i = x_i^*$.

Another simple contract is the linear version. This is the contract actually observed in reality, as when Norway offers REDD contracts to the partnering countries, and it should thus be of particular interest. In this case, the donor commits to pay a district an amount that is linear in the district's choice of x_i:

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

Here, $\overline{x}_i$ is the "baseline" or "reference" level for district i's deforestation level. The contract, which consists of the pair $(t_i, \overline{x}_i)$, implies that district i receives t_i dollars for every unit by which actual extraction x_i is reduced relative to the baseline level $\overline{x}_i$. If $x_i \geq \overline{x}_i$, no payment takes place. The contract is valid for country i regardless of what the other districts decide to do. When discussing linear contracts, we will assume that a district cannot commit to decline

[14] If the donor's payoff was $\widetilde{U}_D = \widetilde{u}_D(X - x) - \tau$, the analysis would be equivalent if we defined $u_D(x) \equiv \widetilde{u}_D(X - x) - \widetilde{u}_D(X)$, since X is a constant. A general nonlinear function $u_D(x)$ can also account for the consumer surplus, which is $ax^2/2$ when demand is linear. Thus, when we simplify to $u_D(x) = -dx$, we ignore the possibility that the donor may value the consumer surplus. This assumption is quite realistic, in our view: after all, the donor should *not* be regarded as a benevolent planner in our positive theory, but rather as an NGO or a single country offering REDD contracts, such as Norway.

payments in the future.[15]

Although we impose the limited-liability assumption that τ_i cannot be strictly negative, it may, in reality, be possible for the donor to penalize a district if it extracts more than what the donor has requested. For example, some tropical countries may receive development aid and this aid can be withheld. In the following, we do allow for this, and we let $f_i \geq 0$ measure how much the donor can credibly withhold or punish if i does not conserve as requested.

4.1 Contracts under Centralization

To build intuition for our results, it is helpful to first study the particularly simple setting in which authority is centralized to a central government, C. In this case, the objective of the donor is to maximize

$$U_D = -dx_C^* - \tau\left(x_C^*\right), \tag{7}$$

subject to the requirement that x_C^* must be a best response for the government given the contract. That is, extracting x_C^* and receiving $\tau\left(x_C^*\right)$ must be weakly better than any other option the government may have:

$$u_C\left(x_C^*\right) + \tau\left(x_C^*\right) \geq \max_{\widehat{x}} u_C\left(\widehat{x}\right) - f_C,$$

where $u_C\left(x\right) = bp\left(x\right)x + \left(v - cp\left(x\right)\right)\left(X - x\right)$, following equation (2). The right-hand side of (IC_C) measures the government's best outside option, that is, the utility it can obtain by freely choosing x without receiving transfers or aid ($f_C = \sum_{i \in N} f_i$).

Substituting a binding (IC_C) into (7), the donor's problem is to maximize:

$$U_D = -dx_C^* + u_C\left(x_C^*\right) - \max_{\widehat{x}} u_C\left(\widehat{x}\right) + f_C. \tag{8}$$

Thus, the donor maximizes the sum of payoffs and implements the first best. The first best is given by (4) if just v is replaced by $v + d$; so it follows that the first best is implemented by a linear contract with $t_C = d$, if $\overline{x}_C$ is so large that (IC_C) holds. By reducing $\overline{x}_C$ until (IC_C) binds, the donor extracts the entire surplus even with linear contracts, and a linear contract is therefore sufficient.

[15]This assumption is not very restrictive: In some earlier notes, when contracts had to be linear, we allowed districts to commit to *decline* any future payments. This condition resulted in a "participation constraint" that, in some situations, became harder to satisfy than the "incentive constraints," discussed below. Since the results were otherwise qualitatively similar, and since it may be questionable whether districts in reality are able to commit to decline future payments, we here simplify the analysis by not allowing the districts to commit in this way. With non-linear contracts, it is in any case straightforward to relax the participation constraint, since the donor can design contracts such that *if* one district rejects the offer, then it will be in the interest of the other districts to select x_i's at levels that would harm the district rejecting the contract.

Proposition 3. *Suppose the donor contracts with a single central government.*

(i) The equilibrium contract leads to the first best:

$$x^* = \frac{(b+c)\,\overline{p} + caX - v - d}{2a\,(b+c)}, \text{ and}$$
$$\tau^* = \frac{d^2}{4a\,(b+c)} - f_C.$$

(ii) This outcome can be implemented by the linear contract:

$$t_C^* = d, \text{ and}$$
$$\overline{x}_C^* = x_C^0 - \frac{d}{4a\,(b+c)} - \frac{f_C}{d}.$$

Naturally, x^* decreases in d, while the transfer must increase. The linear contract is particularly simple as it is similar to a Pigou subsidy.

The baseline $\overline{x}_C$ will be set such that (IC_C) binds and the government is exactly indifferent between choosing x_C^* and ignoring the contract. Note that $\overline{x}_C^* < x_C^0$, so that it is not sufficient to extract marginally less than the business-as-usual quantity in order to receive transfers. If we had $\overline{x}_C^* = x_C^0$, the central government would have been strictly better off with than without the contract, and thus the donor could reduce the transfer without violating the incentive constraint. This result disproves the typical presumption that the reference level should equal the business-as-usual level.[16]

4.2 Contracts under Decentralization

If the donor contracts with n independent districts, the objective is to maximize

$$U_D = -d \sum_{i \in N} x_i^* - \sum_{i \in N} \tau_i\left(\mathbf{x}^*\right), \tag{9}$$

subject to the requirement that $\mathbf{x}^*$ must be implementable and thus the vector of equilibrium extraction given the contracts $\{\tau_i\left(\mathbf{x}\right)\}_{i \in M}$. That is, the donor has to make sure that every district i's incentive constraint holds:

$$u_i\left(x_i^*, x_{-i}^*\right) + \tau_i\left(\mathbf{x}^*\right) \geq \max_{\widehat{x}_i} u_i\left(\widehat{x}_i, x_{-i}^*\right) - f_i.$$

[16] See, for example, **?** or **?**. The latter contribution also discusses why the baseline level may be smaller than the business-as-usual (or historical) deforestation level, since a smaller baseline reduces the amount that needs to be paid.

Here, the outside option at the right-hand side measures the utility i can obtain by freely choosing x_i under the assumption that the other districts will continue to extract their equilibrium quantities.[17]

If we substitute binding (IC$_i$)'s into (9), it is clear that the donor will no longer maximize the sum of payoffs. The reason is that the contracts with one district will influence the outside option for the other districts. The better the outside option is, the more the donor will have to pay. Thus, the donor prefers to design contracts that reduce the payoffs districts can get if they ignore the contract.

Proposition 4. *Suppose the donor contracts with all districts independently.*
(i) The contracts can be written as:

$$\begin{aligned}
\tau_i^* &= \frac{d^2}{a\left(b+c\right)\left(n+1\right)^2} - f_i, \textit{ and} \\
x_i^* &= \frac{\left(b+c\right)\overline{p} + ca\left[\left(n+1\right)X_i - X\right] - v}{a\left(b+c\right)\left(n+1\right)} - \frac{2d}{a\left(b+c\right)\left(n+1\right)^2}, \textit{ implying} \\
x^* &= \frac{n\left(b+c\right)\overline{p} + caX - nv}{a\left(b+c\right)\left(n+1\right)} - \frac{2nd}{a\left(b+c\right)\left(n+1\right)^2}.
\end{aligned}$$

(ii) This outcome can be implemented by the linear contract:

$$\begin{aligned}
t_i^* &= \frac{2d}{n+1}, \text{ and} \\
\overline{x}_i^* &= x_i^0 + \frac{n-3}{4a\left(b+c\right)\left(n+1\right)}t_i^* - f_i/t_i^*.
\end{aligned}$$

(iii) Compared to centralization, $\sum_{D\cup N} U_i$ *is smaller when* $n > 1$*, but also* x *is smaller if and only if:*

$$\frac{e}{d} < -\frac{n-1}{n+1},$$

and U_D *is larger if and only if:*

$$\frac{e}{d} < -\frac{1}{2}\frac{n-1}{n+1}. \tag{10}$$

Part (i) of the proposition shows that a larger d reduces the extraction levels. However, the reduction is small and approaches zero when n grows. The reason is leakage: when one

[17] This is the natural outside option as long as the x_i's are chosen simultaneously. Note that even if the x_i's were gradually increasing over time, the x_i's would effectively be chosen simultaneously if it were difficult for districts to observe the extraction levels at every point in time, before the contracting period has ended. However, if the model were dynamic and extraction levels immediately observable by the neighbors, then it might be easier for a district, $j \in N$, to detect a deviation by another district $i \in N \backslash j$. In this situation, it may be simpler for the donor to reduce i's temptation to deviate, since the contract with j may motivate j to choose a level of x_j that penalizes i when i is on a track to deviate. A satisfactory analysis of this situation requires another model than the one we investigate here.

district extracts less, the other districts prefer to extract more. Thus, when the donor pays one district to extract less, it also has to pay more to all the other $n-1$ districts for any given extraction vector. This expense reduces the donor's willingness to pay when n is large.

Part (ii) complements part (i) by showing that the linear subsidy rate falls when n grows. Linear contracts are sufficient, it turns out, since there is a deterministic and one-to-one relationship between the x_i's and the t_i's, and since the donor must, in any case, ensure that transfers to district i be so large that i cannot achieve a higher payoff by selecting any other x_i, when i takes x_{-i} as given. Of course, linear contracts would not suffice in more general environments with uncertainty or non-concave utility functions, for example.

The subsidy rate t_i^* is robust and remains unchanged if the parameters of the model either change or are unobservable to the donor. This robustness may be one reason for why the linear contract is popular in reality.[18] Note also that the f_i's do not affect the equilibrium allocation of the x_i^*'s; they only reduce the transfer which the donor has to pay.

As under centralization, the baseline $\overline{x}_i$ will be set such that (IC$_i$) binds and district i is exactly indifferent between choosing x_i^* and ignoring the contract. In contrast to centralization, however, the baseline might need to be larger than the business-as-usual level, x_i^0. The reason is that when n is large, the donor is paying so many districts to conserve, and the equilibrium price is so high, that the districts are much more tempted to extract than they would be without any contracts. In this case, the donor must increase the baseline to motivate the districts to conserve.

Part (iii) of Proposition 4 states that decentralization leads to less extraction when property rights are weak. This finding is similar to Proposition 2(ii), but the intuition is different: When the donor pays districts to extract less, the contracts create a negative externality on the other districts when e is small. In particular, there is a negative externality on the other districts' *outside option* which is not internalized by the donor. Instead, the donor benefits when the districts' outside option is worsened, and it therefore asks the districts to extract less than what is socially optimal when property rights are weak.

Part (iii) also shows that the donor benefits when n is large, if just property rights are weak so that the contract with one worsens the outside options of the others. In fact, the donor may benefit from decentralization even when decentralization may increase extraction, since the two thresholds for e/d are not identical: when $e/d \in \left(-\left(n-1\right)/\left(n+1\right), -\left(n-1\right)/2\left(n+1\right)\right)$, the donor benefits from decentralization even though it leads to more extraction.

[18] However, while the equilibrium choices of t_i's are independent of a, b, and c, the baseline level $\overline{x}_i$ should vary with these parameters if the donor wants to ensure that the expenditures be minimized. Note also that it is well known that simple, linear contracts can be optimal in dynamic settings with stochastic shocks being realized over time (**?**).

However, the sum of payoffs is always smaller when $n > 1$, since the first best is implemented when $n = 1$, according to Proposition 3. Thus, if the donor benefits from $n > 1$, it follows that the districts must be worse off.

Part (iii) is important because, 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.[19] Equation (10) shows that the donor benefits from local contracts if and only if property rights are weak.

Interestingly, the threshold for e/d decreases in n. Thus, while the donor may prefer decentralized contracts when the number of districts is relatively small, it might prefer centralized contracts if the alternative would be a very large set of districts to deal with. It is easy to show that the donor would prefer a marginally larger n if and only if:

$$n < n^* = \frac{1 - e/d}{1 + e/d}.$$

The smaller $e/d < 0$ is, the larger is the n^* maximizing the donor's payoff.[20]

Given that we have derived the equilibrium levels of transfers and extractions, it is easy to calculate the donor's payoff for any given n:

$$U_D = -d \frac{n\overline{p}(b+c) + acX - nv - dn/(n+1)}{a(b+c)(n+1)} + \sum_{i \in N} f_i.$$

This expression can be used to derive the donor's preference for n, but also for several other changes. First, given n, the donor benefits if the demand for the resource (measured by $\overline{p}$) is low. This benefit could motivate the donor to support a boycott against the extracted products. Second, note that the donor's payoff increases in v and decreases in c, but the effect of b is ambiguous. This ambiguity is related to the discussion at the end of Section 3 (and equation (5)), where we noted that x increases in b if $c < nv/aX$. This increase is harmful for the donor. In addition, a larger b makes it more expensive to persuade the districts to reduce the x_i's: thus, the condition for when the donor is harmed by a larger b is weaker than (5) and given by $c < nv/aX + dn/aX(n+1)$. In other words, the donor benefits from strengthening property rights insofar as such strengthening means that the protection cost (c) is reduced or that districts' conservation valuation (v) is increased, but *not* necessarily when the districts

[19] If a central government is already active and regulating local governments, it can always undo the donor's offers to the districts; decentralized contracts would then not be an option for the donor. If the central government is absent or passive, however, the donor 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 that central authorities be activated or created.

[20] While Proposition 4(iii) compares $n = 1$ with $n > 1$, we can alternatively consider a marginal increase in n and state that x decreases in n if and only if $e/d < -2(n-1)/(n+1)$, while U_D increases in n if and only if $e/d < -(n-1)/(n+1)$.

will be able to capture a larger fraction (b) of the revenues. In fact, the larger the districts' conservation value v, or the smaller their protection cost c, the smaller is the likelihood that the donor benefits from a large b.

4.3 The Number and Value of Contracts

So far, we have assumed that the donor can either contract with no one (Section 3) or with everyone. However, even if the donor would like to contract with all districts, doing so may be unfeasible for exogenous (or political) reasons. In this subsection, we thus assume that the donor can contract with only a subset $M \subseteq N$ of $m = |M| \leq n$ districts. Clearly, the problem of leakage is larger when m is small: if the donor pays some districts to extract less, the $n - m$ other district will find it optimal to increase extraction. This increase will crowd out the donor's effort.

While we discuss the effect of m below, note that most of the insight discussed already generalizes to our new case. In fact, the following proposition is analogous to Proposition 4.

Proposition 5. *Suppose the donor contracts with $m \leq n$ of the districts.*
(i) Each contract can be written as:

$$\begin{aligned}
\tau_i^* &= \frac{d^2}{a(b+c)(n+1)^2} - f_i, \text{ and} \\
x_i^* &= \frac{(b+c)\overline{p} + ca[(n+1)X_i - X] - v}{a(b+c)(n+1)} - \frac{2d(n+1-m)}{a(b+c)(n+1)^2}, \text{ implying} \\
x^* &= \frac{n(b+c)\overline{p} + caX - nv}{a(b+c)(n+1)} - \frac{2md}{a(b+c)(n+1)^2}.
\end{aligned}$$

(ii) This outcome can be implemented by the linear contract:

$$\begin{aligned}
t_i^* &= \frac{2d}{n+1}, \text{ and} \\
\overline{x}_i^* &= x_i^0 + \frac{4m - 3(n+1)}{4a(b+c)(n+1)} t_i^* - f_i/t_i^*.
\end{aligned}$$

(iii) If m increases, x decreases and U_D increases, but $\sum_{D \cup N} U_i$ decreases if and only if:

$$\frac{e}{d} < -\frac{n-1}{n+1} - 4\frac{n-m}{n^2-1}. \tag{11}$$

Parts (i) and (ii) of the proposition generalize the similar parts of Proposition 4. Naturally, the total extraction level is smaller if m is large.

Part (iii) also states that the donor prefers m to be as large as possible. This result is intuitive, since the donor can always decide to offer nothing to some districts.

The final part of the proposition is therefore most interesting: a larger m can reduce the sum of payoffs. In other words, the donor's contracts with the districts may be harmful for efficiency. The explanation for this is the possibly negative contractual externality. When property rights are weak, one district is harmed when the other districts extract less, as when they are offered conservation contracts by the donor. This negative externality may outweigh the donor's benefit from the contracts, particularly when the donor's damage is relatively small.

Another interpretation of the result is that the contracts may worsen an already existing collective action problem between the districts: When property rights are weak, districts are protecting too much, because they do not internalize the larger enforcement costs on the others. Conservation contracts will reduce extraction even further, and thus they also reduce the sum of payoffs.

Interestingly, there may be a socially optimal number of contracts, m^*. The threshold in (11) depends on m, and the inequality can be rewritten as:

$$m > m^* \equiv \frac{e}{4d}\left(n^2 - 1\right) + \frac{1}{4}\left(n + 1\right)^2.$$

Thus, when $m^* \in (1, n)$, it is socially optimal to increase m up to m^*, but not further. The reason is that when $e/d < 0$ is not that far from zero, business-as-usual leads to too much extraction, but contracting with everyone leads to too little. The first best is then implemented when the donor contracts with a subset only. The subset should be smaller when $e < 0$ falls or d is small: if property rights are weak, it is efficient to contract with fewer districts.

As mentioned, the donor always prefers to increase m, even when doing so reduces the sum of payoffs and thus the districts' utilities. Thus, when property rights are weak, the districts are playing a prisoner dilemma with each other regarding conservation. Each district benefits from reducing extraction when faced with the donor's contract, even though they could all be better off if everyone ignored the donor's contract. However, there is no alternative equilibrium where multiple districts ignore the contract by extracting more simultaneously. The equilibrium is unique because the extraction levels are strategic substitutes, as we noted in Section 3: a larger x_{-i} reduces i's temptation to extract more than x_i^*.

Theoretically, the result that conservation contracts can be harmful is interesting. In practice, however, one may question whether this result would survive in a more general model where we relax the above assumptions. In particular, by assuming perfect knowledge and observability, we allowed the donor to reduce the transfers to the knife-edge case in which each district was exactly indifferent between ignoring and adhering to the contract. If the donor were instead uncertain about some of the parameters of the model, it would be optimal to increase the transfers somewhat in order to make the contract robust to a larger set of parameters. The increase in transfers will naturally reduce the likelihood that the districts are worse off with conservation contracts than without. To investigate these possibilities further, it is thus important to allow for imperfect information and observability in future research.

5 Empirical Predictions

The framework above can be used to study various types of resources and alternative drivers of extractions, but it is motivated in particular by deforestation in the tropics. The model allows for many districts and recognizes that since extracting some of the resource increases the harvest supply, the price declines and so does the monitoring cost for the part that is to be conserved. The externality from one district's conservation effort on others is thus negative if property rights are weak, and positive if they are strong. The sign and the level of this externality determine the effects of decentralization as well as the design of the optimal conservation contract, according to the theory.

To be specific, Proposition 2 predicts that decentralizing authority will reduce deforestation when property rights are weak but increase deforestation when property rights are strong. As discussed in the Introduction, decentralization has been associated with more deforestation in Indonesia but less in the Himalayas. This difference is consistent with our theory if the enforcement cost is larger in the Himalayas, while deforestation is more sales-driven in Indonesia. Anecdotal evidence support this view: "Deforestation in Indonesia is largely driven by the expansion of profitable and legally sanctioned oil palm and timber plantations and logging operations" (**?**: 1328). In the Himalayas, in contrast, "the Forest Department was poorly staffed and thus unable to implement and enforce the national policies, and deforestation increased in the 1960s and 1970s" (**?**: 85).

Future research should carefully test the theoretical predictions. A serious test is beyond the scope of this note, since it will be challenging for several reasons. On the one hand, satellite data on deforestation is increasingly available, and one may also find country-specific data on the number of jurisdictions or administrative decentralization (as in **?**). On the other hand, exogenous changes in decentralization are hard to find. **?** take advantage of the fact that Indonesia had embarked on the decentralization process for a decade: using the differential timing of decentralization in different regions, they estimate the effect on deforestation of raising the number of district. This method can be used also when testing our predictions, but in addition one would need measures of the strength of property rights, or, more specifically, the *enforcement costs*. These enforcement costs should ideally be exogenous, as if they are identified by either geography (f.ex., the distance to the center) or the type of resource, since enforcement costs related to the political system may have changed endogenously as a response to deforestation or the number of jurisdictions. If enforcement costs are difficult to find, a proxy may be the fraction of resource extraction that is illegal: the fraction of resource extraction that is illegal is likely to be large if the enforcement cost is large. Data on illegal activities are naturally also hard to find, as discussed in Section 1.[21]

Proposition 2 further states that districts with a larger resource stock extract a larger fraction of the resource if and only if property rights are weak. The challenges discussed above must be overcome also when testing this prediction. An additional challenge for this prediction is that the variation in jurisdictional size may be endogenous, and the stock that remains is certainly depending on the historical amounts of extraction—unless one can identify more or less exogenous changes in jurisdictional borders.

The analysis of optimal contracts should be interpreted more normatively, in our view, since there are still too few conservation contracts observed in reality, and since there are many reasons for why these may not be optimally designed. In fact, a motivation for our analysis is that donors have little knowledge regarding how conservation contracts should be designed in the best way. That said, it is interesting to note that the various REDD contracts offered by Norway are all characterized by the same subsidy rate per unit of avoided deforestation.[22] This is in line with Propositions 4 and 5, where we showed that the optimal rate under linear contracts is independent of most of the parameters in the model. At the same time, these propositions also showed that the baseline levels—measuring the deforestation levels that must be met for any subsidy to be released—will be rather complicated functions of the parameters. In reality, the baseline levels do vary greatly between the countries, and they are negotiated individually before a contract is signed (**?**).

Perhaps our most interesting result is that the donor benefits from contracting with districts directly if the property rights are weak, but with a central authority if property rights are strong. In many cases, the donor might not be able to decide on the contractual level, as this might be determined by the national government. The preferences of the national

[21] As of 2016, we are not aware of larger data sets available on enforcement costs (rather than on state capacity, more generally) or decentralization of forest management authority (rather than on decentralization, more generally). Regarding the need for exogenous variation in property rights, one may apply similar methods as Acemoglu et al. (2001) who use colonial settlers' mortality rate to obtain exogenous variations in property rights across countries.

[22] The contracts are avaialble at `https://www.regjeringen.no/en/topics/climate-and-environment/climate/climate-and-forest-initiative/id2000712/`.

government can be very different than the desires of the donor, according to results above. In other cases, the donor is given a choice. As of 2016, Norway's policy on REDD is to sign agreements with national governments primarily; regional agreements have been declined. Such a policy is wise, according to our theory, if property rights are perceived to be strong and if one believes that illegal deforestation is not the main problem. If property rights are weak and deforestation is illegal, however, the donor would have been better off with decentralized contracts.

6 Final Remarks

These notes present a simple model of conservation and resource extraction that fits several types of exhaustible resources, and that can be applied whether property rights are strong and extraction is legal and sales-driven, or if instead property rights are weak and extraction is illegal and driven by the cost of protection. In the former case, we predict that decentralization increases extraction, equilibrium conservation contracts will permit too much extraction, and a donor will prefer to contract with a central authority rather than with local districts. In the latter case, when property rights are weak, all these results are reversed. The contrast between the results points out the importance of institutions, geography, and the type of resource that is considered.

Future research can build on this framework in a number of ways. The empirical predictions are testable and data is increasingly becoming available. On the theoretical side, we have abstracted from a number of important issues. This has made the model pedagogical, flexible, and tractable, and we believe it can and should be extended in several directions. By allowing for dynamic considerations, asymmetric information, and incomplete contracts in future analyses, we will deepen our understanding of conservation and how one can best design conservation contracts.

7 Appendix: Proofs

As the proofs below illustrates, our results continue to hold, qualitatively, if the damage function $d(x)$ and the demand function $p(x)$ are nonlinear (NL). Thus, a linear demand function is only a special case, here referred to as $p_L(x) = \overline{p} - ax$. To illustrate this robustness, we start by generalizing Proposition 1. The proofs also allow for heterogeneous v's.

Proposition 1^{NL}. *If c or X_i increases, or v decreases, then x_i increases, x increases, and $p(x)$ decreases. Furthermore, x_j decreases in X_i, $j \neq i$. In the unique equilibrium, the equilibrium x_i's are implicitly given by $x_i = x_i^0$, where:*

$$x_i^0 = \frac{c}{b+c}X_i + \frac{(b+c)\,p(x^0) - v}{-p'(x^0)(b+c)} \in [0, X_i] \text{ and} \tag{12}$$

$$x^0 = \frac{c}{b+c}X + n\frac{(b+c)\,p(x^0) - v}{-p'(x^0)(b+c)} \tag{13}$$

if each stock is sufficiently large and the second-order condition is satisfied:

$$X_i \geq \max\left\{\frac{(b+c)\,p(x^0) - v}{-p'(x^0)\,b}, \frac{v-(b+c)\,p(x^0)}{-p'(x^0)\,c}\right\}; \tag{14}$$

$$\left(-p'(x^0)\right)^2 > \frac{p''(x^0)\left[(b+c)\,p(x^0) - v\right]}{2(b+c)}.$$

Proof of Proposition 1^{NL}. The proofs allow for heterogeneous v_i's, so $u_i(x_i, x_{-i}) = bp(x)\,x_i + (v_i - cp(x))(X_i - x_i)$. The derivative of $u_i(x_i, x_{-i})$ with respect to (w.r.t.) x_i is:

$$(b+c)(p'(x)\,x_i + p(x)) - cp'(x)\,X_i - v_i. \tag{15}$$

Thus, i extracts nothing if the derivative is negative even when $x_i = 0$, which requires:

$$(b+c)\,p(x) - cp'(x)\,X_i - v_i \leq 0 \Rightarrow X_i \leq \frac{v_i - (b+c)\,p(x)}{-p'(x)\,c},$$

while i extracts everything if the derivative is positive also when $x_i = X_i$:

$$(b+c)(p'(x)\,X_i + p(x)) - cp'(x)\,X_i - v_i \geq 0 \Rightarrow X_i \leq \frac{(b+c)\,p(x) - v_i}{-p'(x)\,b}.$$

To ensure that $x_i \in (0, X_i)$, we must thus assume that each stock is sufficiently large:

$$X_i > \max\left\{\frac{v_i - (b+c)\,p(x)}{-p'(x)\,c}, \frac{(b+c)\,p(x) - v_i}{-p'(x)\,b}\right\}. \tag{16}$$

Furthermore, note that the second-order condition (s.o.c.) is satisfied when:

$$(b+c)(p''(x)\,x_i + 2p'(x)) - cp''(x)\,X_i < 0 \Rightarrow$$

$$-p'(x) > \frac{p''(x)\left[(b+c)\,x_i - cX_i\right]}{2(b+c)} = \frac{p''(x)\left[(b+c)\,p(x) - v_i\right]/(-p'(x))}{2} \Rightarrow$$

$$(-p'(x))^2 > \frac{p''(x)\left[(b+c)\,p(x) - v_i\right]}{2(b+c)}, \tag{17}$$

where we used (12). Assuming that s.o.c. is satisfied, (15) decreases in x_i. Since the first-order condition (f.o.c.) is that (15) equals zero,

$$\begin{aligned} x_i &= \frac{(b+c)\,p(x) - cp'(x)\,X_i - v_i}{-p'(x)\,(b+c)} = \frac{c}{b+c}X_i + \frac{(b+c)\,p(x) - v_i}{-p'(x)\,(b+c)} \Rightarrow \\ x &= \frac{c}{b+c}X + \frac{n\,(b+c)\,p(x) - \sum_{i\in N} v_i}{-p'(x)\,(b+c)}. \end{aligned} \tag{18}$$

We can see that if c or X_i increases, or $v_i > 0$ decreases, x must increase and $p(x)$ decrease. For similar reasons, x_i increases. ‖

Proof of Proposition 1. The s.o.c. (17) always holds with $p_L(x)$, since then $p''(x) = 0$. Proposition 1 follows from Proposition 1^{NL} if we substitute with $p_L(x) = \overline{p} - ax$ and solve for the x_i's. For example, with $p_L(x)$, (18) implies:

$$x = \frac{c}{b+c}\frac{X}{1+n} + \frac{n\,(b+c)\,\overline{p} - \sum_{i\in N} v_i}{a\,(b+c)\,(1+n)} \Rightarrow \tag{19}$$

$$\begin{aligned} x_i &= \frac{c}{b+c}X_i + \frac{(b+c)\,\overline{p} - v_i}{a\,(b+c)} - x \\ &= \frac{c}{b+c}X_i + \frac{(b+c)\,\overline{p} - v_i}{a\,(b+c)} - \frac{c}{b+c}\frac{X}{1+n} - \frac{n\,(b+c)\,\overline{p} - \sum_{i\in N} v_i}{a\,(b+c)\,(1+n)} \\ &= \frac{c}{b+c}X_i + \frac{(b+c)\,\overline{p} - acX - (1+n)\,v_i + \sum_{i\in N} v_i}{a\,(b+c)\,(1+n)}. \end{aligned} \tag{20}$$

The threshold on size, (16), is derived in a similar way. Setting $v_i = v$ completes the proof. ‖

Remark 3: On the externality. Let $p(x)$ be nonlinear. Since $i \in N \backslash j$ maximizes (2), the envelope theorem gives:

$$\frac{\partial u_i(x_i, x_{-i})}{\partial(-x_j)} = -p'(x)\,[(b+c)\,x_i - cX_i]\,. \tag{21}$$

When i maximizes u_i by deciding on x_i, x_i is given by (12) and, combined with (21), we get the equilibrium level of externality:

$$\begin{aligned} \frac{\partial u_i\left(x_i^0, x_{-i}^0\right)}{\partial(-x_{-i})} &= \frac{e}{n+1}, \text{ where } e \equiv \left[(b+c)\,p\left(x^0\right) - v\right](n+1) \Rightarrow \\ e &= (b+c)\,\overline{p} - acX - v \quad \text{when } p(x) = p_L(x)\,. \end{aligned}$$

The following lemma is proven in the previous version of this note:

Lemma 1^{NL}. *The equilibrium externality e from reducing extraction decreases in X and increases in b. Further, e is smaller if the districts are weak and c is large, as long as:*[23]

$$\left(-p'\left(x^0\right)\right)^2 > \frac{-nep''\left(x^0\right)/\left(n+1\right)(b+c)}{-p'\left(x^0\right)\left(X - x^0\right)/p\left(x^0\right) - 1} \text{ and } X > x + \frac{p(x)}{-p'(x)}. \tag{22}$$

We next state a generalization of Proposition 2 to the case with nonlinear demand function.

Proposition 2^{NL}. *(i) Small districts extract smaller fractions of their resources if and only*

[23] The first condition in (22) is satisfied if either e or $p''\left(x^0\right)$ is relatively close to zero (as for $p_L(x)$). The second condition requires that the demand curve be so steep at x^0 that a linear approximation would have led to a negative price if the entire stock X were sold on the market: with $p_L(x)$, it reduces to $\overline{p} - aX < 0$.

if $e < 0$*. For* $X_i < X_j$*, we have*

$$\frac{x_i^0}{X_i} < \frac{x_j^0}{X_j} \Leftrightarrow e < 0.$$

(ii) If n *increases, more is conserved if and only if* $e < 0$*:*

$$\frac{\partial x^0}{\partial n} < 0 \Leftrightarrow e < 0.$$

Proof of Proposition 2^{NL}**.**
(i) If we divide both sides of (18) by X_i, we get:

$$\frac{x_i}{X_i} = \frac{c}{b+c} + \frac{1}{X_i}\frac{(b+c)\,p(x) - v_i}{(b+c)(-p'(x))} \Rightarrow \frac{x_i}{X_i} - \frac{x_j}{X_j} = \left(\frac{1}{X_i} - \frac{1}{X_j}\right)\frac{(b+c)\,p(x) - v}{(b+c)(-p'(x))} \text{ if } v_i = v_j = v. \tag{23}$$

(ii) The proof follows, for example, when we summarize the x_i's as given by (12) and substitute in for e: the l.h.s. must increase in x faster than the r.h.s. does for (17) to hold; and x must therefore increase in n if and only if $(b+c)\,p(x) - v > 0$ when $v_i = v_j = v$. $\|$

Proof of Proposition 2. Part (i) follows straightforwardly when substituting in for $p_L(x)$ into (23). Part (ii) follows when differentiating x^0, as expressed in Proposition 1, w.r.t. n. $\|$

We now state and prove that linear contracts are sufficient, also for the case in which the demand function and the damage function are nonlinear.

Lemma 2^{NL}**.** *Linear contracts are sufficient: the outcome of the donor's preferred general contract coincides with the outcome of the donors preferred linear contract.*

Proof of Lemma 2^{NL}**.** If the contract is linear, i's payoff is

$$u_i(x_i, x_{-i}) + \max\{0, t_i(\overline{x}_i - x_i)\} = bpx_i + (v_i - cp)(X_i - x_i) + \max\{0, t_i(\overline{x}_i - x_i)\}, \tag{24}$$

so, when $x_i < \overline{x}_i$, the f.o.c. w.r.t. x_i is given by (18) where v_i is replaced by $v_i + t_i$:

$$x_i = \frac{(b+c)\,p(x) - cp'(x)\,X_i - v_i - t_i}{-p'(x)(b+c)}. \tag{25}$$

The s.o.c. is locally satisfied under the same condition (17) as before.

To see that linear contracts suffice, note, first, that for $\mathbf{x} = \{x_i\}_{i\in N}$ to be implemented by a contract, x_i must be a best response to x_{-i} for every $i \in N\backslash M$. Then, given x and any x_i, $i \in M$, one can solve (25) for t_i to find the subsidy rate which makes x_i optimal for i. Hence, if $\{x_i\}_{i\in N}$ is implementable, it is implementable by linear contracts.

However, the s.o.c. may not hold globally because $\max\{0, t_i(\overline{x}_i - x_i)\}$ is kinked (and convex) at $x_i = \overline{x}_i$. Thus, (25) is a best response for i if and only if the corresponding payoff (24) is larger than what i can achieve by choosing any other $x_i > \overline{x}_i$. The incentive constraint is:

$$\begin{aligned} u_i(x_i, x_{-i}) + \max\{0, t_i(\overline{x}_i - x_i)\} &\geq \max_{\widehat{x}_i} u_i(\widehat{x}_i, x_{-i}) - f_i \Rightarrow \\ \max\{0, t_i(\overline{x}_i - x_i)\} &\geq \max_{\widehat{x}_i} u_i(\widehat{x}_i, x_{-i}) - u_i(x_i, x_{-i}) - f_i. \end{aligned}$$

Clearly, the donor prefers to reduce every $\overline{x}_i$ until this condition binds. By substituting the corresponding equation into the donor's payoff, $-d(x) - \sum_{i\in M} \max\{0, t_i(\overline{x}_i - x_i)\}$, we get that this payoff is exactly the same as for general contracts:

$$-d(x)+\sum_{i\in M}u_i(x_i,x_{-i})-\sum_{i\in M}\max_{\widehat{x}_i}u_i(\widehat{x}_i,x_{-i})+\sum_{i\in M}f_i. \tag{26}$$

Hence, linear contracts are sufficient and they leads to the same quantities of extractions and transfers (and thus payoff to the donor). ‖

Proof of Proposition 3. When $i=C$ is the only district, (26) boils down to:

$$-d(x)+u_C(x)-\max_{\widehat{x}}u_C(\widehat{x})+f_C.$$

Since C's outside option is independent of the contract, $x_C=\arg\max_x -d(x)+u_C(x)$, and x_C is thus first best. From (25), this requires $t_C=d'(x_C)$. The rest of the proposition follows as a special case from Proposition 5. ‖

Proof of Proposition 4. Part (i) and (ii) follow from the more general proof of Proposition 5(i)-(ii): when setting $m=n$ in Proposition 5(i)-(ii), we get Proposition 4(i)-(ii). Part (iii) follows if we set $m=n$ and compare x and u_D for the case in which $n=1$ to the case in which $n>1$. ‖

Proof of Proposition 5.
(i)–(ii): Consider linear contracts. Given the linear $p_L(x)$, we can rewrite (25) by solving for x_i as a function of x_{-i} to get:

$$x_i=\frac{\overline{p}-ax_{-i}}{2a}+\frac{caX_i-v_i-t_i}{2a(b+c)}. \tag{27}$$

It follows that i's optimal response to x_{-i} if i decided to ignore the contract (by not collecting t_i) is:

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

where the price would be lowered from $p=\overline{p}-ax$ to:

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

Thus, if $u_i(x_i,x_{-i})=bp(x)x_i+(v_i-cp(x))(X_i-x_i)$ is i's payoff, minus i's transfer, when i acts according to the contract, then i's payoff from deviating can be written as follows:

$$\begin{aligned}
& u_i\left(x_i^I,x_{-i}\right)=\left[(b+c)p^I-v_i\right]x_i^I+\left(v_i-p^Ic\right)X_i\\
= & \left[(b+c)\left(p-\frac{t_i}{2(b+c)}\right)-v_i\right]\left(x_i+\frac{t_i}{2a(b+c)}\right)+\left(v_i-cp+\frac{ct_i}{2(b+c)}\right)X_i\\
= & u_i(x_i,x_{-i})+\left[(b+c)\left(p-\frac{t_i}{2(b+c)}\right)-v\right]\frac{t_i}{2a(b+c)}-\frac{t_i}{2}x_i+\frac{cX_it_i}{2(b+c)}\\
= & u_i(x_i,x_{-i})+\frac{t_i^2}{4a(b+c)}.
\end{aligned} \tag{28}$$

Thus, the transfer $\tau_i=t_i(\overline{x}_i-x_i)$ plus f_i must be at least $t_i^2/4a(b+c)$ for this deviation to be unattractive to i. When this (IC$_i$) binds, it becomes:

$$t_i(\overline{x}_i-x_i)=u_i\left(x_i^I,x_{-i}\right)-u_i(\mathbf{x})-f_i=\frac{t_i^2}{4a(b+c)}\Rightarrow\overline{x}_i=x_i+\frac{t_i}{4a(b+c)}-f_i/t_i. \tag{29}$$

Given the linear $p_L(x)$, we can sum over the x_i's from (25) and write:

$$x = \frac{n(b+c)\overline{p} + caX - \sum_{i \in N}(v_i + t_i)}{a(b+c)(n+1)}. \tag{30}$$

Thus, the donor maximizes

$$-dx - \sum_{i \in M} \tau_i = -d\left[\frac{n(b+c)\overline{p} + caX - \sum_{i \in N} v_i - \sum_{i \in M} t_i}{a(b+c)(n+1)}\right] - \sum_{i \in M} \frac{t_i^2}{4a(b+c)}.$$

For each t_i, $i \in M$, the first-order condition becomes $t_i = 2d/(n+1)$. The second-order condition trivially holds, so the contracts are symmetric and equal for every $i \in M$. Given this t_i, we can substitute t_i into (30) to get x, substitute into (27) to get x_i, into (29) to find $\overline{x}_i$, and derive the total transfer from $\tau_i = t_i(\overline{x}_i - x_i)$.
(iii) By substituting $t_i = 2d/(n+1)$ into (19), we can see that x decreases in m:

$$x = \frac{n(b+c)\overline{p} + caX - \sum_{i \in N} v_i}{a(b+c)(n+1)} - d\frac{2m}{a(b+c)(n+1)^2}. \tag{31}$$

With x and τ_i, we can also easily write U_D as an increasing function of m:

$$\begin{aligned} U_D &= -d\left(\frac{n(b+c)\overline{p} + caX - \sum_{i \in N} v_i}{a(b+c)(n+1)} - d\frac{2m}{a(b+c)(n+1)^2}\right) - \frac{md^2}{a(b+c)(n+1)^2} + \sum_{i \in M} f_i \\ &= -\frac{ned}{a(b+c)(n+1)} - \frac{dcX}{b+c} + \frac{md^2}{a(b+c)(n+1)^2} + \sum_{i \in M} f_i. \end{aligned}$$

The total sum of payoffs for the donor and all $i \in N$ is $bp(x)x + (v - cp(x))(X - x) - dx$. Since x is given by (31), the derivative w.r.t. m is negative if and only if:

$$\begin{aligned} \left[(b+c)(\overline{p} - 2ax) + acX - v - d\right]\frac{-m2d/(n+1)}{a(b+c)(n+1)} &< 0 \Rightarrow \\ (b+c)(\overline{p} - 2ax) + acX &> v + d, \end{aligned}$$

which, when we substitute in for (31), can be rewritten as (11). ||

Nonlinear $p(x)$. To see how Propositions 3-5 generalize, let $\widehat{x}_i\left(x_{-i}^*\right) = \arg\max_{x_i} u_i\left(x_i, x_{-i}^*\right)$ be i's best outside option, and note that $\widehat{x}_i\left(x_{-i}^*\right)$ is implicitly defined by (12) when one takes into account that $x = x_i + x_{-i}^*$. Given the function $\widehat{x}_i\left(x_{-i}^*\right)$, we can define the equilibrium externality at the outside option:

$$\widehat{e}_i\left(x_{-i}^*\right) \equiv \frac{\partial u_i\left(\widehat{x}_i\left(x_{-i}^*\right), x_{-i}^*\right)}{\partial(-x_{-i})} = (b+c)\,p\left(\widehat{x}_i\left(x_{-i}^*\right) + x_{-i}^*\right) - v.$$

Let *decentralization* mean that n and m increase by the same number, keeping $n - m$ fixed. The following proposition is proven in the previous version of our note:

Proposition 4^{NL}. *Suppose the donor contracts with all $m \le n$ districts.*
(i) If $n = m = 1$, the outcome is first best.
(ii) If $m > 1$, x is too large relative to maximizing $\sum_{i \in D \cup M} u_i$ if $e_i\left(x_{-i}^*\right) > (<) 0 \forall i \in M$. However, x is too small, relative to maximizing $\sum_{i \in D \cup M} u_i$ if $e_i\left(x_{-i}^*\right) < 0 \forall i \in M$.
(iii) Decentralization increases equilibrium x^* if $e_i\left(x_{-i}^*\right) < 0 \forall i \in M$.
(iv) Decentralization increases the donor's equilibrium payoff if $e_i\left(x_{-i}^*\right) < 0 \forall i \in N$.