Online appendices

for

Andreas Wimmer and Chris Miner, "The strategic logic of ethno-territorial competition. Violence against civilians in Africa's civil wars", forthcoming in *Journal of Global Security Studies* 4, 2019

Appendix 1: Details on independent variables

Oil, diamonds, and economically significant mining

We use the geo-coded data on active, on-shore oil wells provided by the Centre for the Study of Civil War (CSCW) at the Peace Research Institute Oslo (Lujala *et al.* 2007). We create a count variable that enumerates the number of producing oil wells in a hexagon (and use other codings for robustness purposes). We also code whether there are diamond deposits in a hexagon (also CSCW data, provided by Gilmore *et al.* 2005) and again use other codings (such as a kilometer distance variable) for robustness purposes. Recently, other kinds of resources such as Tungsten, Coltan, copper, and so forth have come to the fore of the international discussion on the civilian victims of violence. The National Minerals Information Center of the United States Geological Survey (USGS) provides geo-coded data on mining production or deposits around the world and we code a variable that counts the number of economically significant mining sites per hexagon.

Organizational control, external support of rebels, and reliance on natural resource financing

The first two variables are derived from the Non-State-Actor Dataset (NSAD) introduced by Cunningham, Gleditsch, and Salehyan (2009). We use a dichotomous variable that codes whether or not a rebel organization received any kind of financial support from actors situated outside of the country in question. The second variable from the NSAD is the strength of the control that

rebel leaders exercise over the rank-and-file fighters. It is re-coded as 0 if no such control exists, 1 for low, 2 for moderate, and 3 for high control.

To evaluate the natural resource part of Weinstein's argument, we use a dataset assembled by Rustad and Binningsbo (2012). The variable of interest here is whether or not a rebel organization has been funded by natural resources (including drugs). The data ranges up to 2006 only, resulting in missing values compared to other models without this variable. As with the external financing variable, there is no information on whether rebels relied on such resources when they first organized or later on or both—we thus cannot test Weinstein's path dependency argument in fully adequate way.

Territorial gain by rebels or government troops

ACLED lists the location of battles between governments and rebels in which one or the other side gained territory. It also enumerates locations that changed hands without any fighting or where rebel groups established a new base. We combined these variables into two new variables indicating whether government or rebels have gained territory, whether through battle or peacefully. Since the ACLED data refer to a specific spatial point, we don't know the extension of the territory that governments or rebels gained control over. We thus use a count variable that notes the number of territory changes in favor of rebels or in favor of government within a hexagon. To check for robustness, we also coded a distance variable—assuming that the degree of control over territory monotonically decreases with distance to the location of a territory change. The results are substantially very similar.

How well do these two variables capture the logic of Kalyvas' argument? It would obviously be better to have data on actual levels of control over territory by specific actors. It is, however, not that far-fetched to assume that in recently acquired territories, an armed group has neither

established full control yet nor continues to share control with its adversary. To exclude the possibility that we are looking at territories that have recently switched back and forth between rebels and government troops and thus represent a contested territory (where civilians should be safer), we exclude hexagons where territories have changed hands both in favor of rebels and government in the previous 12 months.

Finally, we restricted the universe to guerrilla wars (excluding conventional wars, using the list provided in Kalyvas and Balcells 2010) and to events with a maximum of 5 civilians killed, since a higher death toll could indicate that rebels or government troops committed a massacre, rather than acting upon locally provided information to selectively target potential collaborators, as in Kalyvas' theory. We also coded less restrictive coding of independent and dependent variables, which include violence in all types of civil war, without any restriction on the number of victims, and without disregarding hexagons where territories have changed hands recently. The results are reported in Table 3.

Distance to border between settlement areas of co-ethnics of rebels and government elites

This is one of the two variables associated with the ethno-territorial competition model. In order to make sure we did not code this distance across unrelated conflicts, we limited the calculations to distances from borders of groups in conflict within the corresponding conflict areas. For hexagons outside these areas, we coded the maximum distance we found within conflict areas (2500 km). Results are substantially identical if these hexagons outside conflict polygons are given missing values on the distance variables, thus cutting the number of observations dramatically.

Control variables

We experimented with a very large number of geo-coded data and retained those significantly associated with one-sided violence against civilians, making sure to avoid collinearity problems. The various models include a combination of the following variables: a dummy variable indicating whether hexagon was situated 100 km or closer to the border (based on Robert Hijmans' GADM database of Global Administrative Areas),[1] the km distance (or in some models the travel time) to a city of at least 50 thousand inhabitants (calculated on the basis of the list of cities provided by Nelson 2008), the distance to capital or in some models to the coast, the number of people living in an urban location (data from Gridded Population of the World Data Version 3, produced by the Socioeconomic Data and Applications Center of NASA), the roughness of terrain (from the Shuttle Radar Topography Mission of NASA), the length of roads (from Vector Map Level 0 provided by the National Imagery and Mapping Agency NIMAA), the degree to which the surface consists of sparse vegetation (data from the RCP database produced by the International Institute for Applied Systems Analysis IIASA), the distance to a coast-line, and malaria risk (from the Malaria Atlas Project). We also include a distance to a battlefield without territory changes as an additional variable in order to control for proximity to fighting. The baseline models using various combinations of these variables all produce substantially similar and consistent results: Civilians are more likely to be killed where there are many of them (closer to cities; where the surface is urban; or where the urban population is large; see also (Höglund *et al.* 2016); where access to them is good (road length); where rebels can hide or escape from government forces (close to the border; far away from the capital; where the terrain

[1] We also experimented with a continuous distance to border variable but found that the effect ceases to be consistent after 100 km.

is rough; where the surface is richly vegetated).[2] These various results, while interesting, do not speak directly to any of the major theories of violence that we seek to empirically evaluate. Therefore, they will remain in the background of the analysis and the corresponding terms are not shown in the main tables.

Mean centered and logged variables

To increase model stability, we mean-centered the following variables: distance to oil and diamond production sites, the strength of leadership control of rebel organizations, and travel distance to a city of 50'000 or above. The following variables were logged in order to avoid extremely skewed distributions: violence against civilians in neighboring hexagons (ACLED dataset), distance to battlefields (ACLED dataset; the variable was subsequently mean centered as well), and rough terrain (the variable was subsequently mean centered as well).

[2] Our findings regarding urban population size, distance to borders, and road length are in line with those of Raleigh and Hegre (2009), who offer a more detailed interpretation of the mechanisms that underlie these associations. Only in one of our four baseline models was distance to the capital associated with more violence, as was the case in their models. We also don't find that higher incomes (proxied with nutrition quality of soils, electric light intensity at night, or infant mortality) are associated with an increased likelihood of violence.

Appendix 2: Summary statistics

Model 1	No of observations	Mean	Standard Dev.	Minimum	Maximum
Onesided violence against civilians committed by government or rebel forces	43,133	0.031484	0.1746238	0	1
Presence of oil producing wells	43,133	0.0190805	0.1368097	0	1
Producing lootable diamond deposit	43,133	0.0059119	0.0766625	0	1
Number of economically significant mineral production sites or deposits	43,133	0.0130063	0.1954518	0	13
Model 2					
Onesided violence against civilians by rebel forces	24,359	0.0435157	0.2040192	0	1
Strength of central control over rank-and-file rebel fighters	17,157	2.040333	0.7720524	0	3
Outside financial support for rebels	17,105	0.2572932	0.4371551	0	1
Rebels funded by or fought over natural resources	18,758	0.2150549	0.4108714	0	1
Models 3 and 4 (ACLED data)					
Onesided violence against civilians by government forces with less than 5 deaths	270,351	0.0008729	0.0295327	0	1
Onesided violence against civilians by rebel forces with less than 5 deaths	270,351	0.0028408	0.053223	0	1
No of times government gained terrain (and rebels did not during past 12 months)	270,351	0.0011467	0.0530081	0	12
No of times rebels gained terrain (and government did not during past 12 months)	270,351	0.00199	0.068672	0	10
Modesl 5-7					
Onesided violence against civilians by government forces	36,076	0.0168533	0.1287235	0	1
Onesided violence against civilians by excluded rebel forces	36,076	0.0088702	0.0937642	0	1
Onesided violence against civilians by infighter forces	36,076	0.0044351	0.0664495	0	1
Entire hex population in rebellion	36,076	0.0035758	0.0596917	0	1
Entire hex population included in government	36,076	0.3011143	0.4587486	0	1
Entire hex population included and not in rebellion	36,076	0.2580663	0.4375767	0	1
Polarization between included and rebelious population	36,076	0.0312532	0.1681972	0	1
Polarization between included and excluded population in rebellion	36,076	0.0101394	0.0912043	0	1
Polarization between included (not in rebellion) and infighting population	36,076	0.013127	0.099643	0	1
Km distance to nearest border between included (not in rebellion) and population in rebellion (logged)	35,968	4.44306	1.654992	-6.907755	7.308537
Km distance to nearest border between included and population in rebellion (logged)	35,614	7.405902	1.035492	3.347797	7.824046
Km distance to nearest border between included (not in rebellion) and infighters (logged)	35,486	7.580912	0.7601877	3.358638	7.834151

Appendix 3: Re-analyzing Fjelde and Hultman's results

As reported in the main text, Fjelde and Hultman (2014) find support for their idea that rebels target co-ethnics of government elites and government troops target co-ethnics of rebels. Our results, however, indicate that 1) government troops do not kill more in areas populated by ethnic groups in rebellion; 2) rebels kill *less* in areas populated by included groups represented in government. Fjelde and Hultman's (FH) analysis unfolds in a different universe of cases, uses other control variables, specifies the main independent and dependent variables differently, and is based a different model specification. Which of these differences explain the divergent findings? In what follows, we first show that the conflicting finding regarding government violence is due to the fact that FH do not control for polarization, one of the two key variables in our ethno-territorial competition model. Once one controls for polarization, their results disappear. We then report that their second findings regarding rebel violence are not robust to a different model specification or a different definition of the independent variable. The main problems are multi-collinearity as well as a definition of the independent variable (the presence of included groups) that suffers from measurement problems. When these problems are addressed adequately, their findings regarding rebel violence are reversed … and thus become consistent with those that we report in the paper.

In a third section we outline the differences in research design and definition of the empirical universe and indicate the reasons for our choices—without assuming that FH's choices are unreasonable or inferior. As the first two sections show, the differences in the empirical universes do not generate the differences in findings. Finally, differences regarding variable definitions, control variables, and model specification are discussed.

Differences in the findings regarding government violence

Perhaps we do not find, in contrast to FH, that government are more likely to kill in areas of rebel ethnicities because we also include polarization and distance to border in our models? To assess this possibility, we first rely on our universe of cases since polarization and distance to border are not available in FH's dataset. But we used FH's dependent variable (a count variable of incidences of violence), model specification (negative binomial or zero-inflated negative binomial), definition of independent variable (the presence of at least one excluded group in rebellion), and control variables. We find that 1) without polarization and distance to border, the presence of rebellious ethnic groups increases the number of violent events thus reproducing FH's findings. With polarization and distance to border included in the model, however, the variable is no longer significant at standard levels—in line with the findings that we report in this article.

In a second step, we relied on FH's universe of cases and their exact model specification. Adding polarization to the equation, which is highly significant and with a large coefficient, reduces the significance level of the rebel constituency variable dramatically, but it remains borderline significant at standard levels. Once we introduce country fixed effects, however, it is no longer associated with government violence, while polarization remains highly significant and with a large coefficient (actually the only co-variate in the model that remains significant).

Differences in findings regarding rebel violence

FH report that rebels kill civilians more often in areas where there is at least one ethnic group is allied with government—while our paper finds a significant association in the opposite direction. Additional analysis shows that FH's findings either loose statistical significance or the sign of the coefficient turns from positive to negative—in line with our own results—when slightly

modifying FH's model specification. For this test, we remain in the empirical universe of FH's dataset—the results are therefore not due to a change in the boundaries of the universe.

1) The model with rebel violence as the outcome, as specified by FH, suffers from collinearity problems: many control variables are correlated above .9; the average VIF value is 78 using an OLS regression, while anything above 10 is considered highly problematic. It is therefore not surprising that without GDP as a control variable, the association between the killing of civilians by rebel organizations and the presence of included groups is no longer significant. Without GDP, distance to capital, and rebel strength, rebel deaths are significantly *less* likely where there are included groups. This is consistent with what we find.
2) The substantial findings are also reversed if we use a more precise definition of the independent variable. Using a dummy for the presence of an included group, as FH do, does not tell us whether rebels indeed kill members of these groups (rather than anybody else). A less problematic specification of the IV is to calculate the percentage share of included populations. If coded in this manner, there is again a negatively significant association: Rebels are *less* frequently killing civilians the *larger* the share of the included population, i.e. the government constituency. This is again in line with our findings.[3]

[3] When operating within the empirical universe of this article (and the DV as defined in this article), a continuously defined independent variable produces the same results as under 2 and 3 above: Rebels tend to avoid areas populated by included ethnic groups. This is consistent with the finding that we report in Tables 1 and 2 of the main text (with a differently defined IV: a dummy indicating whether the entire population is included or excluded).

Universe

In the following we outline the main differences between the empirical universes of the two studies and give the reasons for our choices. As the discussion above showed, however, the divergent findings are *not* due to these different model universes.

1) FH use as units of observations squares of 55x55 km size, while our hexagons measure 27.5 x 27.5. We believe that the smaller units are adequate given that the ethnic settlement maps are rather fine-grained such that precision in the measurements matters, especially for evaluating a polarization argument that is sensitive to the size of units of observation. Nor surprisingly and as discussed in the paper, the polarization variable looses significance in the one model on rebel violence when we re-create the universe with a larger, 50-50km square grid, while the rest of the models remain substantially very similar.
2) FH use conflict polygons plus all squares that are within a 100 km radius of these polygons. We do not enlarge the area of observation in this way, thus operating with fewer observations than FH do (reducing their observations to polygons does not substantially alter their results).
3) FH include up to 2 years during which there was no active conflict (according to the 25 year battle death threshold of the UCDP datasets) as long as these years fall in between active conflict periods. These years are not included in our models since we do not know if a conflict was active or not, i.e. if there was any rebel or government activity at all and if the killing of civilians is related to the conflict dynamics. We are therefore using many fewer observations (about half; we note here that FH's finding regarding rebel violence turns insignificant without these intervening years).

4) FH do not reduce to their universe to the squares in which there was at least one politically relevant ethnic group. FH's universe therefore includes units on which, per definition, a logic of "killing the adversary" cannot operate because there are no targets. We therefore operate within a smaller universe (reducing the universe would not affect their results substantially, however).
5) FH do not distinguish between infighting conflicts and rebellions, which is potentially confounding the estimates since infighters target other groups in the governing coalition (e.g. senior partners or junior partners), not only those that hold dominant or monopoly positions, as specified in FH's models.

Variable definition and model specification

1) FH use a dummy variable to identify the presence of at least one included or excluded group on a unit of observation. As discussed above, this then opens the possibility that government troops kill their own civilians in a grid cell where there are also a few members of excluded groups (and the other way around for rebel organizations). A percentage of included/excluded measurement, while not avoiding these measurement problems completely, at least reduces them. Using a dummy for 100% of the corresponding population in a unit of observation, as we do in the article, represents perhaps the best solution. As shown above, using a % measurement of the adversary's co-ethnic population reverses FH's findings regarding rebel violence.
2) The model specification of FH does not include fixed effects for countries. We believe that this is crucial in order to control for time-invariant country-level differences (such as the legal system affecting the punishment for killing civilians, the nature of the conflicts,

etc.). Including country fixed effects is changing FH's results with regard to government violence, as discussed above

3) As control variables, FH include GDP (based on lights data), population density, previous year battle event, whether rebels are secessionist, rebel strength, distance to capital, last year's events in neighboring cells, time since last event (plus time squared and cubed), and events committed by other side last year. We believe that it is meaningful to include additional controls that concern the accessibility and availability of civilians for their potential killers. We doubt, however, that this would substantially affect their results.
4) FH use a count variable as their dependent variable. As discussed in the paper, we prefer a dichotomous dependent variable given the uncertainties in disentangling events from each other and thus arriving at a clear count. Our results hold up as well, however, with a count variable as dependent variable, as shown in Model 4 in Table 3.

Appendix Bibliography

Cunningham, David E., Kristian S. Gleditsch and Idean Salehyan. (2009) It Takes Two: A Dyadic Analysis of Civil War Duration and Outcome. *Journal of Conflict Resolution* 54 (4): 570-597.

Gilmore, Elisabeth, Nils Petter Gleditsch, Päivi Lujala and Jan Ketil Rød. (2005) Conflict Diamonds: A New Dataset. *Conflict Management and Peace Science* 22 (3): 257-292.

Höglund, Kristine, Erik Melander, Margareta Sollenberg and Ralph Sundberg. (2016) Armed Conflict and Space: Exploring Urban-Rural Patterns of Violence. In *Spatializing Peace and Conflict*, ed. A Björkdahl and et al., 60-76. New York: Springer.

Kalyvas, Stathis N. and Laia Balcells. (2010) International System and Technologies of Rebellion: How the End of the Cold War Shaped Internal Conflict. *American Political Science Review* 104 (3): 415-429.

Lujala, Päivi, Jan Ketil Rød and Nadia Thieme. (2007) Fighting over Oil: Introducing a New Dataset. *Conflict Management and Peace Science* 24 (3): 239-256.

Nelson, Andrew. 2008. *Travel Time to Major Cities: A Global Map of Accessibility*. Luxembourg: Office for Official Publications of the European Communities.

Raleigh, Clionadh and Håvard Hegre. (2009) Population Size, Concentration, and Civil War. A Geographically Disaggregated Analysis. *Political Geography* 28 224-238.

Rustad, Siri Aas and Helga Malmin Binningsbo. (2012) A Price Worth Fighting For? Natural Resources and Conflict Recurrence. *Journal of Peace Research* 49 (4): 531-546.