GROUPS IN CONFLICT:

Private and Public Prizes*

Laura Mayoral ⓡ Debraj Ray†

November 2020

Abstract. This paper studies costly conflict over private and public goods. Oil is an example of the former, political and civil rights an example of the latter. Our theory predicts that groups in conflict are likely to be small when the prize is private, and large when the prize is public. We examine these implications empirically by constructing a global dataset at the ethnic group level and studying conflict along ethnic lines. Our theoretical predictions find significant confirmation in this setting, and the analysis sheds new light on group size and collective action in the context of violent conflict.

1. Introduction

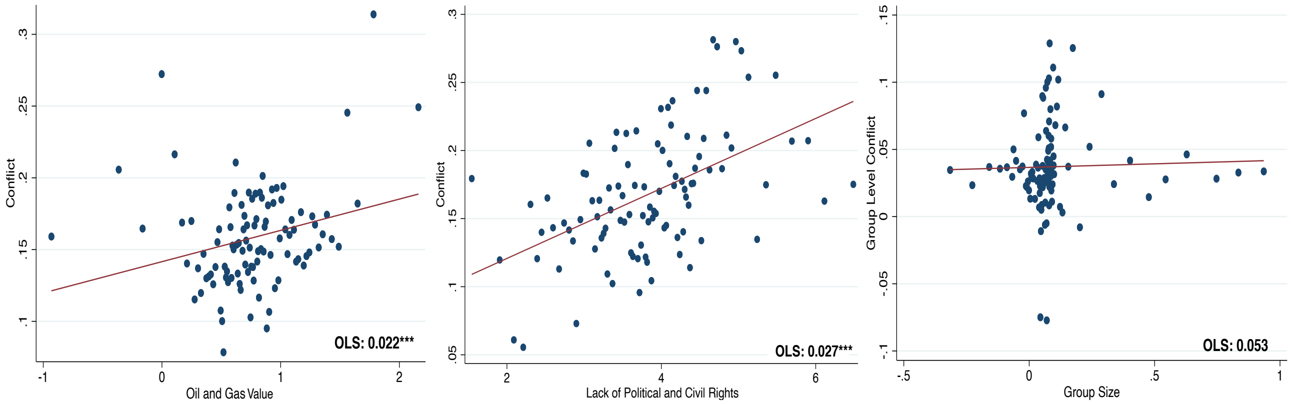
Figure 1 depicts two familiar facts and one not-so-familiar assertion about conflict, to be viewed as correlations rather than causal statements. Panel A displays a binned scatter plot of oil and gas endowments by country and that country's proclivity to conflict, as measured by the probability of armed internal conflict in any year, with per-capita income, population, and country and time fixed effects as controls. The verdict is well known: the presence of natural resources is correlated with the existence of conflict. This diagram would not change if we replaced oil by other resources such as diamonds, or by mineral endowments more generally. This correlation is robust enough to be considered a "stylized fact;" see, for instance, Le Billon (2001), Bannon and Collier (2003), Fearon (2005), Lujala (2010), Dube and Vargas (2013) and Morelli and Rohner (2015).

Panel B depicts a another well-known connection, this time between conflict and an index capturing the lack of political and civil rights. The same controls as in Panel A are deployed. Additionally, we control for oil and gas endowments, for we wish to avoid confounding this relationship with the obvious possibility that both conflict and the absence of rights are positively related to those endowments (and therefore to each other). As in Panel A, we have a statistically significant and positive relationship; another stylized fact.

Panel C switches gears to ethnic groups. It asks: what is the relationship between the size of an ethnic group and its proclivity to enter into conflict with the State in which that group is resident? The answer is that isn't one. This null finding is less familiar, in part because null findings excite less attention. We are all aware of the "tyranny of the majority" (see, e.g. Tocqueville 1835),

*We are grateful to Oeindrila Dube, William Easterly, Joan Esteban and Sahar Parsa for helpful comments on an earlier draft. Ray thanks the National Science Foundation for research support under grant numbers SES-1261560 and SES-1851758. Mayoral gratefully acknowledges financial support from the Generalitat de Catalunya, and the Ministry of Economy and Competitiveness Grant number ECO2015-66883-P.

†Mayoral: Instituto de Análisis Económico (CSIC), Barcelona GSE and University of Gothenburg; Ray: New York University and University of Warwick



(a) Conflict and Oil Reserves **(b)** Conflict and Lack of rights **(c)** Conflict and Group Size

Figure 1. Panel **(a)** shows a binned scatterplot relating civil conflict to the country-level values of oil and gas. Panel **(b)** does the same for civil conflict and an index capturing the lack of political and civil rights. Panel **(c)** shows a binned scatterplot relating civil conflict at the ethnic group level to group size. All graphs control for log GDP per capita, log population and country and year dummies; panel **(b)** also controls for the value of oil and gas. OLS coefficients are noted. *** denotes significance at the 1% level. Details of the construction of these graphs are in Section B.1 of the Online Appendix.

in which a larger group can more easily impose its will on society than its smaller counterpart. That phenomenon expresses itself most clearly under voting, but the suppression of minorities via extra-democratic channels, including coercive means, is also extremely common. And yet, there is a contrasting view, a sort of "tyranny of the minority," according to which small groups are more involved than large groups in lobbying or conflict, because they are more cohesive, and because the same prize goes a longer way *per capita* when groups are small (Pareto 1927, Olson 1965, Chamberlin 1974, McGuire 1974, Marwell and Oliver 1993, Oliver and Marwell 1988, Sandler 1992, Taylor 1987 and Esteban and Ray 2001). The cohesion argument appears to be first voiced in Olson (1965), while the per-capita prize argument was explicitly discussed in Pareto (1927, p.379):

"[A] protectionist measure provides large benefits to a small number of people, and causes a very great number of consumers a slight loss. This circumstance makes it easier to put a protection measure into practice."

Empirically, the two "tyrannies" appear to cancel — or at least, that is one way to read Figure 1(C). That is the starting point of our paper.

Our study focuses on groups that are defined along ethnic lines. Ethnic conflict is a natural choice for the study, as groups demarcated by ethnicity account for between 50–75% of internal conflicts since 1945 (Fearon and Laitin, 2003; Doyle and Sambanis, 2006). We construct a panel dataset at the ethnic group level with global coverage, with 145 countries and 1475 ethnic groups spanning 1960–2006. Our contributions are as follows. First, we show that the resource effect in Panel A of Figure 1 is particularly pronounced when those resources are located on the ethnic homelands of *smaller* groups. In our baseline specification we use oil abundance in the homeland as a proxy for resources. The same results hold for under different robustness checks, such as the presence of mines and minerals, and a larger per-capita land endowment. In line with the Pareto-Olson thesis, smaller groups are more willing to resist incursions by the State that challenge "private" endowments — where "private" is meant in the sense of appropriability by the group in question.

In sharp contrast, the effects of Panel B are particularly pronounced for conflicts that involve *larger* groups. In our baseline specification, we use a pre-sample measure of the *lack of political and civil rights*, constructed by Freedom House, which is a country-level index measuring the rights effectively enjoyed by the citizens. We also employ alternative proxies based on group exclusion from power, the absence or lack of democracy, or a paucity of basic public services. In each of these cases, the effect on conflict is more pronounced for larger, rather than smaller ethnic groups.

Put another way, Panel C of Figure 1 obscures two significant relationships by conflating them into one. By separating them, we gain insight not just into group size and conflict, but also into the effects described in Panels A and B.

Our core findings are backed up by robustness checks and ancillary results. The former include alternative conflict variables, estimation strategies and ways of proxying for the prizes at stake, both private and public. Ancillary results are discussed throughout. For instance, once the private prize and its interaction with group size are accounted for, the coefficient on group size turns positive and significant (it is insignificant if entered on its own). In parallel vein, if per-capita payoffs are held constant, large groups are always more prone to fighting, irrespective of the type of payoff at stake. (Empirically, the role of group size changes dramatically depending on whether one controls for total or per-capita private payoffs, in line with the theoretical predictions.) Finally, conflict over private payoffs is more likely as the number of potential threats to conflict increases.

The world is replete with examples of ethnic conflict over both private and group-public prizes; often mixtures of the two. The typical ethnic conflict could involve a struggle for political power or control (as in Burundi, Bosnia, Liberia, or Zimbabwe), but it can involve secessionist struggles by groups seeking to control their own land or resources (Chechnya, Kashmir, Tamils in Sri Lanka, the Casamance in Senegal, and many other examples). Land and oil are often central among these resources (e.g., the Ijaw conflict in Nigeria, the Darfur conflict, or the Second Civil War in the Sudan). In contrast, a public prize could represent democratic freedoms, untrammeled political power, the imposition of a secular or religious ideology, and it could even represent private gains that are relatively undiluted by the number of *group* recipients; e.g., public-sector jobs or government licenses first reserved for a favored ethnicity. The prize in question may be excludable for the population as a whole (in many cases it is), but as far as the *group* is concerned, it has the dominant features of a public good.

To our knowledge, the interactions between group size and different prizes have not been studied. As already mentioned, it is well known that the presence of natural resources — particularly oil — is correlated with conflict; see the references already cited. Morelli and Rohner (2015) show, additionally, that the concentration of those natural resources in ethnic homelands is related to conflict. But our focus is on the *interaction* between group size and the homeland resource variable, and not just the location of resources. (In addition, as already described, we are equally interested in the public payoff variable and its interaction with group size.) Perhaps the closest relationship is to our own work with Joan Esteban (Esteban and Ray 2011, Esteban, Mayoral and Ray 2012a,b), where we show that ethnic polarization is connected to public prizes, while ethnic fractionalization is connected to private prizes. However, these studies are carried out, not at the level of an ethnic group but at the level of the country. None of the interaction results studied here are addressed there, nor is the question of group size and conflict.

We precede the empirical analysis with a theory that is meant to illustrate these effects — not as some sort of hindsight contribution — but hopefully to be viewed as a natural formalization of these forces. (In fact, the theory was written down long before the empirical research was done.) In our setting, there are several groups that can challenge the payoff implications of some status quo allocation of public or private benefits. A challenge is followed by conflict, with the spoils accruing either to the group — in the case of victory — or to the State, in case the group is defeated. When the prize is private, its per-capita value is affected by group size. When the prize is public at the level of the group, its per-capita value is undiluted by group size. Our two theoretical results parallel the empirical findings to follow: 1. When the prize is private, there is a threshold size *below* which a group is more likely to challenge an equal status quo. 2. When the prize is group-public, there is a threshold size *above* which a group is more likely to challenge an equal status quo.

These two observations lead to the two double-interaction terms — SIZE × PRIVATE PRIZE and SIZE × PUBLIC PRIZE — that we carry through the entire empirical analysis.

The connections between group size and conflict shed light on the logic of collective action, and suggest a rethinking and synthesis of the disparate ideas in Tocqueville, Pareto and Olson, beyond the confines of a democracy to a setting that involves violence and conflict. Second, the effects of group size have implications for the broader question of which social structures are more conflictual. Are these societies with a diverse multitude of small groups, fractured by ethnicity, geography, religion and the like? Or are such societies likely to exhibit deep cleavages that pit one great group against another, or against the State? These are questions of the deepest importance in political economy. Our analysis suggests that they are answered, at least to some degree, by studying the nature of the prize at stake. And third, if the results are convincing, they could serve as indirect pointers "out of sample" to the root causes of a conflict, when these are obscured or ambiguous to the researcher, but the sizes and compositions of the groups involved are more easily observed.

Section 2 introduces a baseline model of conflict. Section 3.1 analyzes the relation between group size and conflict when conflict is over a private prize. Sections 3.2 does the same for public prizes. Our main empirical results are presented in Section 4. Section 5 contains additional variations that examine the robustness of the results. Section 6 concludes. See Appendix for detailed definitions of all variables as well as summary statistics. See Online Appendix for additional empirical results.

2. Theory

Denote by v the total “appropriable resources" of society. These could represent material resources such as oil from a particular geographical location, or public payoffs such as the acquisition of political or cultural power. Assume that the State (or society as a whole) seeks to allocate v widely over the entire community, perhaps according to the wishes of a dominant group. But there is another group, demarcated by ethnicity, geography, religion or occupation, which seeks to retain — or seize — the proceeds of v for itself. For instance, one might think of v as the value of oil reserves located within the homeland of an ethnic group. The State wants to distribute those revenues over the entire country, while the ethnic group might feel that this is “their oil."

The population is normalized to have unit mass. Let X be the set of efficient payoff allocations $\mathbf{x} = \{x(i)\}$ that are generated by v. A special, salient allocation of appropriable resources is the *non-discriminatory* allocation under which everyone receives an equal payoff. We assume that this allocation is feasible.

The “challenger group" in question — call it the Rebel — can accede to the peaceful allocation, or its members can engage in costly conflict. In the case of conflict, we suppose that society is partitioned into two subsets, one of size m (pertaining to the Rebel) and the remainder — the “State" — of size $\overline{m}$ ($m + \overline{m} = 1$), and that they engage in a bilateral conflict. We leave open the question of whether the Rebel “initiates" conflict or “defends itself" against incursions made by the State. That will depend on the situation at hand. For instance, if there is settlement on the Rebel's territory, conflict may be interpretable as defense against State aggression. If the Rebel is fighting to overthrow the State and seize power, then it may be viewed as the aggressor.

Conflict involves — on each side — the expending of effort or funds. The utility cost to an individual from a contribution of r is given by

$$c(r) = (1/\alpha)r^{\alpha}$$

for some $\alpha > 1$.[3] We assume that a leader on each side extracts “effort" from everyone to maximize the per-capita payoff of her coalition.[4] Because the cost of effort provision is strictly convex, the leader will ask for equal effort from each individual, and will make transfers if needed to compensate them. The winner — Rebel or State — obtains full control over the prize.

To map group efforts into win probabilities, we use contest success functions (Tullock 19xx, Skaperdas 1996), so the probability that the Rebel will win is given by

$$p = \frac{mr}{R}, \tag{1}$$

[3]Nothing hangs on the specific choice of cost function, though its strict convexity is important.

[4]To be sure, this neglects the free-rider problem within groups. It is easy to write down variants of our model in which individuals unilaterally provide effort, provided that they at least partially internalize the payoffs of their fellow group members (see Esteban and Ray 2011).

where r is contribution per person in the Rebel, and $R = mr + \overline{mr}$ is the sum of contributions made by both groups (use bars on variables for the State). So, if π stands for the per-capita payoff conditional on winning, and loss payoffs are normalized to zero, the Rebel seeks to maximize

$$\pi \frac{mr}{R} - c(r),$$

A similar problem confronts the State, with payoff $\overline{\pi}$ (or 0) conditional on winning (or losing). A Nash equilibrium of this "conflict game" is fully described by the first-order conditions:

$$\pi m \overline{m} = R^2 \frac{r^{\alpha-1}}{\overline{r}} \text{ and } \overline{\pi} m \overline{m} = R^2 \frac{\overline{r}^{\alpha-1}}{r} \tag{2}$$

for Rebel and State respectively. For the Rebel, rewrite its portion of (2) to observe that

$$r^{\alpha} = \pi p \overline{p},$$

so that its expected payoff from conflict is given by

$$\pi p - c(r) = \pi p - (1/\alpha)\pi p \overline{p} = \pi[kp + (1-k)p^2], \tag{3}$$

where $k \equiv (\alpha - 1)/\alpha$, which lies in $(0, 1)$. This is not closed-form because p, an endogenous outcome, enters this equation. However, notice that the two equations in (2) together imply

$$\frac{r}{\overline{r}} = \left(\frac{\pi}{\overline{\pi}}\right)^{1/\alpha} \equiv \gamma, \tag{4}$$

which can be used along with the contest success function in (1) to conclude that

$$p = \frac{mr}{mr + (1-m)\overline{r}} = \frac{m\gamma}{m\gamma + (1-m)}, \tag{5}$$

where γ is defined in (4). Together, (3) , (4) and (5) describe a full solution to the Rebel's payoff in conflict equilibrium. A parallel expression holds for the State.

Conflict is a threat to peace. We seek conditions under which that threat might manifest itself. Say that a peaceful allocation $\mathbf{x} \in X$ is *blocked* if the expected payoff to the Rebel under conflict exceeds its average payoff under the allocation:

$$\pi[kp + (1-k)p^2] > \frac{1}{m} \int_{\text{Rebel}} x(i). \tag{6}$$

We ask whether small or large Rebels are more likely to be involved in conflict. To address this question, we link the appropriable surplus v to the victory payoffs π and $\overline{\pi}$ for each group.

It should be noted that our approach differs in a basic way from Esteban and Ray (1999, 2001, 2011). These papers consider a situation in which conflict is *ongoing*, not begun, and study the relationship between the level of conflict and group sizes in society. Our model is more special, in that it only deals with bilateral conflict. And yet it is different, in that it asks the question of which configurations are more conducive to the *initiation* of conflict, and in particular the role of group size in shaping those configurations.

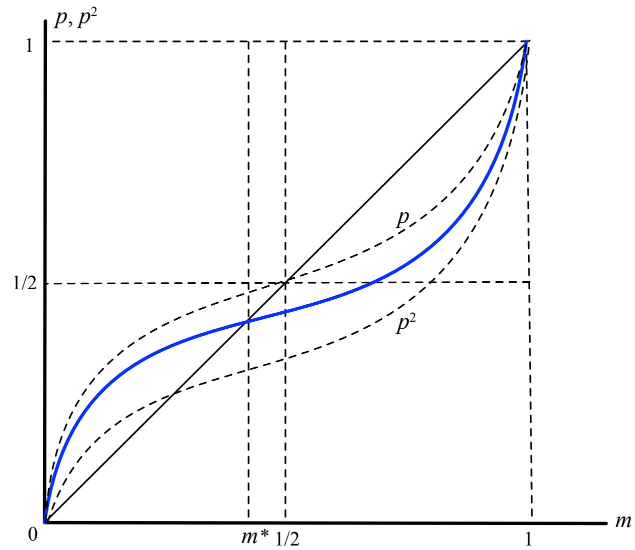


Figure 2. Threshold for Conflict with Private Prize and Non-Discriminatory Allocation.

3. Group Size and Conflict

3.1. Private Goods. Suppose that the prize is private; say, oil located on the homeland of a potential Rebel. Then $X = \{\mathbf{x} | \int x(i)di = v\}$. We assume that the winning group seizes the resources v entirely and excludes losers from the division of the spoils. Therefore, with a Rebel of size m,

$$\pi = v/m \text{ and } \overline{\pi} = v/(1-m).$$

Using this information in (4), we see that $\gamma = \left(\frac{1-m}{m}\right)^{1/\alpha}$, so that by (5),

$$p = \frac{m^k}{m^k + (1-m)^k}, \tag{7}$$

where $k = (\alpha - 1)/\alpha$. Notice from (7) that smaller Rebels are disadvantaged in conflict in the sense that they have a lower probability of winning; p is increasing in m and $p(1/2) = 1/2$. And yet:

Proposition 1. *If the prize is private, there is $m^* \in (0, 1/2)$ such that a Rebel with $m < m^*$ will block the non-discriminatory allocation. Society is conflict-prone in the presence of smaller Rebels.*

Proof. The non-discriminatory allocation gives v to every player. Using (3), conflict payoff is given by $\pi[kp + (1-k)p^2] = v[kp + (1-k)p^2]/m$. So a Rebel of size m will block if

$$kp(m) + (1-k)p(m)^2 > m, \tag{8}$$

where $p(m)$ is given by (7). This function has a "reverse-logistic" shape. It starts above the 45^0 line, crosses it at $n = 1/2$ and then dips below. The derivatives at the two ends are infinite.[5] See Figure

[5] To check these claims, note that $\frac{m^k}{m^k+(1-m)^k} \geq n$ if and only if $m \leq 1/2$ (simply cross-multiply and verify this), and that $p'(m) = \frac{km^{k-1}(1-m)^{k-1}}{[m^k+(1-m)^k]^2}$, which is infinite both at $n = 0$ and $n = 1$.

2, which plots p, p^2 and the convex combination $kp + (1-k)p^2$. With this shape in mind, observe that the left-hand side of (8) starts out higher than the right-hand side for small values of m, but ends up lower. Note that

$$kp(m) + (1-k)p(m)^2 < m,$$

for *any* $m \geq 1/2$.[6] This observation, in conjunction with Figure 2, shows that there is a unique intersection (crossing from above to below) in the interior of $(0, 1/2)$.[7] ∎

What matters is *not* the level of win probabilities or whether these rise or fall with group size. In our theory, while small Rebels fight more intensely (the per-capita stakes are higher), they have a lower probability of winning than big groups do. Thus small groups engage in conflict not because they have a high chance of winning. (They don't.) Rather, they do so because they have a high chance of winning *relative* to their share from the non-discriminatory allocation. That fact is reflected in the reverse-logistic shape of the win probability, derived in the proof of Proposition 1.

We do not interpret this result as a small Rebel deliberately initiating conflict in some "unprovoked fashion." State attempts to control these resources can be viewed by the group in question as an unwarranted infringement of its rights (to the resource). In that case, the correct interpretation is not one of conflict initiation, but rather one of resistance. We bolster this interpretation by studying resources located on the ethnic homeland. We should be also careful not to take Proposition 1 as literally applying to *all* group sizes, however small. Certainly *some* minimum threshold size is needed to even pose a serious threat, and we take our model to apply only above such thresholds.

3.2. Public Goods. Now suppose that v is a proxy for appropriable *public* payoffs, such as political power, human rights, or funding for secular/religious infrastructure. It could even represent private gains from public policies, such as protectionism or job reservations for an ethnic group.

Suppose there are n *disjoint* groups, each with their favorite public good. (We return to the disjointness assumption in Section 3.3.) An allocation is a vector $\mathbf{v} = (v_1, \ldots, v_n)$ of per-capita group-public payoffs for each group. For concreteness, regard v as a budget for different group-specific public goods (e.g., temples or mosques), and say that an individual gets payoff 1 from each unit of the budget spent on her group good; otherwise, she gets zero. (This payoff structure can be easily generalized.) Then, given that each person belongs to just one group, the payoff to a person from a budget allocation $\mathbf{v}$ is v_j, where j is her group membership. X is the set of payoff allocations $\mathbf{x}$ that can arise from all such budget allocations. The non-discriminatory allocation is given by dividing the budget equally across all groups, so that each individual obtains a payoff of v/n in peacetime. We take note of the different from a private prize, where the non-discriminatory allocation is achieved by dividing the budget in proportion to group population.

---

[6] Suppose this is false for some $1 > m \geq 1/2$. By the properties of p already established, we know that $m \geq 1/2$ implies $m \geq p(m)$, so that $km + (1-k)m^2 \geq m$, but this can never happen when $m < 1$, a contradiction.

[7] More formally, the derivative of $kp(m) + (1-k)p(m)^2$ is strictly smaller than 1 at any intersection, so that there can be only one intersection; we omit the details.

A winning Rebel gets to implement its own good, so that $\pi = v$. Assume that if the State wins, it excludes the Rebel and implements the non-discriminatory allocation for everyone else, with payoff $v/(n-1)$. The crucial point is that in the public prize case, group population size is eliminated as a determinant of per-capita payoff. An amount v_j spent on the favorite public good of a group j yields each member of that group v_j no matter what the group size is.

Proposition 2. *If the prize is public, there is $\hat{m} \in (0,1)$ such that a Rebel with $m > \hat{m}$ will block the non-discriminatory allocation. Society is conflict-prone in the presence of larger Rebels.*

Proof. Consider any conflict involving a Rebel of size m and the State of size $\overline{m} = 1 - m$. Then (5) tells us that

$$p(m) = \frac{m\gamma}{m\gamma + (1-m)}, \tag{9}$$

where $\gamma = [\pi/\overline{\pi}]^{1/\alpha} = (n-1)^{1/\alpha}$ is independent of m. Using (6) for the nondiscriminatory allocation with payoff v/n per-capita, we see that the Rebel will wish to engage in conflict if

$$kp(m) + (1-k)p(m)^2 > 1/n. \tag{10}$$

Given (9), the left-hand side of this inequality is monotonically increasing in m. For m close to zero, the inequality must fail because $p(m) \to 0$, and for m close to 1 the inequality must hold because $p(m) \to 1$. Define $\hat{m}$ by equality in the relationship above to complete the argument. ■

Numerical calculations are easy to perform. Combining (9) and (10) and remembering that $\gamma = (n-1)^{1/\alpha}$, the blocking condition for conflict reduces to

$$k\frac{m(n-1)^{1/\alpha}}{m(n-1)^{1/\alpha} + (1-m)} + (1-k)\left[\frac{m(n-1)^{1/\alpha}}{m(n-1)^{1/\alpha} + (1-m)}\right]^2 > \frac{1}{n},$$

and some straightforward but tedious computation eventually reveals that

$$\hat{m} = \left[1 + (n-1)^{1/\alpha}\left\{\frac{(1+\alpha) - \sqrt{(\alpha-1)^2 + \frac{4\alpha}{n}}}{\sqrt{(\alpha-1)^2 + \frac{4\alpha}{n}} - (\alpha-1)}\right\}\right]^{-1}. \tag{11}$$

When there are just two groups and the cost function is quadratic, then the Rebel needs to exceed 61.8% of the population. When there are three groups and $\alpha = 1.2$, then the Rebel needs to exceed 39.7% of the population. We can use (11) to perform these calculations for any number of groups and any cost curvature, but the point should be clear: it is large groups (typically but not always larger than the average) that pose a threat when the potential conflict is over public goods.

To remain within the formal confines of our baseline model, we have assumed that the payoff from a group's favorite public good benefits *just* that group and not others. In no way is this necessary for our result. If a group's victory also means that others in the population benefit, that does not alter the group's calculus, conditional on entering into a conflict with the State. So our setting also includes the case in which a dominant group is fighting to restore democracy or rights to the State, which could spill over to the populace at large. Or alternatively, that group could be seeking group-specific power by seizing the center and continuing an autocratic regime. Either interpretation is consistent with the public goods formulation.

3.3. Non-Discriminatory Peaceful Allocations and Multiple Threats. Our analysis so far presumes that peacetime allocations are non-discriminatory. Of course, Proposition 1 applies even more strongly if society has a reason to favor larger groups to begin with, as it will in a democratic (or voting) scenario. But if the initial allocation is chosen to appease the small groups, then larger groups will have to pay for that appeasement, and matters are more complex.[8]

Suppose that there is a variety of potential markers (religion, caste, occupation, ethnicity, geography, and so on) that might delineate a potential Rebel coalition. To formalize the idea of multiple threats, say that a finite collection C of groups (or potential Rebels) is *balanced* if there is a set of weights in $[0,1]$, $\{\lambda(G)\}_{G\in C}$, such that

$$\sum_{G\in C_i} \lambda(G) = 1 \text{ for every } i \text{ in society,} \tag{12}$$

where C_i is the subcollection of all groups for which i is a member.

Balancedness implies that it is hard to find small groups of individuals who are central to many potential conflicts, and who can therefore be co-opted to shut down a number of conflicts at once. Consider an example illustrating the opposite: suppose that C is fully described by any collection of potential Rebels that contain the special set $[0,1/2]$. Then that collection is not balanced: we relegate the details to a footnote.[9] It contains some distinguished group (in this example, $[0,1/2]$) which is "over-represented" in the collection. In contrast, a balanced collection contains no "over-represented" group. For instance, any partition of $[0,1]$ is a balanced collection (simply use $\lambda(G) = 1$ for all G and verify that the balancing condition is satisfied).[10] We can now state:

Proposition 3. *Assume that the prize is private. Suppose that the collection of all potential Rebels includes a balanced collection C, with each member of cardinality $m < m^*$, where m^* is given by Proposition 1. Then every peaceful allocation, non-discriminatory or otherwise, is blocked by some member(s) of this collection.*

Proof. See Appendix. ■

[8]Discriminatory peacetime allocations are of separate interest because of the Coase Theorem. Because conflict is costly, for each conflictual outcome there is a "peaceful" outcome that Pareto-dominates it, provided that appropriate Coaseian transfers are available. But is there *one* outcome that can *simultaneously* withstand all threats? It is true that conflict is inefficient, but if the variety of potential threats is large relative to the degree of inefficiency, *every* peacetime allocation, discriminatory or not, may be blocked by *some* coalition. This is akin to the problem of an empty core in characteristic function games (Bondareva 1963, Shapley 1967, and Scarf 1967).

[9]For suppose we could find "balancing weights" $\{\lambda(G)\}$; then (12) holds for $i \in [0,1/2]$, but i is contained in every $G \in C$, which implies that $\sum_{G\in C} \lambda(G) = 1$. Now pick any G' with $\lambda(G') > 0$. Because G' is a strict subset of $[0,1]$, there is some individual $j \notin G'$. Given (a.1), it must be the case that $\sum_{G\in C_j} \lambda(G) < 1$, which contradicts balancedness.

[10]Or, if $[0,1]$ is the union of K equally-sized intervals of the form $[i/K,(i+2)/K]$, for $i = 0,\ldots,K-1$, then the collection $\{[0,2/K),[1/K,3/K),[2/K,4/K),\ldots,[(K-2)/K,1),[(K-1)/K,1/K)\}$ has "overlaps" but is also balanced.

Because (as already noted) every partition is balanced, the following corollary applies:

Corollary 1. *Suppose that society can be partitioned into potential Rebels of size $m < m^*$. Then there is no allocation for society that is immune to conflict.*

Note that we do not place any assumptions on the peacetime allocations. They could be *any* allocation of the private good, perhaps discriminating across individuals in the same coalition. And yet, if there is a sufficiently varied multiplicity of small groups all challenging the private prize, society is necessarily unable to find a peaceful allocation that buys off all potential Rebels.[11] Of course, it is possible that for some unequal allocations, a large group may also want to instigate a conflict. But in such a case, *some* small group also will — under the conditions of Proposition 3.

There is a parallel "empty core" argument that we can conduct for public goods. Proposition 2 survives with no essential change whether or not the initial allocation is non-discriminatory. The relevant threshold $\hat{m}$ will change with the size of the proposed peacetime offer, but the qualitative result survives with no alterations. It should also be noted that with arbitrary allocations, there is no need to assume that the groups are pairwise disjoint. That assumption was only used to ensure that a non-discriminatory allocation exists. With group intersections, a non-discriminatory allocation may not exist.[12]

But we also take note of an importance difference between the two cases. With public goods, we need to be especially careful about the transferability of payoffs and exactly what it entails. We have restricted ourselves to the case in which budgets are transferable across groups, not in units of money, but by changing the allocation of public goods. One might allow for a broader class of transfers in which compensatory side-payments of money are made from one group to another in exchange for an uneven distribution of public goods.[13] The analysis of this case is somewhat different, and it is generally more conducive to equilibrium outcomes that are non-conflictual. (The details are available on request from the authors.)

4. Group Size and Conflict: Empirics

Our theory implies that smaller groups are more conflictual if the prize is private, and likewise larger groups are more conflictual if the prize is public. There are several considerations that arise when attempting to empirically implement the theory. These include, but are not limited to, a

[11] The balancedness condition on potential Rebels, while sufficient, is not necessary for our result. For instance, suppose that the cost function is quadratic (so $\alpha = 2$). It is then easy to verify that $m^* = 25\%$. However, it is possible to check that if there are six pairwise disjoint groups of size 10%, conflict is inevitable regardless of the baseline allocation, even if the balancedness condition is not met.

[12] For instance, if there are two groups that intersect but neither group is a subset of other, a non-discriminatory allocation will not exist, as members of the intersection will benefit to a greater degree from any allocation.

[13] This approach with transferability should be used with caution. Public goods are not like oil revenues. Think of ethnic or religious representation, or the sharing of political power. The relative price across objects such as these may be very hard to define. So it may be impossible to conceive of "classical" financial transfers as compensation for the loss of power or culture; see, e.g., Kirshner (2000). What price would those who are thus negated accept as compensation?

suitable definition of "groups," as well as notions of "private" and "public" payoff components. We discuss these issues below as they come up.

4.1. Basic Data. We begin with the data that run through the entire exercise.

4.1.1. Ethnic Groups. To define potential Rebel groups, we use ethnicity. Esteban and Ray (2008) provide arguments for ethnic markers to be salient in conflict. Ethnic conflicts account for 50–75% of internal conflicts since 1945 (Fearon and Laitin 2003, Doyle and Sambanis 2006), and so represent a natural and relatively tractable choice. We use the sample of ethnic groups from the dataset "Geo-Referencing of Ethnic Groups," or GREG (see Weidman, Rod and Cederman 2010). GREG is based on the *Atlas Narodov Mira* or ANV (Bruk and Apenchenko, 1964), created by Soviet ethnographers with the aim of locating and charting ethnic groups worldwide. It provides information on the homelands of 929 groups and it employs a consistent classification of ethnicity with a uniform group list that is valid across state borders.[14] Most homelands are coded as pertaining to one group only, but in some instances up to three ethnic groups share the same territory. Using this information, we can create group-country pairs: that is, we assign ethnic groups to countries depending on the land area occupied by them in each country.[15] That yields 1475 distinct group-country pairs located in 145 countries, to be referred to from now on simply as "group." Our central variable, SIZE, is the size of the (country-specific) group relative to that of the country population.[16]

GREG's settlement patterns are a snapshot from the late 1950s and early 1960s. That has advantages and disadvantages. On the negative side, settlement patterns may be outdated for some parts of the world. Also, as ethnic maps were chartered by Soviet ethnographers during the Cold War, the level of accuracy and resolution varies considerably for different regions in the world. On the positive side, it alleviates concerns that ethnic group locations are endogenous to the conflicts we aim to explain. The locational detail in ANV/GREG also enables us to merge it with other geo-referenced datasets needed for the computation of some of our key group-level variables. When all is said and done, we have a panel dataset at the ethnic group level with global coverage, with information for 145 countries and 1475 ethnic groups over 1960 to 2006.[17]

4.1.2. Conflict. Group-level conflict data come from Cederman, Buhaug and Rod (2009), who use the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al. 2002), checking this against sources that identify ethnic civil wars (such as Fearon and Laitin 2003, Licklider 1995 and Sambanis 2001). Ethnic conflicts are coded based on whether mobilization was shaped by ethnic affiliation, and in such cases, the various groups involved are recorded. Our baseline measure is group-level conflict

[14] The ANV actually contains information for 1248 groups, but 319 of them do not have any territorial basis.

[15] The definition of ethnic group is not clearly stated anywhere in the ANV. It is only possible to infer the coding criteria by comparison with existing data sources on ethnic groups. Fearon (2003) argues that the main grouping criterion in the ANV is the historic origin of language. GREG contains a larger number of groups than alternative sources (such as the Geo-Ethnic Power Relations dataset) as it contains many small-language groups.

[16] Population figures correspond to the early 60's, see Cederman, Buhaud and Rod (2009) for details.

[17] Morelli and Rohner (2015) consult similar sources for ethnic group location and oil fields.

incidence, a binary variable set to 1 in any year that group is involved in an armed conflict against the state, with more than 25 battle-related deaths in that year. To explore the robustness of our results other conflict measures are employed. Among them, group-level conflict *onset* is equal to 1 in a given year if an armed conflict against the state resulting in more than 25 battle-related deaths *begins* in that year.[18] As another variation, we drop the time dimension and use the share of years a group has been involved in conflict against the State (or the share of onset years).

4.1.3. Controls. Throughout, we employ a number of controls, both at the group and at the country level, and depending on the fixed-effects specification that we use. Group-level controls are from Cederman et al. (2009) or directly computed from GREG. MOUNT is an index that captures the group's share of mountainous terrain. GROUPAREA is homeland area (in thousands of square km). DISTCAP measures the group's distance to the country capital. GIP is 1 if the ethnic group is in power in the country, lagged one year. SOILCONST measures the limitations of homeland soil for agriculture. PARTITIONED is 1 if the group's homeland is located in two or more countries. PEACEYRS is the number of years since the last group-level onset and LAG is lagged conflict incidence. At the country level we control for the log of GDP per capita, lagged one year (GDP), the polity index (POLITY), lagged one year, and the log of total population (POP), also lagged one year. GDP and population are taken from the Penn World Tables while POLITY is taken from Polity IV.

4.2. Private Payoffs and the Effect of Group Size. We begin with the analysis of group size in a private-prize setting. We study resources that are located in the homeland of each ethnic group. The underlying presumption is that the State seeks to extract those resources and distribute them more evenly across the country, and that the ethnic group in question can either accept the State policy, or reject it.[19] Alternatively, the group actively seeks conflict to retain those resources. Neither the theory nor the empirics is rich enough to identify the precise perpetrator in these situations, and for our purposes the two are equivalent.

4.2.1. Measures for the Private Prize. We ask if the ethnic homeland is rich in natural resources. In our baseline specification we use oil in the homeland as a proxy for "private prize". This measure, OIL, is computed as follows. First, we obtain geo-referenced information on the location of oil fields and associated discovery dates from Petrodata (Lujala, Rod and Thieme, 2007). Next, we combine the information on group and oil location from GREG and Petrodata, respectively, to construct maps of oil fields at the ethnic group level. Finally, OIL is computed as the log of the ethnic homeland area covered by oil (in thousands of km^2) times the international price of oil. Our results are robust to alternative ways of measuring oil abundance (see Table B.5).

[18]In practice, conflict onset as defined by the PRIO threshold is far from a sharp concept. Before the threshold is crossed, we might have several manifestations of serious conflict (a breakdown in negotiations, an insurgency, a crackdown). Thus, a year of onset is arguably no different from a year of incidence, though to be sure, the factors that contribute to the outbreak of a conflict do not coincide with the ones that continue to feed it (Schneider and Wiesehomeier 2006). This is why we control for lagged conflict in our incidence regressions.

[19]Note that ethnic homelands are *not* countries. A variant considers resources at the national level. We do so in Table B.5, obtaining similar results.

As robustness checks, we also consider mineral availability and land abundance as well as an index of group "resource wealth" (see Table 2), computed via factor analysis on these separate proxies; namely, oil, minerals and land. In all cases, we tie private prizes firmly to ethnic homelands, so that there is a sense of "group heritage." Certainly, the State as a whole can attempt to redistribute the revenues from those resources over the country as a whole, or settle relatively abundant lands with other ethnicities. If violent conflict occurs in that process, our data will pick it up.

4.2.2. Specification. We run the following specification and some variants of it:

$$\textsc{conflict}_{c,g,t} = \beta_1 \textsc{size}_{c,g} + \beta_2 \textsc{priv}_{c,g,t} + \beta_3 \textsc{size}_{c,g} \times \textsc{priv}_{c,g,t} + X'_{c,g,t}\alpha + Y'_{c,t}\delta + Z'\gamma + W'_t\eta + \epsilon_{c,g,t} \tag{13}$$

for countries c, groups g, and dates t. Our main outcome variable CONFLICT is "conflict incidence."[20] The variable PRIV is our measure of private prize — OIL will be the leading baseline used here — and its interaction with group size is of particular interest. Our theory predicts that β_3, the coefficient associated with SIZE $\times$ PRIV, is negative, implying that smaller groups are more likely to be involved in conflict as the private prize in the homeland becomes more abundant.

Unless otherwise stated, we always employ group- and country-level controls ($X_{c,g,t}$ and $Y_{c,t}$ respectively), a vector Z of fixed effects imposed either for countries or for (country-specific) groups, and year dummies W_t. With country fixed effects, identification for the interaction term SIZE $\times$ OIL is achieved, because we have variation in ethnic groups within countries — so that SIZE varies — and also intertemporal variation in oil prices or in known reserves, so that OIL varies. With group fixed effects, the former variation must be sacrificed, but the latter is still available.

We estimate equation (13) by OLS. The reason for fitting a linear probability model (rather than a non-linear specification, such as probit or logit) is that our key variables are interactions and interpreting them in nonlinear models isn't straightforward, as Ai and Norton (2003) point out.[21] For completeness, we study nonlinear variants in Section 5.8.

Robust and clustered standard errors are computed in all cases. We follow Abadie et al. (2017) and cluster errors according to the clustering of the assigned treatment. That is, whenever the "treatment" of interest is at the group (country) level, we cluster errors at the group (country) level as well. Our results are robust to other clustering strategies such as two-way clustering (at the country and ethnic group level, where the latter considers all the territories occupied by the same group, even if they belong to different countries). See Table B.6 in the Online Appendix.

[20] In the robustness check section, we consider an alternative specification where the unit of analysis is the group — not the group-year as in our main analysis — and the dependent variable is the share of conflict years, see Table B.4.

[21] In linear models, the coefficient of the interaction term has a direct interpretation, as it is just the value of the cross derivative of the dependent variable with respect to the variables in the interaction. However, this logic does not extend to nonlinear models: the cross derivative in this case is a more complicated object. As shown by Ai and Norton (2003), its value depends on all the covariates of the model and the sign does not necessarily coincide with the sign of the coefficient of the interaction, see Section B.3 of the Online Appendix for a more detailed discussion.

	Dependent Variable: **Conflict Incidence**							
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
SIZE	-0.015	0.032	0.066***		-0.019	0.049**	0.080***	
	(0.307)	(0.101)	(0.001)		(0.221)	(0.017)	(0.000)	
OIL	0.448**	0.684***	0.771***	-0.094				
	(0.040)	(0.009)	(0.007)	(0.778)				
OIL_{0-25}					-0.002	-0.002	-0.004***	-0.005***
					(0.176)	(0.153)	(0.007)	(0.002)
OIL_{25-50}					-0.000	-0.000	-0.001	-0.004**
					(0.798)	(0.803)	(0.701)	(0.043)
OIL_{50-75}					0.003*	0.004**	0.004**	-0.001
					(0.082)	(0.017)	(0.023)	((0.748)
$\text{OIL}_{>75}$					0.004**	0.006***	0.007***	0.002
					(0.041)	(0.010)	(0.007)	(0.447)
SIZE × OIL		-13.628***	-14.433***	-8.680**				
		(0.000)	(0.000)	(0.011)				
SIZE × OIL_{0-25}						-0.014	0.047	0.051
						(0.831)	(0.539)	(0.591)
SIZE × OIL_{25-50}						-0.063	-0.058	0.075
						(0.151)	(0.398)	(0.454)
SIZE × OIL_{50-75}						-0.152***	-0.154***	-0.107**
						(0.000)	(0.000)	(0.020)
SIZE × $\text{OIL}_{>75}$						-0.130***	-0.135***	-0.100***
						(0.000)	(0.000)	(0.003)
POLITY			-0.002**	-0.003*			-0.002**	-0.003**
			(0.016)	(0.056)			(0.016)	(0.045)
GIP			-0.004**	-0.000			-0.003**	-0.000
			(0.044)	(0.853)			(0.049)	(0.895)
GDP			0.001	0.004**			0.001	0.005**
			(0.156)	(0.049)			(0.195)	(0.036)
POP			0.001	-0.000			0.002	-0.000
			(0.589)	(0.745)			(0.462)	(0.941)
GROUPAREA			0.000				0.000	
			(0.271)				(0.530)	
SOILCONST			-0.001				-0.001	
			(0.299)				(0.235)	
DISTCAP			0.002***				0.002***	
			(0.000)				(0.000)	
MOUNT			0.002*				0.003*	
			(0.069)				(0.064)	
PARTITIONED			-0.001				-0.001	
			(0.296)				(0.337)	
LAG	0.895***	0.895***	0.894***	0.924***	0.895***	0.894***	0.894***	0.924***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.844	0.844	0.847		0.844	0.844	0.848	
Obs	64839	64839	55289	55289	64839	64839	55289	55289
Fixed effects	CFE	CFE	CFE	GFE	CFE	CFE	CFE	GFE
Estim. Method	OLS	OLS	OLS	GMM	OLS	OLS	OLS	GMM

Table 1. Group Size and Conflict: Private Prize. This table regresses conflict incidence on group size and indices of private prizes, along with interactions between these variables. All regressions contain year dummies and country fixed effects (CFE) except for Columns 4 and 8 that replace the latter by group fixed effects (GFE). Dummy variables OIL_{i-j} equal 1 if the value of oil reserves is between the ith and the jth percentile of the distribution of OIL, conditional on having oil in the homeland. Regressions are estimated by OLS except for Columns 4 and 8, which are estimated using system GMM (Blundell and Bond, 1998). The time period considered is 1960-2006. p-values are reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

4.2.3. Baseline Results for Private Prizes. In Table 1, the dependent variable is conflict incidence. Each Column reports on a different linear probability specification, all with lagged conflict, year fixed effects and country or (country-specific) group fixed effects. Column 1 regresses INCIDENCE on only two variables: group size (SIZE) and our baseline measure of private prize, group-level oil abundance (OIL).[22] The abundance of oil in the ethnic homeland is positively associated with conflict incidence involving that ethnicity. As already observed, this is a well-established correlation. The coefficient of SIZE is insignificant. This is precisely what the theory would lead us to expect, as it predicts that the *unconditional* effect of group size on conflict is ambiguous. Column 2 introduces the interaction of SIZE and OIL. The coefficient of the interaction term is negative and significant, as predicted by the theory. Column 3 adds on controls to the regression in Column 2, with no change in the results. Column 4 replaces the country dummies in the previous specifications by (country-specific) group fixed effects. The inclusion of group fixed effects contributes to the reduction of potential bias from omitted variables, which is a good thing. The drawback is that all time-invariant controls drop out from the regression, including one of our key variables, SIZE. Nevertheless, it is still possible to test the key hypothesis pertaining to SIZE × OIL. Column 4, estimated using system GMM (Blundell and Bond 1998), shows that the interaction of SIZE and OIL remains negative and significant.

Columns 1–4 impose the restriction that the marginal effect of SIZE on conflict is linear in oil, and shows that the presence of oil attenuates the effect of size on conflict. But the theory makes a sharper prediction: that the marginal effect of SIZE actually turns *negative* as the prize becomes increasingly private. We could see this by extrapolation of the interaction effect for large private prizes, but that might take our particular specification too literally. Instead, we re-do these Columns using a more flexible specification, one in which linearity is not imposed. In Columns 5–8, we employ four dummies that correspond to the quartiles of the distribution of OIL for the groups that have oil in their homeland (the omitted category corresponding to groups without oil).

Column 5 shows that the effect of group size is not statistically different from zero when just the oil dummies and country and year fixed effects are introduced. Column 6 adds interactions of group size and each of the oil dummies and shows that they are negative and significant in groups with large quantities of oil. Column 7 adds controls to the previous specification and Column 8 replaces country fixed effects by group fixed effects. Similar results are found.

How large are these effects? We postpone this discussion to Section 4.3.5, once we have included both private and public prizes in the analysis. Additionally, this Section revisits the oil-dummy specification studied here in Columns 5-8, showing that the marginal effect of SIZE on conflict is positive and significant in the absence of oil, but that the effect decreases as the amount of oil in the homeland becomes more abundant, *and that it eventually becomes negative* for groups with abundant oil reserves. Our claimed effect is not driven by merely extrapolating a linear specification.

[22]For convenience, the coefficients of SIZE and its interactions have been multiplied by 10 in all tables.

This table — as well as the variations to come — contains an additional observation. With the interaction of group size and oil in place, SIZE *alone* generally turns positive and significant, and its absolute magnitude is decidedly larger than in Columns that do not contain that interaction. The absolute value of the coefficient of SIZE in Column 3 (Column 7) is almost 4 times larger than that in Column 1 (Column 5). This observation suggests that once a significant private component of the prize is "removed" (i.e., controlled for), group size is positively related to conflict.

4.2.4. Alternative Measures of Private Prizes. Table 2 considers other proxies of privateness (Columns 1–7) as well as an additional implication of our theory (Columns 8–11). Our main alternatives are MINES, which is mineral availability in the ethnic homeland, as well as HOME, the area of the ethnic homeland as a fraction of country area. Both resources can be "seized" via redistribution or settlement. For the measure of HOME, there is nothing to add. For MINES, we use geo-referenced data on mining activities around the world since 1980.[23] For each year and mine, we know whether that mine is active or not, and the specific minerals produced by it. As in Berman et al (2015), we focus on 13 minerals for which we have world price data,[24] which we take from the World Bank's commodity price database. For each group, year and mineral, we create a dummy variable that is one if the group has at least one active mine of that mineral. To introduce information on mineral prices, we multiply each of the mineral dummies by (the log of) its international price, normalized by (log) price in 1980. MINES is the sum of the resulting quantities for each group and year.

Column 1 introduces MINES as an additional regressor. As with OIL, the coefficient of SIZE is not significant. Column 2 adds the interaction of the mines index and group size which has a negative and significant coefficient, as predicted.[25] Column 3 introduces group fixed effects and uses system GMM to estimate the resulting equation. The results are robust.

Column 4 shows that the coefficient of SIZE is insignificant with the introduction of HOME, along with OIL, country and year fixed effects, and the controls. But in line with our findings, Column 5 shows that the interaction of SIZE and HOME is indeed negative and significant, suggesting that small groups are more likely to be involved in conflict as the value of HOME increases. We cannot consider a variation with group fixed effects as HOME has no time variation.

Column 6 uses an alternative proxy of privateness, PRIVINDEX, computed using factor analysis on the three prize indicators oil, land and mineral abundance, along with country fixed effects. Column 7 is similar but uses group fixed effects instead. In both cases, similar results are obtained.

Recall our remark at the end of Section 4.2.3: that SIZE becomes positive and significant after the interaction with the private prize is introduced. With a similar perspective in mind, Columns 9–11 explore our result from another angle. The theory rests on the idea that large groups are less likely to fight because payoff per head is relatively low. The implication is that if one controls for *per*

[23] The source is the *Raw Material Data* (IntierraRMG, 2015). Since data on mining activity starts in 1980, our sample in these regressions focuses on the period 1980–2006.

[24] These are bauxite, coal, copper, diamonds, gold, iron, lead, nickel, platinum, phosphates, silver, tin and zinc.

[25] Similar results hold when only information on mine availability is used to compute the mines proxy.

	Dependent Variable: **Conflict Incidence**										
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
SIZE	0.013	0.053		0.033	0.168***	0.066**		0.022	0.038**	0.034**	0.034**
	(0.586)	(0.118)		(0.576)	(0.005)	(0.013)		(0.200)	(0.017)	(0.028)	(0.028)
SIZE× MINES		-0.015**	-0.081**								
		(0.049)	(0.015)								
SIZE× HOME					-0.370***						
					(0.001)						
SIZE× PRIVINDEX						-0.049***	-0.033**				
						(0.001)	(0.031)				
OIL	0.551*	0.521	17.707**	0.551**	0.407*			0.556**			
	(0.073)	(0.195)	(0.012)	(0.022)	(0.092)			(0.028)			
MINES	-0.000	0.000	-0.006**								
	(0.672)	(0.996)	(0.022)								
HOME				-0.002	0.016						
				(0.823)	(0.111)						
PRIVINDEX						0.002**	-0.000				
						(0.017)	(0.882)				
OIL PC									-0.001		0.012
									(0.971)		(0.611)
LAND PC										-0.001**	-0.001**
										(0.036)	(0.035)
POLITY	-0.001	-0.001	-0.002	-0.002**	-0.002**	-0.002**	-0.004	-0.002**	-0.002**	-0.002**	-0.002**
	(0.169)	(0.408)	(0.497)	(0.017)	(0.019)	(0.020)	(0.126)	(0.016)	(0.017)	(0.018)	(0.018)
GIP	-0.002	-0.002	-0.009	-0.004**	-0.005***	-0.004**	0.002	-0.003**	-0.003*	-0.003*	-0.003*
	(0.344)	(0.411)	(0.104)	(0.050)	(0.006)	(0.036)	(0.296)	(0.050)	(0.064)	(0.054)	(0.054)
GROUPAREA	-0.000	0.000				0.000		-0.000	0.000		
	(0.796)	(0.399)				(0.333)		(0.629)	(0.318)		
GDP	0.004*	0.004	-0.003	0.001	0.001	0.001	0.005	0.001	0.001	0.001	0.001
	(0.095)	(0.246)	(0.444)	(0.162)	(0.160)	(0.193)	(0.108)	(0.148)	(0.133)	(0.133)	(0.133)
POP	-0.004	-0.004	-0.003**	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001
	(0.545)	(0.724)	(0.036)	(0.538)	(0.566)	(0.744)	(0.851)	(0.526)	(0.609)	(0.774)	(0.771)
SOILCONST	-0.001*	-0.001*		-0.000	-0.001	-0.001		-0.000	-0.000	0.000	0.000
	(0.062)	(0.076)		(0.402)	(0.224)	(0.256)		(0.417)	(0.625)	(0.719)	(0.733)
DISTCAP	0.002***	0.002		0.002***	0.002***	0.002***		0.002***	0.002***	0.002***	0.002***
	(0.002)	(0.167)		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
MOUNT	0.002	0.003		0.002*	0.002*	0.002*		0.002*	0.002	0.002	0.002
	(0.162)	(0.110)		(0.084)	(0.075)	(0.093)		(0.081)	(0.137)	(0.193)	(0.193)
PARTITIONED	-0.001	-0.001		-0.001	-0.001	-0.001		-0.001	-0.001	-0.001	-0.001
	(0.653)	(0.671)		(0.303)	(0.281)	(0.281)		(0.302)	(0.342)	(0.356)	(0.357)
LAG	0.887***	0.887***	0.895***	0.894***	0.894***	0.894***	0.897***	0.894***	0.895***	0.895***	0.895***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.838	0.838		0.847	0.847	0.847		0.847	0.847	0.847	0.847
Obs	33325	33325	33325	54486	54486	54486	32753	55289	55289	55289	55289
Fixed effects	CFE	CFE	GFE	CFE	CFE	CFE	GFE	CFE	CFE	CFE	CFE
Estim. Method	OLS	OLS	GMM	OLS	OLS	OLS	GMM	OLS	OLS	OLS	OLS

Table 2. Group Size and Conflict: Alternative Private Prize Specifications. This table regresses conflict incidence on group size and indices of private and public prizes, along with interactions between subsets of these variables. Columns 1–3 use mineral availability, Columns 4–5 use land-based measures, and Columns 6–7 use an index of privateness, as described in the text. Columns 8–11 show how the magnitude and the significance of SIZE change if one controls for *total* or *per capita* private payoffs. All Columns contain country fixed effects (CFE) and year dummies, and have been estimated by OLS, expect for Columns 3 and 7 which contain group fixed effects (GFE) and have been estimated using system GMM (Blundell and Bond, 1998). **p-values** (based on robust standard errors clustered at the group level) are reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

capita payoffs (rather than for total payoffs, as in the specifications above), large groups should *unambiguously* be more conflict-prone. To explore this prediction, Columns 9–11 control for per capita private payoffs. To facilitate comparison with the results so far, Column 8 controls once again for OIL, our baseline measure of total private payoff, and shows that group size is insignificant.[26] Column 9 is identical to Column 8 but replaces OIL by OIL PC, computed by dividing OIL by group population. In this case, the coefficient of SIZE is significant and more than 70% larger than in Column 8, suggesting that larger groups are more prone to conflict once per capita payoffs are held constant. Columns 10 and 11 show that a similar result also holds when land per capita (Column 10) and land and oil per capita (Column 11) are introduced in the regression.[27]

4.3. Public Payoffs and the Effect of Group Size. We now turn to public prizes. Our baseline has in mind the seizure of political power. That conflict could be a *struggle for democracy*, in which a group spearheads a movement for broad-based freedoms and rights. Or it could be a *struggle for autocratic control*, in which a group either seeks the reins of power to favor its own ethnicity, or to impose its religion or norms on the rest of society. Both prizes are public prizes as far as the *group* is concerned, though in the first case a positive externality is imposed on others in the society, while in the second case the externality is generally negative. In both the autocratic and democratic conflicts, our theory predicts that a larger group is likely to be more active. The spillovers are unimportant for this result: see the very end of Section 3.2 for a discussion of this point.

4.3.1. Measures of the Public Prize. Our contention is that absence of rights is a good proxy for either notion of the public prize. Under the democratic struggle, the group is at the vanguard of a rebellion that seeks to *restore* political and civil rights, and there is obviously more to restore if there is less of it in the first place. In an autocratic conflict, the absence of political and civil rights is suggestive of the power to be had if an ethnic group can take control of the State. It would be easier to formulate and implement policies that disproportionately benefit members of the group in power.[28] Therefore, in our baseline specification our variable is LACK RIGHTS, a composite index from Freedom House that measures the absence of political and civil rights. It is designed to capture the real-world rights and freedoms enjoyed by individuals. In formulating the index, both the legal guarantees of rights as well as actual practices are taken into account.[29]

LACK RIGHTS is the average of two individual indices, the lack of political rights and the lack of civil rights; see Freedom House for additional details.[30] For robustness we will also consider these indices individually (see Columns 3 and 4 of Table 3). We deliberately take LACK RIGHTS off the shelf so as to avoid any implication that the components or weights are chosen to suit our purpose. We are also aware that there are concerns of endogeneity: for instance, conflict can lead to changes in the level of rights. Therefore, we *only* consider pre-sample values of the index

[26] Column 8 is different from Column 1 as the latter also controls for MINES.

[27] We cannot perform a similar calculation using MINES as we don't have a measure of total mineral production.

[28] The resulting individual benefits might be private, but the overall prize — viewed as heightened access to power — is still in the nature of a group-specific public good.

[29] We renormalize the index to lie between 0 and 1, where 0 indicates the lowest rights provision.

[30] The political rights index is elaborated taking into account i) the freedom in the electoral process, ii) political pluralism and participation, and iii) the functioning of the government. The civil rights index evaluates i) the freedom of expression and belief, ii) associational and organizational rights, iii) the rule of law, and iv) personal autonomy and individual rights. See https://freedomhouse.org/report/methodology-freedom-world-2018 for additional details.

(and in addition we control for past conflict in all our regressions). Specifically, LACK RIGHTS is computed by averaging the values of the Freedom House index from the first year where it is available (which is 1972) to 1975, and is then employed in regressions using post-1975 data only. The resulting measure is "assigned" to all the ethnic groups in the country, so that LACK RIGHTS is a time-invariant country-level index.

We check the robustness of our results by considering alternative proxies for the publicness index as well as alternative ways of operationalizing the above-described measures; see Table 3. For instance, we use a group-level proxy for publicness, which is group exclusion, and is defined as the average over the period 1960–1975 of a dummy variable that captures whether a group is excluded from national power (Cederman et al, 2009). Our results are robust to using these alternative definitions.

4.3.2. Specification. We add an interaction term for public prizes to our previous specification:

$$\text{CONFLICT}_{c,g,t} = \beta_1 \text{SIZE}_{c,g} + \beta_2 \text{SIZE}_{c,g} \times \text{PRIV}_{c,g,t} + \beta_3 \text{PRIV}_{c,g,t} + \beta_4 \text{SIZE}_{c,g} \times \text{PUB}_c + X'_{c,g,t}\alpha + Y'_{c,t}\delta + Z'_c\gamma + W'_t\eta + \epsilon_{c,g,t} \tag{14}$$

for countries c, groups g, and dates t. As before, our main outcome variable CONFLICT is "conflict incidence." The new variable is an index PUB of a public prize. The theory states that β_4, the coefficient on the index between this prize and group size, is positive.

Group- and country-level controls are employed as before, but this time the vector Z_c of fixed effects only pertains to countries, and not to groups. The imposition of country fixed effects means that the only source of variation for the interaction term SIZE $\times$ PUB is changes in ethnic groups within countries, because PUB is a *country-level*, time-invariant indicator. That means that we cannot use group fixed effects for regressions that involve such invariant public prizes. The remarks about the clustering of standard errors in Section 4.2.2 also apply to the analysis at hand.

4.3.3. Baseline Results With Public Prizes. Tables 3 summarizes our baseline results for public prizes. It imposes $\beta_2 = 0$ in equation (14) and so focuses exclusively on the interaction of SIZE and different indices of publicness, chief among them being LACK RIGHTS. All specifications contain lagged conflict, country and year fixed effects and are estimated by OLS.

Column 1 contains three regressors: OIL, SIZE and the interaction of SIZE with LACK RIGHTS, our baseline measure of publicness.[31] Recall that LACK RIGHTS is a pre-sample time-invariant index, so it is not an independent regressor and is subsumed in the country fixed effects. But the *interaction* of SIZE and LACK RIGHTS is, of course, well-defined, it has the predicted positive sign and it is highly significant. Column 2 introduces similar controls as the previous tables; the result is robust.

[31] As already mentioned before, standard errors have been clustered at the country level, respecting the fact that the "treatment" PUB is assigned at the country level. Our results are robust to the consideration of other clustering strategies, such as clustering at the country level or using two-way clustering, with errors clustered at the group (as opposed to country-group) *and* country level. See Table B.6 in the Online Appendix.

	Dependent Variable: **Conflict Incidence**							
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
SIZE	-0.073***	-0.045	-0.044	-0.041	-0.037	-0.020	-0.047*	0.028
	(0.007)	(0.152)	(0.173)	(0.160)	(0.186)	(0.493)	(0.098)	(0.386)
SIZE × LACK RIGHTS	0.097**	0.106**						
	(0.028)	(0.019)						
SIZE× CIVIL RIGHTS			0.112**					
			(0.028)					
SIZE× POL. RIGHTS				0.095**				
				(0.013)				
SIZE× AUTOC					0.116***			
					(0.007)			
SIZE× EXCLUDED						0.112***		
						(0.008)		
SIZE× CHILD MORTALITY							0.006**	
							(0.032)	
SIZE× PUBINDEX								0.035**
								(0.021)
EXCLUDED						0.002		
						(0.305)		
PUBINDEX								0.101
								(0.426)
PRIVINDEX								0.001
								(0.212)
OIL	0.631*	0.647	0.642	0.651	0.673*	0.519*	0.647	
	(0.095)	(0.109)	(0.112)	(0.108)	(0.083)	(0.053)	(0.103)	
GIP		-0.003	-0.003	-0.002	-0.003		-0.002	0.001
		(0.253)	(0.249)	(0.259)	(0.250)		(0.357)	(0.885)
GDP		0.003	0.003	0.003	0.002	0.003**	0.003	0.003
		(0.324)	(0.324)	(0.324)	(0.339)	(0.025)	(0.352)	(0.345)
POP		-0.001	-0.001	-0.001	-0.002	-0.001	-0.001	-0.000
		(0.878)	(0.877)	(0.878)	(0.821)	(0.726)	(0.877)	(0.958)
GROUPAREA		-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
		(0.901)	(0.867)	(0.914)	(0.520)	(0.772)	(0.698)	(0.798)
POLITY		-0.001	-0.001	-0.001	-0.001	0.001	-0.001	-0.001
		(0.516)	(0.516)	(0.516)	(0.514)	(0.396)	(0.456)	(0.447)
SOILCONST		-0.000	-0.000	-0.000	-0.001	-0.001	-0.001	-0.000
		(0.507)	(0.512)	(0.510)	(0.186)	(0.117)	(0.209)	(0.571)
DISTCAP		0.002*	0.002*	0.002*	0.002	0.002***	0.002	0.003*
		(0.089)	(0.090)	(0.088)	(0.128)	(0.001)	(0.118)	(0.091)
MOUNT		0.002	0.002	0.002	0.003*	0.003*	0.003*	0.002
		(0.117)	(0.112)	(0.123)	(0.063)	(0.071)	(0.077)	(0.194)
PARTITIONED		-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
		(0.439)	(0.440)	(0.440)	(0.409)	(0.529)	(0.446)	(0.472)
LAG	0.898***	0.900***	0.900***	0.900***	0.899***	0.898***	0.898***	0.900***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.850	0.855	0.855	0.855	0.852	0.855	0.850	0.853
Obs	41314	38341	38341	38341	39394	45330	36839	35755

Table 3. Group Size and Conflict: Public Prize. This table regresses conflict incidence on group size and indices of public prizes. Alternative proxies are considered for the public prize. All regressions contain year dummies and country fixed effects, and have been estimated by OLS. Robust standard errors clustered at the country level have been computed in all Columns, except in Columns 7 and 8, as EXCLUDED is a group-level variable (results are robust to clustering errors at the country level). **p-values** (based on robust standard errors clustered at the country level) are reported in parentheses. For convenience, the coefficients of SIZE × CHILD MORTALITY are multiplied by 10. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Columns 3 and 4 consider separately the two indices employed to compute our baseline measure LACK RIGHTS. These "subindices" measure the lack of civil rights and the lack of political rights, respectively (see Appendix for definitions). Identical results are obtained.

Columns 5–8 consider alternative proxies of publicness. Columns 5 and 6 use proxies that aim to capture the lack of access to political power. The first proxy (Column 5) is similar to that employed by Esteban et al. (2012a, 2012b) and rests on the idea that there are large gains to seizing power when groups are excluded from it. Specifically, the more "autocratic" a country is, the lower is the sharing of power and the larger the number of citizens/groups thereby excluded. Consequently, the value of controlling the State — for both the democratic and autocratic reasons discussed earlier — is higher. The measure of lack of access to power that we use is AUTOC, a composite measure of autocracy from Polity IV.[32] As in the case of LACK RIGHTS, we *only* consider pre-sample values of the autocracy index. Specifically, AUTOC is computed by averaging the values of the autocracy index from 1960 to 1975, which is then employed in regressions using post 1975 data. The resulting measure is "assigned" to all the ethnic groups in the country, so that AUTOC is a time-invariant country-level index. Moreover, we control for past conflict in all our regressions.

The second proxy (Column 6) considers a group-level measure of publicness, based on whether the group is excluded from State power. We construct an index for exclusion, EXCLUDED, in a similar fashion as AUTOC, i.e., by averaging the values of a yearly dummy for exclusion over the period 1945-1970. Identical results are obtained. The interactions of SIZE and the different prize proxies in Columns 5 and 6 keep their expected signs and are significant.

Column 7 employs CHILD MORTALITY as a public prize proxy; see Appendix for the exact definition and sources. We interpret this index as a measure of low provision of services, and take it as a proxy for a large gain under the democratic struggle interpretation. As before, we only use pre-sample values to compute this measure. Column 7 shows that the interaction of group size and child mortality rates is positive and significant; Finally, Column 8 introduces a composite proxy of publicness, PUBINDEX, obtained by applying factor analysis on five indicators of (pre-sample) lack of public goods: lack of political rights, lack of civil rights, the level of autocracy, group exclusion from central power and child mortality rates. Our conclusions remain unchanged.

In Section 4.2.3 we observed (as a minor remark) how the coefficient on SIZE alone turned positive in the presence of an interaction between group size and the private prize. In similar vein, and with the interaction SIZE × PUB included instead, the coefficient on SIZE alone turns negative. Again,

[32]See Polity IV for details. We renormalize AUTOC to lie between 0 and 1, where 1 indicates the highest degree of autocracy. The Polity IV manual summarizes the index thus: "[We] define [autocracy] operationally in terms of the presence of a distinctive set of political characteristics. In mature form, autocracies sharply restrict or suppress competitive political participation. Their chief executives are chosen in a regularized process of selection within the political elite, and once in office they exercise power with few institutional constraints . . . Our operational indicator of autocracy is derived from codings of the competitiveness of political participation, the regulation of participation, the openness and competitiveness of executive recruitment, and constraints on the chief executive."

we do not intend this observation as any more than a passing comment, as in no case have we controlled for *all* aspects of public and private prizes, but it is nevertheless suggestive.

4.3.4. Private and Public Prizes Together. Recall that Table 3 imposed $\beta_2 = 0$ in (14) so as to exclusively study the interaction of SIZE and publicness. Table 4 frees up β_2 and considers simultaneously the interactions of SIZE with indices of privateness *and* publicness. As before, all specifications in Table 4 contain lagged conflict, country and year fixed effects and are estimated by OLS.

Column 1 shows, with country and year fixed effects but with no other controls, that the interaction of SIZE and OIL is negative, the interaction of SIZE and LACK RIGHTS is positive and that both are significant. Column 2 adds controls, obtaining identical results. Column 3 shows that our results are robust to considering the flexible specification for oil in Table 1. Column 4 switches the private prize to MINES, which is mineral availability in the ethnic homeland. Column 5 switches the private prize again to HOME, the area of the ethnic homeland as a fraction of country area. Column 6 use the above-defined composite indices of privateness (PRIVINDEX) and publicness (PUBINDEX), obtained by applying factor analysis on the different proxies of privateness and publicness employed here and in Tables 1 and Tables 3. Column 7 returns to oil as the private prize, while switching the public prize to the AUTOC variable. Columns 8 and 9 considers EXCLUSION and CHILD MORTALITY as alternative public proxies, as described earlier, while retaining OIL for the private prize.

It is remarkable that very similar results are obtained in all cases. These are robust findings.

In keeping with the analysis in Table 4, we note the obvious: while the data are replete with conflicts over private and public payoffs, the two are sometimes closely intertwined. For instance, even a conflict as seemingly primordial as Rwanda was permeated with economic looting, such as land grabs, under the cover of ethnic violence. The Second Civil War in the Sudan is about different cultural and religious identities, but it is also — to some degree — about oil; so is the Chechnyan War. The Zimbabwean conflict is about identity and political power, but it is also about land, and so on. In the light of these expected complications, it is of interest that the two interaction predictions made by the theory hold up separately and robustly across different variations.

4.3.5. Coefficient Magnitudes. With both private and public interaction effects in place, we are in a position to provide a sense of the magnitudes of the estimated coefficients. We use the estimates from Column 2 from Table 4, which contain our two baseline indices for public and private prizes along with other controls. The estimated marginal effects of group size, coming as they do from interactions, must depend on the values of LACK RIGHTS and OIL, so we give a couple of examples here and refer the reader to Figure for more. For LACK RIGHTS $= 0$ and a high value of oil (at the 95th percentile) an increase of one standard deviation in SIZE decreases the unconditional probability of conflict incidence by 4.9%. Similarly, if OIL $= 0$ and LACK RIGHTS is high ($= 1$), an increase of one standard deviation in SIZE increases the probability of conflict by 7.3%.

	Dependent Variable: **Conflict Incidence**								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
SIZE	-0.002	0.011	0.022	0.004	0.110	0.084**	0.014	0.031	0.012
	(0.915)	(0.656)	(0.412)	(0.893)	(0.118)	(0.019)	(0.587)	(0.305)	(0.656)
SIZE× LACK RIGHTS	0.068*	0.083**	0.083*	0.086*	0.067				
	(0.062)	(0.035)	(0.050)	(0.086)	(0.134)				
SIZE× OIL	-14.455**	-12.836**					-12.554**	-13.350***	-11.696**
	(0.036)	(0.026)					(0.025)	(0.001)	(0.046)
SIZE × OIL_{0-25}			0.033						
			(0.761)						
SIZE × OIL_{25-50}			0.182						
			(0.551)						
SIZE × OIL_{50-75}			-0.166**						
			(0.027)						
SIZE × $\text{OIL}_{>75}$			-0.118***						
			(0.005)						
SIZE× PRIVIND						-0.046**			
						(0.016)			
SIZE× PUBINDEX						0.023*			
						(0.080)			
SIZE × MINES				-0.013*					
				(0.099)					
SIZE × HOME					-0.331**				
					(0.034)				
SIZE × AUTOC							0.097**		
							(0.013)		
SIZE × EXCLUDED								0.098**	
								(0.015)	
SIZE × CHILD MORT.									0.004
									(0.137)
OIL	0.887*	0.828*		0.482	0.515		0.841*	0.710**	0.796*
	(0.062)	(0.069)		(0.251)	(0.165)		(0.052)	(0.018)	(0.074)
$\text{OIL}_{>75}$			0.007**						
			(0.029)						
OIL_{50-75}			0.005*						
			(0.064)						
OIL_{25-50}			-0.002						
			(0.600)						
OIL_{0-25}			-0.003						
			(0.393)						
MINES				0.000					
				(0.938)					
PUBINDEX						0.110			
						(0.386)			
HOME					0.014				
					(0.186)				
EXCLUDED								0.003	
								(0.233)	
POLITY		-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.001	-0.001
		(0.511)	(0.512)	(0.494)	(0.532)	(0.437)	(0.509)	(0.405)	(0.452)
GIP		-0.003	-0.003	-0.002	-0.004	0.001	-0.003		-0.002
		(0.246)	(0.253)	(0.380)	(0.101)	(0.867)	(0.243)		(0.361)
GDP		0.003	0.003	0.005	0.003	0.003	0.002	0.003**	0.003
		(0.324)	(0.320)	(0.199)	(0.346)	(0.356)	(0.340)	(0.025)	(0.352)
POP		-0.001	-0.001	-0.003	-0.001	-0.001	-0.002	-0.001	-0.001
		(0.878)	(0.884)	(0.786)	(0.864)	(0.919)	(0.823)	(0.711)	(0.879)
GROUPAREA		0.000	0.000	0.000		0.000	0.000	0.000	0.000
		(0.289)	(0.338)	(0.340)		(0.340)	(0.684)	(0.297)	(0.582)
SOILCONST		-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001*	-0.001
		(0.399)	(0.358)	(0.264)	(0.333)	(0.387)	(0.121)	(0.067)	(0.153)
DISTCAP		0.003*	0.002*	0.002	0.003*	0.003*	0.002	0.002***	0.002
		(0.085)	(0.085)	(0.123)	(0.074)	(0.088)	(0.121)	(0.001)	(0.112)
MOUNT		0.002	0.002	0.002	0.002	0.002	0.003*	0.003*	0.003*
		(0.107)	(0.100)	(0.191)	(0.124)	(0.168)	(0.057)	(0.061)	(0.071)
PARTITIONED		-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
		(0.444)	(0.431)	(0.693)	(0.431)	(0.477)	(0.411)	(0.531)	(0.450)
LAG	0.898***	0.900***	0.900***	0.890***	0.900***	0.900***	0.899***	0.898***	0.898***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.850	0.855	0.855	0.843	0.855	0.853	0.852	0.855	0.850
Obs	41314	38341	38341	32026	37659	35755	39394	45330	36839

Table 4. Group Size and Conflict: Private and Public Prize Specifications. This table regresses conflict incidence on group size and indices of public prizes. All regressions contain year dummies and country fixed effects, and have been estimated by OLS. Robust standard errors clustered at the country level have been computed in all Columns. **p-values** (based on robust standard errors clustered at the country level) are reported in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

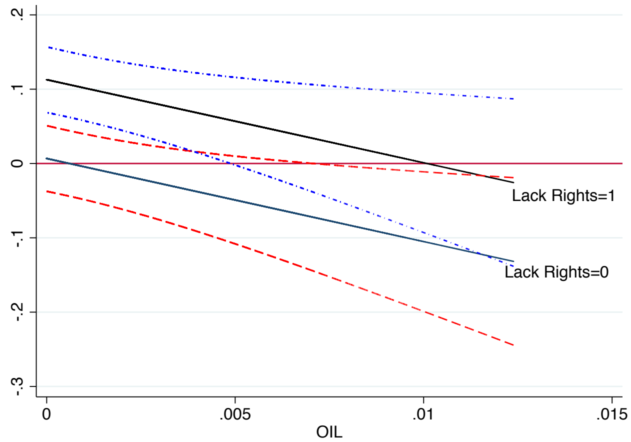


Figure 3. Marginal Effects of Group Size on Conflict. This graph depicts the marginal effect of group size on conflict incidence as a function of OIL for two different values of LACK RIGHTS: 0 (bottom solid line) and 1 (top solid line). Confidence bands at the 90% level are also depicted. Estimates from Table 4 (Column 2) have been employed to construct the graphs.

We've seen that the presence of oil attenuates the effect of size on conflict. But as already mentioned, the theory makes a sharper prediction: the marginal effect of SIZE actually turns *negative* as the prize becomes increasingly private. (The opposite is true when the prize is public.) The two examples in the previous paragraph are provided with this in mind. More generally, Figure 3 plots the marginal effect of SIZE on INCIDENCE computed using the estimates from Column 2 in Table 4. The marginal effect is a function of both OIL and LACK RIGHTS, and the plot displays this marginal effect as a function of OIL (on the x-axis), for the minimum and maximum values of LACK RIGHTS (i.e., LACK RIGHTS=$\{0,1\}$). The dashed lines represent 90% confidence bands. In line with the theory, the figure shows that the marginal effect of size is negative or positive, depending on the values of the public and private payoffs. For a small value of LACK RIGHTS and moderate or large values of OIL, an increase in group size has a negative and significant effect on conflict incidence. The opposite is true when LACK RIGHTS is high and OIL is small; now the marginal effect of SIZE is positive and significant. However, it is not significantly different from zero when either both prizes are small or when both are large. Table B.1 in the Online Appendix makes these points in an even simpler way by taking binary cuts for private and public prizes and comparing the four cells that result.

Recall that our baseline specification imposes the restriction that the marginal effect of SIZE on conflict is a linear function of oil. One might therefore respond that the above observation stems from that assumed linearity. However, Columns 3 and 4 in Table 4 show that similar results are found when a more flexible specification, one in which linearity is not imposed, is employed. Recall that these Columns employ four dummies that correspond to the quartiles of the distribution of OIL for the groups that have oil in their homeland (thus, the omitted category corresponds to groups that do not have oil). In this case, the marginal effect of SIZE on conflict is given by the sum of the coefficient of SIZE and that of the variables SIZE$\times$ OIL$_j$, where OIL$_j$ is equal to 1 if the group's oil is in quartile j. The marginal effect of SIZE on conflict is positive and significant in the absence of oil, but the effect decreases as the amount of oil in the homeland becomes more abundant and it eventually becomes negative for groups with abundant oil reserves. In particular, using the results in Column 4, we can reject at the 1% level the assertion that the sum of the coefficient of SIZE and that of SIZE× OIL_{50-75} (or SIZE× $\text{OIL}_{\geq 75}$) is greater than or equal to zero.[33] So this effect is not driven by merely extrapolating a linear specification.

This negative relationship between size and conflict when the prize is private — transforming itself into a positive relationship as the prize turns public — is a central finding of both the theory and the empirical analysis. Section 5.1 argues that it is hard to think of any alternative explanation that would generate the same joint pattern.

5. Variations

The evidence so far shows a robust link between the probability of conflict, group size and the nature of the payoffs. In this section, we consider several variations on the baseline exercises. First, we consider alternative explanations that could rationalize our empirical findings and provide evidence against them (Section 5.1). Second, following up on the theory in Section 3.3, we study the implications of multiple threats to conflict (Section 5.2). In similar vein, we address the possibility that groups could form alliances in conflict (Section 5.3). We then go on to consider alternative measures of conflict (Section 5.4), other ways of measuring oil wealth (Section 5.5), and the possibility that our results are due to omitted variables more generally, finding little support for this (Section 5.6). We end with some statistical variations: two-way clustering of standard errors instead of country- or group-level clustering (Section 5.7), the use of a nonlinear model; specifically logit (Section 5.8), and robustness to dropping different regions of the world (Section 5.9). Some of the tables are provided in the Online Appendix.

5.1. Alternative Explanations. One possible interpretation of our results invokes differences in conflict technology rather than the structure individual payoffs generated by group sizes and the nature of the prize. According to this view, large groups (with or without oil) would have easier access to the funds needed to engage in conflict against the State. However, small groups would find it particularly useful to have oil in their homeland to purchase weapons, hire mercenaries, etc., which otherwise would be beyond their means. As in the case of our theory, this explanation would generate a heterogeneous impact of group size on conflict: small groups would tend to fight less than large groups, unless they have oil. However, this alternative explanation would fail to generate the *negative relationship* between size and conflict shown in Tables 1 and 4 and in Figure 3. While the effect of size on conflict would be attenuated by oil, the net effect of group size must always remain positive.

Another interpretation of our findings might emphasize the potential confounding role of oil concentration. When a small ethnic group has oil in its homeland, then it is likely, controlling for

[33] Similar results are also true if the estimates in Column 7 (with a moderate value of LACK RIGHTS, as suggested by the theory) are considered.

	Dependent Variable: **Conflict Incidence**							
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
SIZE	0.065***	0.014	0.065***	0.014	0.142***	0.033	0.077***	0.025
	(0.006)	(0.576)	(0.005)	(0.578)	(0.002)	(0.602)	(0.001)	(0.339)
SIZE × OIL	-15.007**	-14.491**	-16.306*	-17.488	-26.752***	-18.602*	-17.060***	-16.350**
	(0.016)	(0.027)	(0.063)	(0.112)	(0.000)	(0.063)	(0.000)	(0.016)
SIZE × LACK RIGHTS		0.081**		0.077**		0.199*		0.094**
		(0.033)		(0.045)		(0.094)		(0.027)
OIL	0.794*	0.900*	0.768*	0.829*	0.908**	1.096**	0.925***	0.980**
	(0.061)	(0.050)	(0.071)	(0.077)	(0.011)	(0.027)	(0.004)	(0.019)
OIL CONCENTRATION	-0.014	0.013						
	(0.158)	(0.304)						
OIL (SHARE)			0.001	0.003				
			(0.820)	(0.649)				
GIP	-0.003	-0.002	-0.003	-0.002			-0.004*	-0.003
	(0.141)	(0.304)	(0.127)	(0.306)			(0.068)	(0.208)
GDP	0.000	0.003	0.001	0.003	0.002*	0.003	0.002*	0.004
	(0.790)	(0.252)	(0.393)	(0.289)	(0.059)	(0.251)	(0.066)	(0.173)
POP	0.001	-0.000	0.001	-0.000	-0.002	0.000	-0.000	-0.005
	(0.822)	(0.985)	(0.797)	(0.968)	(0.632)	(0.991)	(0.879)	(0.623)
GROUPAREA	0.000	0.000	0.000	0.000	0.000	-0.000	0.000	0.000
	(0.288)	(0.247)	(0.268)	(0.256)	(0.689)	(0.555)	(0.296)	(0.270)
SOILCONST	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.001
	(0.645)	(0.588)	(0.636)	(0.590)	(0.366)	(0.850)	(0.518)	(0.416)
DISTCAP	0.002*	0.002*	0.002*	0.002*	0.002***	0.003*	0.002***	0.002*
	(0.098)	(0.084)	(0.099)	(0.085)	(0.000)	(0.063)	(0.000)	(0.092)
MOUNT	0.002**	0.002	0.002**	0.002	0.002	0.001	0.002	0.002
	(0.034)	(0.125)	(0.033)	(0.116)	(0.292)	(0.373)	(0.129)	(0.160)
PARTITIONED	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.282)	(0.546)	(0.280)	(0.542)	(0.303)	(0.328)	(0.219)	(0.650)
LAG	0.894***	0.900***	0.893***	0.900***	0.896***	0.900***	0.889***	0.895***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
C	-0.010	-0.036	-0.033	-0.031	-0.011	-0.046	-0.013	0.024
	(0.917)	(0.734)	(0.699)	(0.777)	(0.836)	(0.790)	(0.794)	(0.854)
R^2	0.846	0.854	0.846	0.854	0.860	0.862	0.839	0.847
Obs	57559	39992	57559	39992	41158	34198	49892	34647

Table 5. Variations: Oil Concentration, Excluded Groups and Small Ruling Elites. This table regresses conflict incidence on group size and indices of private and public prizes, along with interactions between subsets of these variables as suggested by the theory. OIL CONCENTRATION is computed as the Herfindahl index of oil reserve distribution across groups. OIL SHARE is the share of oil in the homeland of the group. To compute Columns 5 and 6, excluded groups have been dropped from the sample. Excluded groups are those with a value of (pre-sample) EXCLUDED larger than 0.5. To compute Column 7, country/years in which the size of the ruling elite in autocraticies is small (as compared to the ruling elite in non-autocracies) have been dropped from the sample (see the main text for details). All regressions contain year dummies and country fixed effects. **p-values** (based on robust standard errors clustered at the country level) are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

total oil reserves, multiple ethnic homelands have access to oil rather than a single homeland. That in itself could lead to violence. To the extent that the distribution of oil reserves in the country is stable over time, the country fixed effects included in all our specifications will partly eliminate this effect. Nevertheless, as our regressions consider a large time span, new oil discoveries could change distributional patterns within the country. Columns 1 and 2 in Table 5 add to our baseline specifications (Column 3 in Table 1 and Column 2 in Table 4) a (time-varying) country-level oil concentration index. We do not find a significant effect of concentration and this variable does not change the interaction effect of group size and oil: the coefficients and their significance are very similar as in the baseline table.

Morelli and Rohner (2015) study the relationship between conflict and the *concentration* of natural resources in ethnic homelands. They show that the larger a group's share of oil, the larger the probability of conflict onset. This is a completely different prediction from ours; it is orthogonal to what we do. That said, and because bigger groups are more likely to have a larger share of national oil, we check that oil share is not a confounder in our regressions. Columns 3 and 4 add to our baseline specifications the share of oil as computed by Morelli and Rohner (i.e., the surface of an ethnic group's territory covered with oil and gas as a percentage of total country surface covered with oil and gas). Oil share is far from being significant in these specifications. Our conclusions survive unchanged, although the coefficient of SIZE × OIL is estimated more noisily in Column 4 (p-value is .11), when the public prize interaction is also in the regression.[34]

Our theoretical results stress the fact that if the initial allocation is non-discriminatory — or if it is biased *against* small groups — then the latter are more likely to be involved in conflict if the payoff is private. An alternative interpretation, however, would run as follows: it's reasonable to think that large groups are stronger and, as a result, more likely to be in power. Thus, they can get a disproportionately large share of the rents from the center, and therefore they are less likely to rebel. Although this argument is clearly related to ours, the underlying mechanism is somewhat different: small groups rebel because they are more likely to be excluded from power and, as a result, they do badly under the initial allocation. This alternative explanation is, however, at odds with our empirical results: in that case we should see small groups fighting more, not just on account of oil, but simply because they are treated worse. But there is no evidence of that at all: in *all* our specifications, group size per se is either insignificant or positive whenever significant.

Absent a direct measure of the initial allocation, we do control in all our regressions for whether a group is included or excluded from political power. In addition, we have also considered whether our results are robust when only *excluded* groups are considered.[35] Columns 5 and 6 in Table 5 restrict the sample to groups excluded from power and shows that our conclusions continue to hold when only those groups are in the sample.

Or it could be argued that the index LACK RIGHTS is typically high in authoritarian regimes, which often tend to have a ruling elite made up of minorities. So it would be possible that in those regimes conflict is initiated by majorities that want to take over power from these minorities. Indeed, the average size of the group(s) in power in autocracies is smaller than in less autocratic regimes. To rule out this possibility, we have dropped from the sample countries whose ruling elites are small (as compared to ruling elites in non-autocratic countries). More specifically, we have divided

[34] Morelli and Rohner (2015) consider onset rather than incidence. We've also checked that in onset regressions the significance of the share of oil vanishes once one controls for total group oil, but again, this is not our focus.

[35] More specifically, we drop from the sample groups with a value of EXCLUDED (a pre-sample average of an exclusion dummy over the years 1960–1975) larger than 0.5. Results are very robust to other ways of defining exclusion.

the sample into autocratic and non-autocratic countries (defined as those with autocracy index higher/lower than 5) and we have dropped from the sample autocratic countries where the size of the ruling elite is smaller than the median of the size of the group(s) in power in non-autocracies. Then, we've re-run our baseline specification with this reduced sample. Our results remain robust to this variation, see Columns 7 and 8 in Table 5.

Finally, one could, of course, posit something with no particular conceptual foundation: that oil is somehow special for smaller groups, or that small groups have a better conflict technology, or that small groups are particularly fond of secession and so fight harder for their freedom. However, these *ad hoc* alternatives must all contend with the simple observation that we explore in Columns 9-11 of Table 2: that controlling for the *per-capita value* of a private prize, group size is positively related to conflict. None of these explanations can simultaneously explain why the relationship turns positive when per-capita controls for private wealth are imposed.

The one reasonable (and related) argument that does generate a negative relationship between group size and conflict in the presence of private goods is the free-rider argument first described by Olson (1965). Small groups are better capable of cohesion. As already discussed, this argument complements the one based on per-capita payoffs first described by Pareto (1827), that we emphasize in our theory. The problem is that the free-rider theory works well for one side of our observations but not the other. It would have no prediction for group size and *public* prizes, where the observed relationship is positive.

Apart from the evidence against the above arguments, it is to be noted that they pertain to pieces of our main result and not the entire prediction. The negative relationship between size and conflict under a private prize — morphing into a positive relationship as the prize turns public — is our central finding. It is not easy to think of an alternative explanation with the same *joint* pattern.

5.2. Multiple Threats and Conflict. Because fighting is costly, individual threats can always be appeased provided that appropriate Coaseian transfers are available. However, Section 3.3 shows that in conflicts over private payoffs the existence of a variety of potential threats might make conflict unavoidable. This is so because it might be impossible to find a set of transfers that *simultanously* appeases all threats. The empirical implication of this result is that, in the presence of private payoffs, we should observe a positive relationship between conflict and the number of groups that can threaten peace, where these groups can be defined along several dimensions (ethnicity, class, geography, etc). Without pretending to explore this issue in depth here — as we lack information on group affiliation on dimensions other than ethnicity — our empirical setting allows us to partially examine this prediction. Because attention is restricted to groups defined along ethnic lines, we proxy the multiplicity of threats by an ethnic fractionalization index. Then, we explore whether groups in more ethnically fragmented countries are more likely to be involved in conflicts over private payoffs.[36] Results are presented in Table 6. Column 1 introduces the

[36]The fractionalization index measures the probability that two individuals chosen at random in country j belong to two different ethnic groups. The polarization index is a measure of social "antagonism," and its key ingredients

interaction of an index of fractionalization, FRAC, and OIL in a regression of conflict incidence on all the controls. The interaction is positive and significant, as predicted by the theory.[37] This conclusion is robust to introducing in the regression the interactions of SIZE and OIL (Column 2) and SIZE and LACK RIGHTS (Columns 3).[38] For comparison, Columns 4–6 reestimate Columns 1–3 replacing FRAC × OIL by the interaction of a polarization index, POL, and OIL. Our theory doesn't have any prediction for this interaction, which turns out to be insignificant in our regressions.

5.3. Alliances in Conflict. In some cases, *alliances* of groups could form. For instance, in the First Sudanese Civil War, also known as the Anyanya Rebellion, a conglomeration of the Acholi, Bari, Dinka, Lotuko, Madi, Nuer and the Zande from South Sudan came together, albeit in an alliance marked by substantial infighting. Other alliances are not hard to find: e.g., ethnic alliances exist in the Casamance conflict in Senegal or in the Liberian war that toppled the Taylor government.

As already described, the data we use code ethnic groups in conflict against the State. In the case of alliances, *each* ethnic group is so coded. As expected, the dataset has a number of such conflicts. Now, several of these conflicts are genuinely separate conflicts, and some are not. It is unclear how one might approach this problem comprehensively without running into severe issues of endogeneity in the definition of a "coalition."

Without pretending to satisfactorily solve this dilemma, one can run a rough variant of our exercise by mechanically combining all multiple instances of conflict. Table B.3 presents regressions with public and private prizes along the lines of the tables above but uses an alternative definition of group size, $\text{SIZE}_{\text{COAL}}$. This variable is defined as follows: for peace years, $\text{SIZE}_{\text{COAL}}$ and SIZE coincide. For years where *some* group is in conflict, $\text{SIZE}_{\text{COAL}}$ adds up the size of *all* the groups in conflict in the same country and year. In this way we try to capture the possibility that there exists an alliance between the fighting groups. The variable OIL_{COAL} is defined in a similar way: in peace years, OIL and OIL_{COAL} are identical. In case of conflict, $\text{SIZE}_{\text{COAL}}$ adds up the oil in the homelands of all the groups in conflict in that country and year. Our conclusions are robust to this variation.

5.4. Alternative Measures of Conflict. Table B.4 considers alternative measures of conflict: Columns 1–3 use conflict *onset* in a panel set-up, while Columns 4–7 drop the time dimension of the data. In these Columns the dependent variables are the share of years over 1975–2006 in which a group has been involved in conflict against the State (Columns 4–6), and the share of onset years (Column 7). Qualitatively, the results are very similar to those described above. The interactions of group size and the publicness/privateness indicators have the predicted sign and are highly significant. There is only one exception: when the dependent variable is conflict

are intergroup distances (how alien groups are from each other) and group size (an indicator of the level of the group identification). Both indices are computed using data from Fearon (2003), and linguistic distances are employed to capture intergroup distances, see Esteban et al. (2012) for details on their computation.

[37] This result is closely related to Esteban et al. (2012), who show that more fragmented countries tend to fight more on account of oil. The results in Table 6 show that a similar result also holds when considered at the ethnic group level.

[38] Similar results are found if other ways of defining the private payoff are considered, such as oil at the country rather than at the group level or the index of private payoffs introduced in Table 1.

	Dependent Variable: **Conflict Incidence**					
	[1]	[2]	[3]	[4]	[5]	[6]
SIZE	0.041**	0.071***	0.016	0.022	0.072***	-0.005
	(0.030)	(0.001)	(0.553)	(0.272)	(0.003)	(0.887)
FRAC × OIL	3.267***	2.979***	2.879*			
	(0.000)	(0.001)	(0.056)			
POL × OIL				1.645	1.131	3.872
				(0.735)	(0.814)	(0.494)
SIZE × OIL		-10.137**	-8.187		-16.410***	-14.312**
		(0.014)	(0.230)		(0.000)	(0.034)
SIZE × LACK RIGHTS			0.081*			0.125**
			(0.061)			(0.016)
OIL	-1.304***	-0.991*	-0.925	0.609*	0.879**	0.767
	(0.009)	(0.074)	(0.427)	(0.081)	(0.018)	(0.211)
POLITY	-0.002**	-0.002**	-0.001	-0.002***	-0.002***	-0.001
	(0.017)	(0.017)	(0.506)	(0.009)	(0.009)	(0.268)
GIP	-0.004**	-0.004**	-0.003	-0.003	-0.003*	-0.002
	(0.032)	(0.029)	(0.205)	(0.110)	(0.099)	(0.518)
GDP	0.001	0.001	0.003	-0.000	-0.000	0.001
	(0.142)	(0.148)	(0.343)	(0.794)	(0.781)	(0.863)
POP	0.001	0.001	-0.001	0.005*	0.005	0.009
	(0.772)	(0.800)	(0.881)	(0.099)	(0.116)	(0.452)
GROUPAREA	0.000	0.000	0.000	-0.000	0.000	0.000
	(0.516)	(0.114)	(0.192)	(0.598)	(0.369)	(0.240)
SOILCONST	-0.000	-0.000	-0.000	-0.001	-0.001	-0.001
	(0.482)	(0.382)	(0.511)	(0.478)	(0.379)	(0.469)
DISTCAP	0.002***	0.002***	0.003*	0.002***	0.002***	0.003*
	(0.000)	(0.000)	(0.083)	(0.000)	(0.000)	(0.088)
MOUNT	0.002	0.002	0.002	0.003*	0.003**	0.003**
	(0.130)	(0.113)	(0.177)	(0.052)	(0.042)	(0.045)
PARTITIONED	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.350)	(0.344)	(0.482)	(0.168)	(0.169)	(0.283)
LAG	0.894***	0.894***	0.900***	0.896***	0.896***	0.903***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
C	-0.020	-0.016	-0.017	-0.058	-0.053	-0.173
	(0.640)	(0.712)	(0.890)	(0.363)	(0.408)	(0.419)
R^2	0.847	0.847	0.855	0.851	0.851	0.857
Obs	54762	54762	37957	45362	45362	29387

Table 6. Private Payoffs, Fractionalization, Polarization and Conflict. This table regresses conflict incidence on group size, indices of private and public prizes, along with interactions of a Fractionalization index (FRAC) and OIL (Columns 1–3) as well as a Polarization index (POL) and OIL (Columns 4–6). All regressions contain year dummies and country fixed effects. FRAC and POL are country-level indices of ethnic fractionalization and polarization, respectively. **p-values**, based on standard errors clustered at the country level, are reported in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

onset and the two interactions are in the regression (Column 3), both interactions maintain their predictive signs but become insignificant. It is important to notice, however, that the number of onset observations in the post 1975 sample employed in that regression is only 130 (0.35% of all observations). In contrast, in the same period there are 1,680 conflict observations, around 4.5% of all observations. This implies that onsets in the post 1975 sample are rare events. As a result coefficients are estimated with considerable error. Also notice that both interactions are significant in the cross-section, where the dependent variable is the share of onset years (Column 7).

5.5. Alternative Ways of Measuring Oil Wealth. Table B.5 considers alternative ways of measuring oil wealth. Columns 1–3 consider oil rents at the national level, rather than just the oil that lies in the homeland of the group. Column 1 shows that the same results on the interaction between group size and oil goes through for national oil rents. That said, the level effects suggest clearly that it is resources under the spatial control of a particular group that are highly linked to conflict involving *that group*. Group oil matters for conflict involving that group; controlling for that, national oil does not. Moreover, Columns 1 and 2 show that, once the interaction between group oil and group size is introduced in the regression, it is significant, while the corresponding interaction between group size and national oil falls silent. This suggests that, although our conclusions are robust to considering national oil rents, group-level oil seems to be a better proxy for the private payoff. Columns 4 and 5 use an alternative measure of oil abundance, OIL (SURFACE), based on the ethnic homeland's area covered by oil. Identical conclusions are obtained.

5.6. Assessing the Importance of the Omitted Variable Bias. Despite our attempts to control for a large number of potential confounders, we still cannot completely rule out the possibility that unobserved variables are biasing our results. However, it is possible to assess the likelihood that our observed effect is solely due to selection bias. To that effect, we apply a technique recently developed by Oster (2017), which builds on the work by Altonji, Elder and Taber (2005) and Bellows and Miguel (2008). This method allows to determine the degree of selection on unobservables relative to observables (denoted by δ) that would be necessary to explain away the result. If the set of observed controls is representative of all possible controls, then a large value of δ suggests that it is implausible that omitted variable bias explains away the entire effect. Altonji et al. (2005) and Oster (2017) suggest the use of a cut-off of δ (δ^*) equal to 1. This value means that selection on unobservables would need to be at least as important as that on observables to produce a treatment effect of zero. Thus, if for example a value of $\delta = 2$ is obtained, it would suggest that the unobservables would need to be twice as important as the observables to produce a treatment effect of zero. One reason to favor the cut-off $\delta^* = 1$ is that researchers typically choose the controls they believe ex ante are the most important (Angrist and Pischke, 2010) and thus situations where selection on unobservables is larger than that of the observed controls are deemed unlikely.

The statistic employed to compute δ is designed to evaluate the stability of the variable(s) of interest to the introduction of controls. More specifically, it is a function of the coefficient of the variable of interest estimated in a full model (that contains all controls), the same coefficient obtained in a restricted model with no (or few) controls, the R^2s obtained in these regressions and $R^2_{\max}$, the R^2 from a hypothetical regression of the outcome on treatment and both observed and unobserved controls. If the outcome can be fully explained, then $R^2_{\max} = 1$. However, as acknowledged by

Oster (2017), in most empirical settings it seems likely (due, for example, to measurement error) that the outcome cannot be fully explained even if the full control set is included.

In our case, the variables of interest are either SIZE × OIL or SIZE × LACK RIGHTS. The full model corresponds to the specification containing both interactions and the full list of controls (Column 2 in Table 4). Since our variables of interest are interactions, the restricted model is one where the only controls are the variables included in the corresponding interaction.[39] The values of δ are quite sensitive to the choice of the (unobserved) value of $R^2_{\max}$. Then, we have computed the maximum value of $R^2_{\max}$ we can use in order to obtain values of δ larger than 1. This value turns out to be quite large (around 0.93 for both interactions). Since the variables employed in our regressions are clearly not perfectly measured, we believe that this is a reasonable value for the maximum R^2 that can achieved given the quality of the available data. Therefore, we argue that the latter value of $R^2_{\max}$ gives support to our claim that omitted variable bias is not likely to explain our results.

5.7. Alternative Clustering Strategies. As explained in Section 4.2.2, throughout the paper we follow Abadie et al (2017) and cluster errors at the group or at the country level, depending on whether the corresponding treatment is assigned at the group or at the country level. Our results are firmly robust to other clustering strategies; for instance, to two-way clustering. Table B.6 recomputes our baseline regressions but this time standard errors are adjusted for clustering at the ethic homeland *and* at the country level. Notice that since ethnic homelands are often split by an international border, the latter dimensions are not nested. Our conclusions are robust to considering alternative clustering schemes.

5.8. Nonlinear Models. Since our baseline dependent variable is binary, we have re-estimated our baseline specifications using a logit specification. Table B.2 in the Online Appendix presents the results. All equations contain the full list of controls as well as country and year fixed effects, but differ on the interactions included in them: Column 1 includes the interaction of SIZE and OIL, Column 2 that of SIZE and LACK RIGHTS, while Column 3 considers both of them. The coefficients of the interactions of SIZE and the public and private payoffs maintain the expected signs and remain significant. In nonlinear specifications, however, one has to be cautious when interpreting the change in two interacted variables, as Ai and Norton (2003) pointed out. The Online Appendix discusses this issue in more detail and shows that our conclusions still hold when nonlinear estimation is considered.

5.9. Dropping Regions of the World. Table B.7 drops observations from particular regions of the world: former USSR countries (Columns 1 and 2), Asia (Columns 3 and 4), Middle East (Columns 5 and 6), West-South Africa (Columns 7 and 8), East and Central Africa (Columns 9 and 10), and Latin America (Columns 11 and 12). For each region, the first (second) Column replicates Column

[39] That is, SIZE and OIL or SIZE for the private and public interactions, respectively.

4 (5) in Table 1; i.e., we consider specifications with and without the interaction of group size and the public prize. Results are generally robust. The only exception is when East and Central Africa observations are dropped. In this case, the interaction of SIZE and OIL is significant in Column 9, but when the interaction of SIZE and LACK RIGHTS is introduced as well (Column 10), it ceases to be so (while the p-value of the public interaction is 0.14). To put this result in perspective, however, notice East and Central Africa is by far the most conflictual region in our sample: it represents 15% of all observations but 30% of all conflict observations. Thus, it is not surprising that the coefficients are estimated less precisely when a large number of observations is dropped.

6. Conclusion

In the introduction to his essay, "On Liberty," John Stuart Mill (1859) writes:

"Society . . . practices a social tyranny more formidable than many kinds of political oppression, since, though not usually upheld by such extreme penalties, it leaves fewer means of escape, penetrating much more deeply into the details of life, and enslaving the soul itself. Protection, therefore, against the tyranny of the magistrate is not enough; there needs protection also against the tyranny of the prevailing opinion and feeling, against the tendency of society to impose, by other means than civil penalties, its own ideas and practices as rules of conduct on those who dissent from them . . . "

Mill is referring to the tyranny of the majority, a notion that also appears in the writings of John Adams and in the Federalist Papers, in the 18th century, and then amplified and used more extensively by Alexis de Tocqueville (1835). Arrayed against this distinguished company are Wilfredo Pareto and Mancur Olson, who emphasize the power of *minorities* to cohere around a cause. We discussed their contributions in detail above.

In this paper we've studied a model of social conflict, in which the conflict may be over a *public* or a *private* good. The main result, that we explore empirically through a variety of specifications, is that conflict is more likely in the presence of a private prize when the group is small, and it is more likely in the presence of a public prize when the group is large. By using a global panel dataset at the ethnic group level we find powerful and robust empirical support for these claims. This is our approach to reconciling Tocqueville-Mill with Pareto-Olson.

Our approach can be extended to other questions of interest. Specifically, as already hinted at in Section 3.3, one can develop a theory of conflict in which there are *multiple* potential threats to peace from different groups, thereby generating conflict "in equilibrium" even in the presence of inefficiencies. Because a multiplicity of groups are typically formed using ethnic markers, such a theory could also help us understand why ethnic conflict might be salient.[40]

[40] Other factors that bear on the salience of ethnic violence includes the greater visibility of ethnicity (Caselli and Coleman 2013), or the ability of an economically unequal ethnic group to to exploit the synergy of money and labor when engaging in conflict (Esteban and Ray 2008).

References

Abadie, A., S. Athey, G. W. Imbens and J. Wooldridge (2017), "When Should You Adjust Standard Errors for Clustering?", Mimeo.

Ai, C. R. and E. C. Norton (2003), "Interaction terms in logit and probit models," *Economics Letters* **80**: 123–129.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy* **80**:151–184.

Angrist, J. D., and J. Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, **24**: 3-30.

Baliga, S. and T. Sjostrom (2004), "Arms Races and Negotiations, *Review of Economic Studies* **71**, 351–369.

Bannon, I. and P. Collier, eds. (2003), *Natural Resources and Violent Conflict*, Washington D.C.: The World Bank.

Bellows, J., and E. Miguel (2009), "War and Local Collective Action in Sierra Leone". *Journal of Public Economics* **93**:1144?1157.

Berman N., M. Couttenier, D. Rohner and M. Thoenig (2015), "This mine is mine! How minerals fuel conflict in Africa," Mimeo.

Besley T. and T. Persson (2011), "The logic of political violence," *Quarterly Journal of Economics* **126**: 1411–1445.

Bester, H. and K. Wärneryd (2006), "Conflict and the Social Contract," *Scandinavian Journal of Economics* **108**, 231–249.

Blundell R. and S. Bond (1998), "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics* **87**: 115–143.

Bondareva, O. N. (1963), "Some applications of linear programming methods to the theory of cooperative games", *Problemy Kybernetiki* **10**: 119–139.

Bruk, S., and V. S. Apenchenko (eds) (1964), *Atlas Narodov Mira [Atlas of the Peoples of the World]*. Moscow: Glavnoe Upravlenie Geodezii i Kartogra.

Cederman, L.-E., H. Buhaug and Rod, J. K. (2009), "Ethno-Nationalist Dyads: A GIS-Based Analysis," *Journal of Conflict Resolution* **53**: 496-525.

Chamberlin, J. (1974), "Provision of Collective Goods as a Function of Group Size," *American Political Science Review* **68**: 707–716.

Doyle, M. and N. Sambanis (2006), *Making War and Building Peace*. Princeton: Princeton University Press.

Dube, O. and J. Vargas (2013), "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," *Review of Economic Studies* **80**, 1384–1421.

Esteban, J. and D. Ray (1999), "Conflict and Distribution," *Journal of Economic Theory* **87**: 379–415.

Esteban, J. and D. Ray (2008), "On the Salience of Ethnic Conflict," *American Economic Review* **98**: 2185–2202.

Esteban, J. and D. Ray (2001), "Collective Action and the Group Size Paradox," *American Political Science Review* **95**: 663–672.

Esteban, J. and D. Ray (2011), "Linking Conflict to Inequality and Polarization," *American Economic Review* **101**: 1345–74.

Esteban, J., Mayoral, L. and D. Ray (2012),"Ethnicity and Conflict: An Empirical Investigation," *American Economic Review* **102**: 1310–1342.

Fearon, J. (1995), "Rationalist Explanations for War," *International Organization* **49**, 379–414.

Fearon, J. (2003), "Ethnic and cultural diversity by country," *Journal of Economic Growth* **8**:195-222.

Fearon, J. (2005), "Primary Commodity Exports and Civil War," *Journal of Conflict Resolution* **49**, 483–507

Fearon, J. and D. Laitin (2003), Ethnicity, Insurgency, and Civil War," *American Political Science Review* **97**: 75-90.

Fischer, G., F. Nachtergaele, S. Prieler, H. van Velthuizen, L. Verelst, and D. Wiberg (2008), *Global Agro-ecological Zones Assessment for Agriculture*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

François, P., I. Rainer, and F. Trebbi (2015), "How Is Power Shared in Africa?", *Econometrica* **83**, 465–503.

Freedom House (2014), *Freedom of the World*, dataset, https://freedomhouse.org/report/freedom-world-aggregate-and-subcategory-scores.

Garfinkel, M. and S. Skaperdas (2000), "Conflict Without Misperceptions or Incomplete Information: How the Future Matters," *Journal of Conflict Resolution* .**44**, 793–807

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002), "Armed Conflict 1946-2001: A New Dataset," *Journal of Peace Research* **39**: 615–637.

IntierraRMG (2015), "SNL Metals and Mining", dataset, http://www.snl.com/Sectors/metalsmining/Default.aspx.

Jackson, M. and M. Morelli (2007), "Political Bias and War," *American Economic Review* **97**, 1353–1373.

Kirshner, J. (2000), "Rationalist Explanations for War?," *Security Studies* **10**: 143–150.

Le Billon, P. (2001), "The Political Ecology of War: Natural Resources and Armed Conflicts," *Political Geography* **20**, 561–584.

Licklider, R. (1995), "The Consequences of Negotiated Settlements in Civil Wars, 1945-1993," *American Political Science Review* **89**: 681-690.

Lujala, P. (2010), "The Spoils of Nature: Armed Civil Conflict and Rebel Access to Natural Resources," *Journal of Peace Research* **47**, 15–28.

Lujala, P., J. K.Rod and N. Thieme (2007), "Fighting over Oil: Introducing a New Dataset," *Conflict Management and Peace Science* **24**: 239-56.

McGuire, M.C. (1974), "Group Size, Group Homogeneity and the Aggregate Provision of a Pure Public Good under Cournot Behavior," *Public Choice* **18** (Summer): 107–126.

Marwell, G. and P. Oliver (1993), *The Critical Mass in Collective Action. A Micro-Social Theory*. Cambridge, UK: Cambridge University Press.

Mill, J. S. (1859), *On Liberty*, London: Parker and Son.

Morelli, M. and D. Rohner (2015), "Resource Concentration and Civil Wars," *Journal of Development Economics* **117**: 32–47.

Oliver, P. and G. Marwell (1988), "The Paradox of Group Size in Collective Action: A Theory of the Critical Mass. II," *American Sociological Review* **53**: 1–8.

Olson, M. (1965), *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Oster, E. (2017), "Unobservable Selection and Coefficient Stability: Theory and Evidence". *Journal of Business and Economic Statistics*, June 2017.

Pareto, V. [1906] (1927), *Manual of Political Economy*. New York: A.M. Kelley.

Polity IV (2014), "Political Regime Characteristics and Transitions", dataset, http://www.systemicpeace.org/inscrdata.html.

Powell, R. (2004), "The Inefficient Use of Power: Costly Conflict with Complete Information," *American Political Science Review* **98**, 231–241.

Powell, R. (2006), "War as a Commitment Problem," *International Organization* **60**, 169–203.

Religion and State Project ARDA, Round 2, Dataset: http://www.thearda.com/ras/.

Sambanis, N. (2001), "Do Ethnic and Non-Ethnic Civil Wars Have the Same Causes? A Theoretical and Empirical Enquiry," Journal of Conflict Resolution **45**: 259-282.

Sánchez-Pagés, S. (2009), "Conflict as Part of the Bargaining Process," *Economic Journal* **119**, 1189–1207.

Sandler, T. (1992), *Collective Action: Theory and Applications*. Ann Arbor, MI: University of Michigan Press.

Scarf, H. (1967), "The Core of an n-Person Game," *Econometrica* **35**, 50–69.

Schneider, G.and N. Wiesehomeier (2006), "Ethnic Polarization, Potential Conflict, and Civil Wars: Comment," unpublished manuscript, University of Konstanz.

Shapley, Ll. S. (1967), "On Balanced Sets and Cores," *Naval Research Logistics Quarterly* **14**: 453–460.

Skaperdas, S. (1996) "Contest Success Functions," *Economic Theory* **7**: 283–290.

Slantchev, B. (2003), "The Power to Hurt: Costly Conflict with Completely Informed States," *American Political Science Review* **47**, 123–135.

Taylor, M. (1987) *The Possibility of Cooperation.* Cambridge, UK: Cambridge University Press.

Tocqueville, A. de (1835), *Democracy in America*, Chicago, IL: University of Chicaco Press (2000 edition).

Vreeland, J. R. (2008). "The Effect of Political Regime on Civil War," *Journal of Conflict Resolution*, **52**: 401–425.

Weidmann, N. B., Rod, J.K. and Cederman, L. -E. (2010), "Representing Ethnic Groups in Space: A New Dataset," *Journal of Peace Research* **47**: 491-499.

Wooldridge, J.M. (2003), "Cluster-Sample Methods in applied Econometrics," *The American Economic Review* **93**, Papers and Proceedings of the One Hundred Fifteenth Annual Meeting of the American Economic Association, 133–138.

Wucherpfennig, J., Weidmann, N., Girardin, L., Cederman, L. and A. Wimmer (2011), "Politically relevant ethnic groups across space and time: Introducing the GeoEPR dataset," *Conflict Management and Peace Science* **28**: 423–437.

Appendix

This Appendix contains a proof of Proposition 3 (Section A.1), definitions of all the variables in the empirical analysis (Section A.2) as well as a table of summary statistics (Section A.3).

A.1. Proof of Proposition 3. Suppose that the conditions in the proposition are met, but that there is indeed an unblocked allocation **x**. For every group $G \in C$, we have

$$\int_{j \in G} x(j) \geq v[kp(m) + (1-k)p(m)^2] > vm. \tag{a.1}$$

Pick a collection of balancing weights $\{\lambda(G)\}_{G \in C}$. Multiplying each side of (a.1) by $\lambda(G)$, and adding over all groups in C, we see that

$$\sum_{G \in C_j} \lambda(G) \int_{j \in G} x(j) > \sum_{G \in C_j} vm\lambda(G).$$

Because $\{\lambda(G)\}_{G \in C}$ are balanced weights, this implies $\int_j x(j) > v$, a contradiction. ■

A.2. Variable Definitions. We provide below the definitions and sources of the variables employed in the empirical analysis.

Conflict INCIDENCE: group-level dummy variable = 1 for those years where an ethnic group is involved in armed conflict against the state resulting in more than 25 battle-related deaths. Source: Cederman, Buhaug and Rod (2009); CBR henceforth.

Conflict ONSET: group-level dummy variable = 1 in a given year if an armed conflict against the state resulting in more than 25 battle-related deaths involving that ethnic group newly starts. For ongoing wars, ONSET is coded as missing. Source: CBR.

SHARE OF CONFLICT YEARS: group-level variable that captures the share of years a group has been in conflict against the State in the period 1960–2006. Source: CBR.

SHARE OF ONSET YEARS: group-level variable that captures the share of years a group has started in conflict against the State (onset years) in the period 1960–2006. Source: CBR.

SIZE: relative size of the group, source: CBR.

OIL: log of the homeland area covered by oil (in thousands of square kilometres) times the international price of oil. To avoid taking the log of zero, 1 has been added to all observations. Source: Oil fields: Petrodata (Lujala et al. 2007). Oil prices: the World Bank.

OIL (AREA): log of homeland area covered by oil (in '000 km^2). To avoid taking the log of zero, 1 has been added to all observations. Source: Petrodata (Lujala et al. 2007).

OIL (SHARE): ratio of OIL(AREA) to homeland area. Source: Petrodata and GREG.

OIL CONCENTRATION: Herfindahl index of oil reserve distribution across groups. Source: Petrodata and GREG.

OIL (COUNTRY): log of the area of the country covered by oil (in thousands of square kilometres) times the international price of oil. To avoid taking the log of zero, 1 has been added to all observations. Source: information on oil fields comes from Petrodata (Lujala et al. 2007). Data on oil prices comes from the World Bank.

MINES: mineral availability in the ethnic homeland, computed as follows: we consider 13 minerals (bauxite, coal, copper, diamond, gold, iron, lead, nickel, platinum, phosphate, silver, tin and zinc) for which international price data is readily available. For each mineral, year and ethnic group, we create a dummy variable that is one if the group has at least one active mine of that mineral. Then, each dummy is multiplied by its normalized international price, constructed as the log of its international price divided by the log of its price in 1980 (the year when the data starts). MINES is the sum of the resulting values. Data on mineral availability comes from the *Raw Material Data* dataset, (IntierraRMG, 2015) whereas data on mineral prices is provided by the World Bank.

PRIVINDEX: Index of privateness computed via factor analysis on the variables OIL, MINES and HOME.

AUTOC: country average of the Polity IV autocracy index for the years 1960 to 1975.

EXCLUDED: average over the period 1960-1975 of *excluded*, a dummy variable that is 1 if the ethnic group is in power in a given country and year (source: CBR).

CIVIL RIGHTS: Lack of civil liberties from Freedom house. We rescale the original index so it is measured between 0 and 1 (where 0 indicates highest level of civil liberties). For each country, we average its value from 1972 to 1975 and assign the resulting quantity to all post 1975 years.

POLITICAL RIGHTS: Lack of political liberties from Freedom House, rescaled to lie between 0 and 1 (where 0 indicates highest level of civil liberties). For each country, we average its value from 1972 to 1975 and assign the resulting quantity to all post 1975 years.

LACK RIGHTS: average of CIVIL RIGHTS and POLITICAL RIGHTS.

CHILD MORTALITY: Number of deaths of children under 5 per 1000 live births. For each country, we consider the average of this quantity over the period 1960 to 1975 and assign it to all subsequent years. Source: UNICEF Global Databases.

PUBINDEX: Index of publicness computed via factor analysis on the variables POLITICAL RIGHTS, CIVIL RIGHTS, AUTOC , EXCLUDED and CHILD MORTALITY.

GIP: dummy variable = 1 if group is in power in country and year (lagged one year), source: CBR.

GROUPAREA: area of the ethnic homeland (in thousands of square kilometres), source: GREG.

HOME: area of the ethnic homeland relative to total area of the country, source: GREG.

LAND PC: log of the total area divided by group population, source: GREG.

SOILCONST: measures limitations that homeland soil presents to agriculture, constructed using the Harmonized World Soil Database from Fischer et al., (2008). Fisher et al. (2008) construct a global grid of 38 nutrient availabilities ranked from 1 (no or slight constraints) to 4 (very severe constraints), and also including categories 5 (mainly non-soil), 6 (permafrost area) and 7 (water bodies). SOILCONST is the average of the cell values pertaining to the group's homeland.

DISTCAP: group's distance to the country capital, source: CBR.

MOUNT: 0-1 index capturing the group's share of mountainous terrain, source: CBR.

PEACEYEARS: number of years since the last group-level onset, source: CBR.

LAG: lagged conflict incidence, source:CBR.

PARTITIONED: dummy variable = 1 if ethnic homeland covers two or more countries, source: GREG.

GDP: log of (country-level) GDP per capita, lagged one year. Source: Penn World Tables.

POP: log of total country population (POP), lagged one year. Source: Penn World Tables.

POLITY: Polity 2 index, lagged one year. Source: Polity IV.

A.3. Summary Statistics. Table A.1 provides summary statistics for the main variables.

	Summary Statistics				
	Obs	Mean	SD	Min	Max
INCIDENCE	64001	0.04	0.19	0.00	1.00
ONSET	61928	0.00	0.06	0.00	1.00
SHARE CONFLICT	1475	.030	.123	0	.982
SIZE	64001	0.10	0.23	0.00	1.00
OIL	64001	0.00	0.001	0.002	0.01
OIL PC	62103	0.00	0.00	0.00	0.17
MINES	65639	0.57	1.42	0.00	13.00
HOME	61968	0.09	0.20	0.00	1.01
PRIVINDEX	61968	-0.00	0.70	-0.41	5.37
LACK RIGHTS	42950	0.64	0.28	0.00	1.00
CIVIL RIGHTS	42950	0.62	0.27	0.00	1.00
POLITICAL RIGHTS	42950	0.66	0.31	0.00	1.00
CHILD MORTALITY	60669	100.88	50.60	12.54	211.01
AUTOC	45870	0.53	0.29	0.00	1.00
EXCLUDED	63544	0.86	0.34	0.00	1.00
PUBINDEX	38049	0.00	0.96	-2.17	1.32
GIP	64001	0.14	0.35	0.00	1.00
GROUPAREA	64001	84.28	406.74	0.04	7354.72
AREA PC	62103	-9.73	2.00	-15.90	1.55
SOILCONST	64001	1.62	0.78	0.00	6.15
DISTCAP	64001	0.92	1.03	0.00	7.97
MOUNT	64001	0.37	0.36	0.00	1.00
PARTITIONED	64001	0.62	0.48	0.00	1.00
GDP	56945	7.75	1.16	5.08	11.16
POP	61893	17.08	1.81	11.73	20.98
POLITY	58120	-0.09	0.70	-1.00	1.00

Table A.1. Variable Description. Summary statistics for the main empirical variables. See Section A.2 for definitions.

Online Appendix [Not For Publication]

This Appendix presents additional discussion and tables probing the robustness of the results in the main text. Section B.1 provides details about the construction of Figure 1. Following the discussion in Section 4.3.5, Table B.1 provides a simple visualization of our main result. It divides the sample in four groups (countries with/without private/public prizes) and compares for each of the bins the average sizes of groups that have experienced conflict over the sample with those that haven't. Table B.3 allows for group alliances; see Section 5.3. Table B.4 considers additional conflict measures; see Section 5.4. Table B.5 considers alternative ways of measuring oil wealth; see Section 5.5. Table B.6 shows that results are robust to clustering errors at the country level and at the ethnic homeland and country level (two-way clustering); see Section 5.7. Section B.3 and Table B.2 present results estimated by maximum likelihood in a logit specification; this complements Section 5.8 in the main text. Table B.7 considers the robustness of our results to dropping regions of the world for two different specifications of the oil variable, as described in Section 5.9.

B.1. More Details for Figure 1. This section describes the data employed in Figure 1. The unit of analysis in panels a) and b) is the country-year. Both panels show conflict in their Y axis, which is defined as a dummy variable that is one for those years where a party is involved in armed conflict against the state resulting in more than 25 battle-related deaths. (Source: UCDP/PRIO). The x-axis in panel (a) is a measure of country-level oil and gas value, obtained from Ross and Paasha (2015)[1] and in panel (b) is the average of the (lack of) political and civil rights indices provided by Freedom House.

The unit of analysis in panel c) is the group-year. The Y axis also shows conflict (see the definition of Conflict INCIDENCE above) while the X axis is group size (SIZE).

The graphs are binned scatters and several control variables have employed in its computation. In particular, Panels a) and b) control for log GDP per capita (Penn World Tables), log population (Penn World Tables) and year and country fixed effects. Panel b) adds to the former controls the oil variable employed in Panel a).

B.2. Average Percentage Sizes of Groups in Conflict. The informal analysis here helps visualize our data in the simplest possible way. We classify countries in four groups, depending on whether they are rich/poor in oil reserves and on whether they have low/high levels of LACK RIGHTS.[2] For each category, Table B.1 provides the average size of groups that have never experienced conflict and, in parentheses, similar information corresponding to groups that have experienced conflict in the period considered in our sample. Our theory makes predictions for the main diagonal and

[1] Ross, M. and M. Paasha (2015), "Oil and Gas Data, 1932-2014", https://doi.org/10.7910/DVN/ZTPW0Y.

[2] To do that, we have first dropped the time dimension of the data by considering for each group the first non-missing value of OIL and LACK RIGHTS. A country is classified as oil rich (poor) if the the oil reserves corresponding to the group with largest reserves in the country are above the overall median. Analogously, a country is classified as low (high) level of LACK RIGHTS if the first non-missing value in the sample is below (above) the median.

this simple analysis confirms the main implications of the theory: the average size of groups in conflict changes dramatically when conflict is over public prizes (where the average size of groups in conflict is 0.16) or over private prizes (where the same figure falls to 0.04).

	Average %-Size of Groups	
	LACK RIGHTS low	LACK RIGHTS high
Oil high	6.7% (**4.0%**)	8.1% (**7.0%**)
Oil low	18.1% (**6.3%**)	8.0% (**16.0%**)

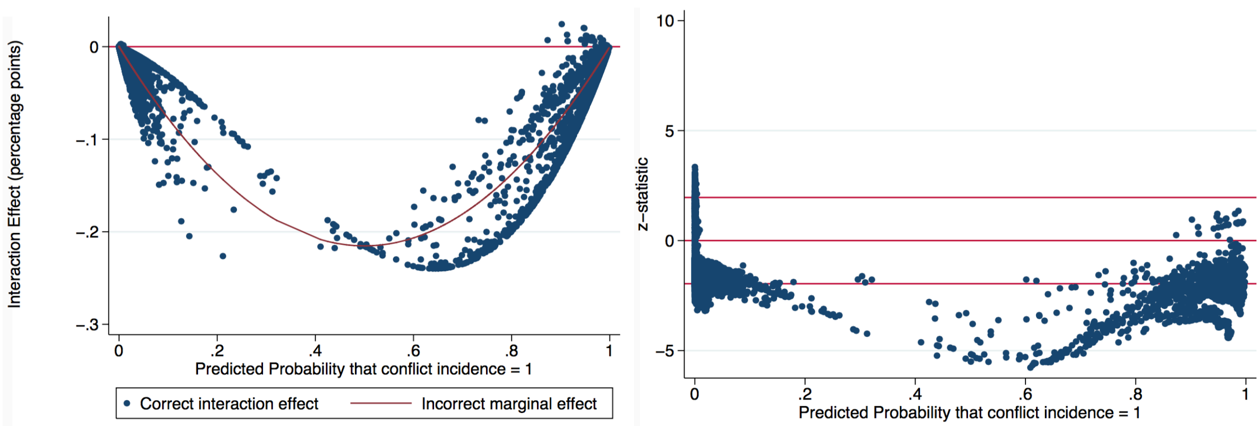
Table B.1. Average Sizes of Groups in Peace and in Conflict. Countries are placed in four categories depending on whether they have high/low values of oil reserves and on whether they have high/low values of LACK RIGHTS. For each bin, it reports the average percentage size of groups that have never experienced conflict in our sample as well as the average percentage size of groups that *have* experienced conflict (in parentheses). See Footnote 2 for details on how countries have been assigned to bins.

	Dependent Variable: **Conflict Incidence**		
	[1]	[2]	[3]
SIZE	23.387***	-15.634	-0.179
	(0.000)	(0.184)	(0.988)
OIL	243.897***	152.334***	213.912***
	(0.000)	(0.001)	(0.000)
SIZE × OIL	-8383.401***		-7403.100***
	(0.000)		(0.000)
SIZE × LACK RIGHTS		40.962**	35.839**
		(0.018)	(0.048)
GDP	0.376**	0.338	0.350*
	(0.024)	(0.101)	(0.093)
POP	1.772**	0.828	0.850
	(0.039)	(0.447)	(0.436)
GROUPAREA	0.000	-0.000	-0.000
	(0.523)	(0.270)	(0.975)
GIP	-0.409	-0.094	-0.065
	(0.106)	(0.720)	(0.799)
SOILCONST	-0.193	-0.118	-0.212*
	(0.167)	(0.311)	(0.081)
DISTCAP	0.565***	0.486***	0.508***
	(0.000)	(0.002)	(0.001)
MOUNT	0.613***	0.587**	0.686***
	(0.009)	(0.019)	(0.007)
PARTITIONED	-0.150	-0.099	
	(0.323)	(0.540)	
LAG	7.270***	7.255***	7.229***
	(0.000)	(0.000)	(0.000)
C	-46.125***	-27.108	-27.751
	(0.007)	(0.211)	(0.201)
Obs	27344	20960	20960

Table B.2. Group Size and Conflict: Non-Linear Models. Conflict incidence is regressed on group size and indices of private and public prizes, along with interactions between subsets of these variables as implied by the theory. Estimation uses maximum likelihood in a Logit model. All regressions contain year dummies and country fixed effects. Robust standard errors clustered at the group level have been computed. p-values are reported in parentheses: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B.3. Interactions in Nonlinear Models. Table B.2 re-estimates our baseline models using a logit specification. The coefficient of the interactions has the expected signs and are highly significant.

Interpreting the coefficients associated with interactions is straightforward in linear models, as they are simply the appropriate cross-partial derivatives of the dependent variable with respect to the relevant variables in the interaction. However, this logic does not extend to nonlinear models, as shown by Ai and Norton (2003). In non-linear models, the cross-partial derivative does not admit a simple interpretation, and important differences arise with respect to the linear case. First, the "true sign" of the interaction does not need to equal the sign of the cross-partial derivative. Second, the significance of that interaction cannot be tested with a simple t-test on the coefficient of the interaction term (in the regression). Third, given the nonlinearity, the value of the interaction term depends on all the independent variables of the model. See Ai and Norton(2003) for a discussion.



(a) Nonlinear Interactions: Cross-Partials

(b) Nonlinear Interactions: z-Statistics

Figure B.1. Interpreting Interactions in Nonlinear Models: Logit. These graphs depict the value of the cross-partial derivative of conflict incidence with respect to OIL and SIZE (Panel a) and the the z-statistics associated with those estimates (Panel b), for each of the points in the sample. Estimates from Table B.2 (Column 1) have been employed to compute the estimates.

To overcome these difficulties and in order to facilitate the interpretation of the interactions reported in Table B.2, we have evaluated the cross-partial derivative at each of the points in our sample. Panel (a) in Figure B.1 plots the derivative of the dependent variable with respect to SIZE and OIL, using the specification in Column 1, Table B.2. This figure shows that the cross-derivative is negative for most observations in our sample, a result that mimics the one obtained for the linear case. Panel (b) in Figure B.1 plots the z-statistics associated with the cross-partial derivative for each of the points in the sample, together with confidence bands (at the 90 per cent level). This figure shows that the effect is significant in most cases. Similar results are found when interpreting the interaction of SIZE and LACK RIGHTS. In this case, the cross-partial derivative is positive and significant for most of the observations.[3]

[3] For the sake of brevity, we don't report the corresponding graphs as they are very similar to those associated with SIZE and OIL, but they are available upon request.

	Dependent Variable: **Conflict Incidence**					
	[1]	[2]	[3]	[4]	[5]	[6]
$\text{SIZE}_{\text{COAL}}$	0.221***	0.405***	0.623***	0.765**	1.100***	1.193***
	(0.000)	(0.000)	(0.000)	(0.015)	(0.002)	(0.002)
OIL_{COAL}	0.002***	0.003***	0.004***	0.003**	0.004***	0.004***
	(0.000)	(0.000)	(0.000)	(0.022)	(0.003)	(0.004)
$\text{SIZE}_{\text{COAL}} \times \text{OIL}_{\text{COAL}}$		-0.054***	-0.047***		-0.069***	-0.048*
		(0.000)	(0.000)		(0.005)	(0.079)
$\text{SIZE}_{\text{COAL}} \times$ LACK RIGHTS				-1.147***	-1.285***	-1.227***
				(0.006)	(0.003)	(0.003)
POLITY			-0.002**			-0.000
			(0.022)			(0.702)
GIP			-0.021***			-0.016***
			(0.000)			(0.002)
GDP			0.001			0.002
			(0.274)			(0.342)
POP			0.002			0.000
			(0.360)			(0.947)
GROUPAREA			-0.000***			-0.000**
			(0.002)			(0.039)
SOILCONST			-0.001			-0.001
			(0.202)			(0.605)
DISTCAP			0.004***			0.005**
			(0.000)			(0.027)
MOUNT			0.003			0.002
			(0.125)			(0.387)
PARTITIONED			-0.001			-0.000
			(0.570)			(0.942)
LAG	0.886***	0.885***	0.876***	0.877***	0.873***	0.870***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
C	-0.027***	-0.039***	-0.095***	-0.024***	-0.034***	-0.067
	(0.000)	(0.000)	(0.005)	(0.007)	(0.001)	(0.608)
R^2	0.845	0.845	0.850	0.853	0.853	0.859
Obs	64839	64839	55289	41314	41314	38341

Table B.3. Groups and Conflict: Alliances. This table regresses conflict incidence on group size (allowing for coalitions) and indices of private and public prizes, along with interactions between subsets of these variables as implied by the theory. All regressions contain year dummies and country fixed effects. *p*-values in parentheses: $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Dependent Variable: **Conflict Onset** [1–3]; **% of Years in Conflict** [4-6]; **% of Onset Years** [7]						
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
SIZE	0.018	0.051***	0.019	0.226	0.613**	0.091	0.080
	(0.191)	(0.001)	(0.382)	(0.318)	(0.030)	(0.766)	(0.199)
OIL	0.698***	0.867***	0.685*	7.549**	9.340**	10.320**	3.169*
	(0.004)	(0.002)	(0.072)	(0.038)	(0.024)	(0.028)	(0.073)
SIZE × OIL		-11.006***	-5.179		-126.967**	-130.765**	-32.260*
		(0.000)	(0.294)		(0.028)	(0.040)	(0.051)
SIZE × LACK RIGHTS			0.039			1.003**	0.170*
			(0.244)			(0.014)	(0.054)
POLITY	-0.001	-0.001	-0.000				
	(0.172)	(0.172)	(0.915)				
GDP	0.001	0.001	0.002	-0.039	-0.027	-0.070	-0.027
	(0.288)	(0.307)	(0.391)	(0.358)	(0.517)	(0.336)	(0.250)
POP	0.002	0.002	0.007	-0.237	-0.242	-0.494	-0.185
	(0.174)	(0.203)	(0.165)	(0.401)	(0.393)	(0.294)	(0.234)
GROUPAREA	-0.000*	-0.000	-0.000	-0.000	0.000	0.000	0.000
	(0.072)	(0.682)	(0.283)	(0.558)	(0.599)	(0.270)	(0.621)
GIP	-0.002	-0.002	-0.001	-0.045**	-0.046**	-0.049**	-0.012*
	(0.125)	(0.112)	(0.460)	(0.027)	(0.023)	(0.020)	(0.074)
SOILCONST	-0.000	-0.000	-0.001	-0.006	-0.006	-0.004	-0.001
	(0.633)	(0.486)	(0.300)	(0.287)	(0.201)	(0.619)	(0.486)
DISTCAP	0.001***	0.001***	0.001	0.013	0.013	0.024	0.008
	(0.000)	(0.000)	(0.179)	(0.193)	(0.186)	(0.143)	(0.141)
MOUNT	0.002	0.002*	0.002	0.024**	0.025**	0.026**	0.007*
	(0.100)	(0.085)	(0.242)	(0.022)	(0.016)	(0.033)	(0.070)
PARTITIONED	-0.001	-0.001	-0.000	-0.009	-0.009	-0.012	-0.004*
	(0.134)	(0.130)	(0.973)	(0.293)	(0.282)	(0.268)	(0.098)
PEACEYRS	-0.001***	-0.001***	-0.002***				
	(0.000)	(0.000)	(0.000)				
C	0.020	0.022	-0.061	3.209	3.164	6.600	2.484
	(0.999)	(.)	(0.426)	(0.401)	(0.410)	(0.302)	(0.238)
R^2	0.033	0.034	0.048	0.373	0.375	0.377	0.240
Obs	53357	53357	36761	1427	1427	1297	1296

Table B.4. Group Size and Conflict: Alternative Dependent Variables. This table regresses conflict onset (Columns 1–3), the share of conflict years (Columns 4–6) and the share of onset years (Column 7) on group size and indices of private and public prizes, along with interactions between subsets of these variables as implied by the theory. All regressions contain country fixed effects and Columns 1–3 include year dummies as well. *p*-values in parentheses: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Dependent Variable: **Conflict Incidence**				
	[1]	[2]	[3]	[4]	[5]
SIZE	0.053***	0.054***	-0.005	0.056***	-0.000
	(0.006)	(0.005)	(0.834)	(0.006)	(0.986)
OIL	0.741***	0.844***	0.856*		
	(0.008)	(0.003)	(0.057)		
OIL COUNTRY	-1.143	-1.238*	0.615		
	(0.105)	(0.077)	(0.823)		
OIL (SURFACE)				0.002***	0.003**
				(0.008)	(0.036)
SIZE × OIL COUNTRY	-10.171***	21.960	31.127		
	(0.006)	(0.182)	(0.177)		
SIZE × OIL		-34.888**	-41.824*		
		(0.038)	(0.086)		
SIZE × LACK RIGHTS			0.083**		0.084**
			(0.034)		(0.032)
SIZE × OIL (SURFACE)				-0.029***	-0.031**
				(0.001)	(0.030)
GDP	0.001	0.001	0.003	0.001	0.003
	(0.192)	(0.159)	(0.312)	(0.134)	(0.326)
POLITY	-0.002**	-0.001**	-0.001	-0.002**	-0.001
	(0.021)	(0.022)	(0.548)	(0.017)	(0.515)
POP	0.001	0.001	-0.001	0.001	-0.001
	(0.773)	(0.746)	(0.892)	(0.614)	(0.870)
GROUPAREA	0.000	0.000	0.000	0.000	0.000
	(0.570)	(0.470)	(0.363)	(0.572)	(0.634)
GIP	-0.003*	-0.004**	-0.003	-0.004**	-0.003
	(0.060)	(0.034)	(0.179)	(0.041)	(0.236)
SOILCONST	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.320)	(0.284)	(0.407)	(0.301)	(0.429)
DISTCAP	0.002***	0.002***	0.003*	0.002***	0.002*
	(0.000)	(0.000)	(0.072)	(0.000)	(0.098)
MOUNT	0.002*	0.002*	0.002	0.002*	0.002
	(0.067)	(0.068)	(0.122)	(0.076)	(0.104)
PARTITIONED	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.305)	(0.270)	(0.414)	(0.316)	(0.475)
LAG	0.895***	0.894***	0.900***	0.894***	0.900***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
C	-0.026	-0.026	-0.016	-0.036	-0.014
	(0.475)	(0.460)	(0.916)	(0.269)	(0.930)
R^2	0.847	0.847	0.855	0.847	0.855
Obs	55289	55289	38341	55289	38341

Table B.5. Group Size and Conflict: Alternative Ways of Measuring Oil Wealth. This table regresses conflict incidence on group size and indices of private and public prizes, along with interactions between subsets of these variables as suggested by the theory. All regressions contain year dummies and country fixed effects. p-values are reported in parentheses. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Dependent Variable: **Conflict Incidence**									
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
SIZE	-0.015	0.032	0.066***	0.080***	0.066**	-0.073***	-0.002	0.011	0.022	0.084**
	(0.372)	(0.192)	(0.006)	(0.002)	(0.019)	(0.007)	(0.916)	(0.658)	(0.413)	(0.019)
OIL	0.448	0.684*	0.771*			0.631*	0.887*	0.828*		
	(0.143)	(0.082)	(0.072)			(0.099)	(0.065)	(0.075)		
SIZE × OIL		-13.628**	-14.433**				-14.455**	-12.836**		
		(0.023)	(0.014)				(0.038)	(0.028)		
SIZE × OIL_{0-25}				0.047					0.033	
				(0.582)					(0.763)	
SIZE × OIL_{25-50}				-0.058					0.182	
				(0.392)					(0.549)	
SIZE × OIL_{50-75}				-0.154***					-0.166**	
				(0.004)					(0.026)	
SIZE × $\text{OIL}_{>75}$				-0.135***					-0.118***	
				(0.001)					(0.006)	
SIZE × PRIV. IND.					-0.049**					-0.046**
					(0.017)					(0.018)
SIZE × LACK RIGHTS						0.097**	0.068*	0.083**	0.083**	
						(0.029)	(0.065)	(0.035)	(0.050)	
SIZE × PUBLIC INDEX										0.023*
										(0.080)
$\text{OIL}_{>75}$				0.007**					0.007**	
				(0.014)					(0.024)	
OIL_{50-75}				0.004*					0.005*	
				(0.095)					(0.067)	
OIL_{25-50}				-0.001					-0.002	
				(0.834)					(0.600)	
OIL_{0-25}				-0.004					-0.003	
				(0.202)					(0.399)	
PRIVATE INDEX					0.002*					0.002
					(0.067)					(0.134)
PUBLIC INDEX										0.110
										(0.356)
POLITY			-0.002	-0.002	-0.002			-0.001	-0.001	-0.001
			(0.213)	(0.215)	(0.233)			(0.512)	(0.513)	(0.437)
GIP			-0.004*	-0.003	-0.004*			-0.003	-0.003	0.001
			(0.098)	(0.105)	(0.084)			(0.224)	(0.231)	(0.856)
GDP			0.001	0.001	0.001			0.003	0.003	0.003
			(0.433)	(0.476)	(0.469)			(0.321)	(0.317)	(0.354)
POP			0.001	0.002	0.001			-0.001	-0.001	-0.001
			(0.810)	(0.739)	(0.883)			(0.878)	(0.883)	(0.919)
GROUPAREA			0.000	0.000	0.000			0.000	0.000	0.000
			(0.331)	(0.529)	(0.354)			(0.290)	(0.324)	(0.321)
SOILCONST			-0.001	-0.001	-0.001			-0.001	-0.001	-0.001
			(0.455)	(0.407)	(0.396)			(0.409)	(0.368)	(0.399)
DISTCAP			0.002	0.002	0.002			0.003*	0.002*	0.003*
			(0.117)	(0.125)	(0.129)			(0.084)	(0.084)	(0.087)
MOUNT			0.002**	0.003**	0.002**			0.002	0.002	0.002
			(0.022)	(0.016)	(0.030)			(0.109)	(0.102)	(0.169)
PARTITIONED			-0.001	-0.001	-0.001			-0.001	-0.001	-0.001
			(0.306)	(0.348)	(0.296)			(0.445)	(0.431)	(0.478)
LAG	0.895***	0.895***	0.894***	0.894***	0.894***	0.898***	0.898***	0.900***	0.900***	0.900***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
C	-0.002	-0.005**	-0.035	-0.033	-0.028	-0.002	-0.010	-0.015	-0.020	-0.142
	(0.355)	(0.021)	(0.601)	(0.626)	(0.777)	(0.800)	(0.166)	(0.924)	(0.902)	(0.473)
R^2	0.844	0.844	0.847	0.848	0.847	0.850	0.850	0.855	0.855	0.853
Obs	64839	64839	55289	55289	54486	41314	41314	38341	38341	35755

Table B.6. Robustness to Clustering Errors at the Country Level and at the Country and Group Level (Two-Way Clustering). This table regresses conflict incidence on group size and indices of private and public prizes, along with interactions between subsets of these variables as suggested by the theory. All regressions contain year dummies and country fixed effects. Robust standard errors clustered at the country level in Columns 1–5 and at the country and group level (two way clustering) have been computed. In the latter case, a group living in two different countries is considered as one group, as opposed to the rest of the exercise where ethnic groups are country specific. **p-values**, based on robust standard errors clustering at the group and country level, are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Dependent Variable: **Conflict Incidence**											
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
SIZE	0.067***	0.007	0.067***	0.006	0.060***	0.004	0.048**	0.018	0.054**	0.016	0.094***	0.033
	(0.001)	(0.795)	(0.002)	(0.801)	(0.003)	(0.855)	(0.024)	(0.486)	(0.014)	(0.483)	(0.001)	(0.329)
OIL	0.775**	0.838*	0.944***	1.102***	0.671**	0.707	0.592**	0.649	0.627**	0.633	1.054***	1.132*
	(0.011)	(0.070)	(0.006)	(0.010)	(0.022)	(0.143)	(0.032)	(0.136)	(0.023)	(0.209)	(0.004)	(0.058)
SIZE × OIL	-15.148***	-12.398**	-16.180***	-16.704***	-13.430***	-10.574*	-11.333***	-11.640**	-8.499**	-6.772	-21.314***	-18.699**
	(0.000)	(0.028)	(0.000)	(0.004)	(0.001)	(0.071)	(0.004)	(0.043)	(0.019)	(0.211)	(0.000)	(0.021)
SIZE × LACK RIGHTS		0.092**		0.096**		0.082*		0.074*		0.041		0.079*
		(0.049)		(0.045)		(0.050)		(0.059)		(0.140)		(0.074)
POLITY	-0.002**	-0.001	-0.001	-0.000	-0.002**	-0.001	-0.003***	-0.001	-0.001*	-0.001	-0.001	-0.000
	(0.010)	(0.435)	(0.133)	(0.845)	(0.025)	(0.557)	(0.000)	(0.261)	(0.079)	(0.240)	(0.364)	(0.974)
GIP	-0.004**	-0.003	-0.002	-0.001	-0.003	-0.002	-0.003	-0.002	-0.007***	-0.005**	-0.003*	-0.002
	(0.047)	(0.250)	(0.241)	(0.639)	(0.116)	(0.426)	(0.210)	(0.361)	(0.001)	(0.027)	(0.090)	(0.346)
GDP	0.002	0.003	0.000	0.004	0.001	0.002	0.000	0.001	0.002**	0.004	0.002	0.002
	(0.109)	(0.302)	(0.886)	(0.377)	(0.179)	(0.374)	(0.650)	(0.805)	(0.023)	(0.191)	(0.135)	(0.414)
POP	0.003	0.001	0.001	-0.001	0.001	-0.001	-0.003	-0.010	-0.003	-0.005	0.005**	0.004
	(0.275)	(0.935)	(0.669)	(0.883)	(0.577)	(0.906)	(0.225)	(0.297)	(0.183)	(0.599)	(0.023)	(0.541)
GROUPAREA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.156)	(0.353)	(0.517)	(0.301)	(0.328)	(0.395)	(0.341)	(0.372)	(0.181)	(0.114)	(0.441)	(0.374)
SOILCONST	-0.000	-0.001	0.001	0.000	-0.000	-0.000	-0.000	-0.000	-0.002***	-0.002***	-0.001	-0.001
	(0.697)	(0.396)	(0.416)	(0.774)	(0.459)	(0.675)	(0.370)	(0.550)	(0.000)	(0.004)	(0.243)	(0.301)
DISTCAP	0.002***	0.003*	0.000	0.001	0.002***	0.003*	0.002***	0.002*	0.002***	0.002	0.002***	0.003*
	(0.000)	(0.085)	(0.435)	(0.291)	(0.000)	(0.077)	(0.000)	(0.086)	(0.002)	(0.173)	(0.000)	(0.094)
MOUNT	0.002	0.003	0.002	0.001	0.002	0.002	0.002	0.002	0.003**	0.004**	0.003*	0.003*
	(0.108)	(0.102)	(0.167)	(0.664)	(0.118)	(0.198)	(0.102)	(0.228)	(0.042)	(0.021)	(0.077)	(0.089)
PARTITIONED	-0.001	-0.001	-0.002	-0.001	-0.002	-0.002	-0.001	-0.001	0.001	0.001	-0.002	-0.001
	(0.291)	(0.446)	(0.186)	(0.380)	(0.105)	(0.194)	(0.429)	(0.475)	(0.371)	(0.176)	(0.196)	(0.378)
LAG	0.895***	0.900***	0.885***	0.893***	0.893***	0.899***	0.903***	0.911***	0.886***	0.891***	0.895***	0.899***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
C	-0.055	-0.043	-0.022	-0.013	-0.036	-0.010	0.038	0.164	0.028	0.043	-0.105**	-0.094
	(0.118)	(0.724)	(0.621)	(0.925)	(0.409)	(0.929)	(0.303)	(0.333)	(0.484)	(0.786)	(0.017)	(0.384)
Dropped Reg.	EX-USSR	EX-USSR	ASIA	ASIA	MID. EAST	MID. EAST	W-SAFR.	W-S AFR.	E-C AFR.	E-C AFR.	LATIN AM.	LATIN AM.
R^2	0.850	0.854	0.851	0.856	0.846	0.854	0.861	0.869	0.844	0.849	0.829	0.841
Obs	53181	37504	39004	26829	52637	36473	45984	31513	45840	31580	44673	31092

Table B.7. Dropping Regions of the World. This table reproduces Columns 3 and 8 from Table 1 dropping regions of the world. Regions dropped are: former USSR countries (Columns 1 and 2), Asia (Columns 3 and 4), the Middle East (Columns 5 and 6), West-South Africa (Columns 7 and 8), East and Central Africa (Columns 9 and 10), and Latin America (Columns 11 and 12). All regressions contain year dummies and country fixed effects. *p*-values are reported in parentheses. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.