

Location, Location, Location: The Ethno-Geography of Oil and the Onset of Ethnic War

Hui Li¹ · Shiping Tang¹

Received: 28 February 2017 / Accepted: 7 March 2017 / Published online: 14 March 2017
© Fudan University and Springer Science+Business Media Singapore 2017

Abstract Challenging and improving upon existing studies, we develop a more integrated and fine-grained theory regarding oil and the onset of ethnic war and present systematic quantitative evidences for our theory. We contend that it is the ethno-geographical location of oil rather than oil income, rent, or relative distribution/concentration that really connects oil with the onset of ethnic war. When the core territory of a minority group has a significant amount of oil, the minority group is more likely to rebel against a central state dominated by another group and oil is strongly associated with the onset of ethnic war. In contrast, when oil is located with the core territory of a dominant majority group or that a country has a fairly even distribution of ethnic groups and no group can lay an exclusive claim to oil, oil is not associated with a higher risk of ethnic war. We construct two new indicators regarding the ethno-geographical location of oil from two different sources and test our hypotheses with the two new indicators. Statistical results strongly support our hypotheses. Together with evidences from case studies with process-tracing that demonstrate the mechanisms singled out by our theory really did operate in driving ethnic wars in an accompanying paper, we provide a more complete and close-to-definitive answer to the question whether and how oil is associated with the onset of ethnic war. Our exercise also points to a broader theory regarding the ethno-geography of commodity-type mineral resource with the onset of ethnic war.

Keywords Ethno-geography · Oil · Location · Ethnic war · Onset

H. Li and S. Tang contribute equally to the paper and their names are listed alphabetically.

Electronic supplementary material The online version of this article (doi:[10.1007/s41111-017-0062-2](https://doi.org/10.1007/s41111-017-0062-2)) contains supplementary material, which is available to authorized users.

✉ Shiping Tang
twukong@yahoo.com; twukong@fudan.edu.cn

¹ Fudan University, Shanghai, China

1 Introduction

Several recent studies have uncovered a conditional association between oil and the onset of ethnic war.¹ Challenging and improving upon them, we develop a more integrative and fine-grained theory regarding the conditional relationship between oil and the onset of ethnic war. We contend that it is the ethno-geographical location of oil that really connects oil with the onset of ethnic war, rather than a country's exact amount of oil income, rent (e.g., Collier and Hoeffler 1998, 2004), local production (e.g., Sorens 2011; Hunziker and Cederman 2012; Asal et al. 2016); or (relative) concentration and distribution (e.g., Le Billon 2001; Morelli and Rohner 2015).² More concretely, when the core territory of a minority group possesses a significant amount of oil, the minority group is more likely to rebel against a central state dominated by another group and oil is strongly associated with the onset of ethnic war. In contrast, when oil is located with the core territory of a dominant majority group or that a country has a fairly even distribution of ethnic groups and no group can lay an exclusive claim to oil, oil is not associated with a higher risk of ethnic war (for details, see Sect. 2 below).

To test our theory, we construct two new indicators that integrate the geographical location of oil with the geographical distribution of ethnic groups. Our new indicators avoid not only all the pitfalls of endogeneity associated with earlier measurements of oil rent or income (absolute or proportional) as incisively criticized by Ross (2006, pp. 275–277, 2012, pp. 17–22), but also some key pitfalls associated with more recent indicators on local production and relative concentration of oil (e.g., Sorens 2011; Hunziker and Cederman 2012; Morelli and Rohner 2015; Asal et al. 2016). We test our empirical hypotheses with our new indicators and statistical results strongly support our hypotheses.

In a forthcoming accompanying paper (Tang et al. 2017), we deploy comparative case studies with process-tracing to demonstrate that the mechanisms singled out by our theory really did drive ethnic wars, whereas ethnic peace usually results when these mechanisms lay dormant. Our case studies also strengthen our case against other studies that center upon local production or relative concentration of oil. Together, our two studies crystallize that it is the ethno-geographical location of oil that truly matters for the onset of ethnic war rather than oil alone, the amount of oil income/rent/(local) production,³ or relative concentration. Our studies thus provide a close-to-definitive answer to the question whether and how oil is associated with the onset of ethnic war. Our exercise also points to a broader theory that connects

¹ Following convention in the literature, we use oil to denote both oil and gas unless specified otherwise. By ethnic war, we mean an organized conflict between two ethnic groups with both groups having fielded an army or at least a militia and the total war causality has reached the threshold of 1000 deaths.

² Note that location is different from relative concentration or distribution. To our best knowledge, the only study that emphasizes the location of oil is Smith (2014). Smith, however, did not provide any systematic evidences and cites Sorens (2011) and Hunziker and Cederman (2012), which we criticize in more detail in Appendix E in supplementary material, as supporting evidences. Although Asal et al. (2016) seem to single out oil location, their key variable is really centered upon “(local) production”. Our study offers more integrated and fine-grained theorization and more robust and fine-grained empirical results than Asal et al. (2016). For details, see Sect. 3 below.

³ Although we only control for oil production per capita in our regressions, we indirectly control for oil income and rent by doing so because income and rent depend on oil production.

the ethno-geography of other commodity-type mineral resource with the onset of ethnic war.

Before we proceed further, a key caveat is in order. Many earlier studies on oil and conflict (ethnic or non-ethnic) deploy aggregate data at the country level and thus do not explicitly link the subnational properties of oil with the subnational location of ethnic groups. Most of these studies conclude that oil is generally a curse (for reviews, see Ross 2006, 2012, 2014; Blattman and Miguel 2010; Van der Pleog 2011). These results, however, have been questioned from time to time (some key studies are cited in Appendix E in supplementary material). Because our study deploys subnational data and explicitly links the subnational location of oil with the subnational location of ethnic groups, our study differs fundamentally from these earlier studies. We thus do not engage with this literature based on aggregate data extensively, although we do cite some of them when appropriate. We concur with Smith (2014) call that the two levels (i.e., national and subnational) should be synthesized. Indeed, our results below, which cover both the group-level and the state-level, provide a possible explanation why results from earlier studies with aggregate national data have been inconsistent.

The rest of the article is structured as follows. Section 2 outlines our new theory, lays out our empirical hypotheses, and paves the way for contrasting our exercise with several recent studies. Section 3 identifies the shortcomings within two recent studies that have reported seemingly similar results, focusing on theorization, measurement, and technical issues. Section 4 provides a brief description of our key explanatory variables and then presents quantitative results for our theory. Section 5 draws implication and concludes.

2 Oil and the Onset of Ethnic War: An Integrative Theory

In a series of excellent reviews (2004a, b, 2006), Michael Ross pointed out that despite some kind of consensus that oil does impact ethnic war (and non-ethnic civil war), results have been inconsistent within the first wave of literature. Ross reasoned that three deficiencies might have caused these inconsistencies (see also Hegre and Sambanis 2006; Smith 2014; Koubi et al. 2014). First, earlier studies stay with aggregate national level data, even though ethnic war is often a subnational phenomenon. Most prominently, earlier studies do not contain any geographical data regarding either resources or ethnic groups. Second, earlier studies have deployed oil production, income, and rent as instruments for oil and yet all these instruments suffer from serious problems of endogeneity (Humphreys 2005; Ross 2006; for earlier evidences, see Brunnschweiler and Bulet 2009; Mitchell and Thies 2012). Different authors have also used different data sets of civil war or ethnic war. Finally, many studies have proposed a wide range of causal mechanisms that purportedly link oil with ethnic war but have not explicitly tested these mechanisms.

The first deficiency has now been largely corrected, thanks to the pioneering study by Le Billon (2001) and Buhaug and Gates (2002). More and more recent studies of natural resources and civil wars now deploy Geographic Information

System (hereafter, GIS) data. As a result, our understanding of oil and ethnic war has become far more fine-grained.

Regarding the remaining two deficiencies, we hold that they are so tightly interconnected that we cannot expect to correct one without correcting the other, even with GIS data. Fundamentally, the two deficiencies are underpinned by inadequate theorizing about oil and the onset of ethnic war, aggravated by the lumping together ethnic civil war and non-ethnic civil war. Without adequate theorizing, we are often at a loss about what should be measured, how things should be measured, what mechanisms should be tested, and how they should be tested. Consequently, some clear-cut measurements of oil have not been tried: most authors still measure oil in quantity, as oil rents, oil vs. total GDP, oil vs. total export (e.g., Collier and Hoeffler 1998, 2004; Fearon and Laitin 2003), (local) oil production (e.g., Humphreys 2005; Brunnschweiler and Bulet 2009; Wimmer et al. 2009; Cederman et al. 2010; Hunziker and Cederman 2012; Asal et al. 2016), oil income per capita (Ross 2006, 2012), or (relative) concentration/distribution (Morelli and Rohner 2015). Meanwhile, besides several exceptions (e.g., Ross 2004a, pp. 38, 60–61; Smith 2014; Tang 2015; Paine 2016) insisting that some mechanisms are not mutually exclusive, most authors still pit different mechanisms against each other (i.e., greed, grievance, or opportunity) rather than integrating underlying factors, immediate drivers, and mechanisms into a coherent theory.

We thus aim to correct the two interconnected deficiencies together, by developing a more integrative theory regarding oil and the onset of ethnic war. Critically, building upon elements and insights from the existing literature on natural resources and civil conflict (esp. Le Billon 2001; Ross 2004a, 2012) and the literature on the nexus of ethnic domination/subordination and resentment/hatred (e.g., Cederman et al. 2013; Horowitz 1985; Wimmer 2013), we advance a more integrated theory regarding oil and the onset of ethnic war.

Our theory contends that it is the ethno-geographical location of oil that truly connects oil with the onset of ethnic war. When the core territory of a (subordinate) minority group possesses (a significant amount of) oil,⁴ this minority group is more likely to rebel against a central government dominated by another group, seeking greater autonomy or outright secession, *ceteris paribus*. Oil with such an ethno-geographical location is thus more likely to trigger an ethnic war or intensify an ongoing ethnic war.⁵ Consequently, countries with oil located within the core territories of minority groups are more likely to experience ethnic war. In contrast, when oil is located with the core territory of a dominant majority group or that a

⁴ By “the core territory of a minority group”, we are implying that the minority group is a concentrated group. Contrary Asal et al. (2016), we do not believe that a group has to be excluded from the center in order to rebel. Instead, we believe that subordination, exclusion, and downgrading can all contribute to a minority group’s grievance and hence the onset of ethnic rebellion. In an unpublished paper, Paine (2016) developed a formal bargaining model between the center and an ethnic group with oil, without assuming a minority group being excluded. Although Paine’s empirical results are consistent with what we report here, his empirical results are based on giant oil fields. Yet, ethnic groups may not need to possess giant oil fields in order to rebel. Although the quantity of oil must be significant enough for a minority group to contemplate rebellion, it is difficult to gauge a threshold quantity because resent can be constructed.

⁵ In another paper, we examine how the ethno-geographical location of oil intensifies ongoing ethnic wars.

country has a fairly even distribution of ethnic groups and no group can lay an exclusive claim to oil, oil is not associated with a higher risk of ethnic war.

Our theory further proposes two interlocked major mechanisms linking the ethno-geographical location of oil with ethnic war.⁶ On the one hand, when oil is discovered within the core territory of a minority group, the central government (dominated by another group or other groups) almost inevitably tries to control the resources for two reasons. First is simple economic interest (i.e., “greed”): every state desires to control more resources and revenues. Second, the central government seeks to preempt the minority group from controlling the resources and revenues partly because the central government fears that the minority group may seek greater autonomy. This is most severe when there had been earlier episodes of ethnic tension, or worse, earlier ethnic war between the group that dominates the central government and the minority group. These two dynamics almost inevitably lead to the central government to tighten its grip on the minority’s core territory and its oil, via (para-)military deployment, forced, or induced migration of the majority group to the core territory of the minority group, and usually both. Indeed, due to the technology and capital-intensive nature of oil production, even without encouragement from the central government, the extraction of oil almost inevitably attracts an influx of immigrant workers, usually in the form of ethnic aliens (from the majority group or other countries) with more technological and linguistic skills plus political and business connections. The result is an “internal colonialization” of the core territory of the minority group by the majority group (Hechter 1970), which, in turn, increases the risk of “sons of the soil” conflict (Weiner 1978; Fearon and Laitin 2011).

On the other hand, even without earlier episodes of ethnic tension and conflict, the minority group will resent the central government—dominated by another group—takes what rightly “belongs” to the minority group away from its natural owner. *Put it crudely, the minority group will inevitably hold that oil discovered within their core territory to be their oil.* The influx of immigrant workers as ethnic aliens and the fact that immigrants usually take up most of the high-pay jobs available only add to the resentment by the local minority group in the form of “relative deprivation”, partially driven by the fact or perception that the income gap between the minority group and the majority group widens. Worse, oil extraction, and processing almost inevitably entail severe environmental degradation, and oil companies, whether multinationals or state-controlled, almost never compensate the local people enough, and do enough to protect the environment. These dynamics generate greater resentment by the “native” minority group against the “alien” majority group. Finally, elites within the minority group can employ the expected oil revenue to broadcast the bright prospect of greater autonomy or secession, with looting and extorting being a source of income for financing the war effort if oil production has already commenced.

Oil located within the core territory of a minority group, therefore, impacts both the minority group and the state-controlled by another (majority) group. Most

⁶ Again, in Tang et al. (2017), we use case studies with process-tracing, which are better at establishing mechanisms, to demonstrate that these two mechanisms really do drive ethnic war.

critically, adding the two sides of the dynamics together results in a powerful mixture of immediate drivers of ethnic war. More specifically, oil located in the core territory of a subordinate minority group impacts fear of secession (by the majority), resentment (by the minority), interest or greed (both sides), and possibly capability (for the minority especially). *And if some hatred between the minority group and the majority group already exists, oil located within the core territory of a subordinate minority group would impact five of the seven immediate drivers of ethnic war which, in turn, will drive the two sides into a spiral of escalating tension and mutual distrust, and eventually war. As such, our theory predicts that oil located within the core territory of a (subordinate) minority group should be strongly and positively associated with the onset of secessionist ethnic war.*

From our new theory, two hypotheses for quantitative exercises,⁷ each with two subhypotheses, can be derived:

H1A-G: At the group level, when oil is located within the core territory of a minority group, this minority group is more likely to rebel against the central government dominated by another group, seeking greater autonomy or outright secession, ceteris paribus.

H1B-G: At the group level, oil located within the core territory of a minority group may have little impact on the risk of governmental civil war.

H2A-C: At the country level, countries with oil located within the core territory of a minority group are more likely to experience secessionist ethnic war, ceteris paribus.

H2B-C: At the country level, countries with oil located within the core territory of a minority group are no more likely to experience war of ethnic infighting that is for the control of the government.

3 A Critique of Existing Theorizing and Empirical Efforts

Oil and civil conflict is an extremely crowded field, and several studies have touched upon some of the elements within our theory and advanced one or two of the hypotheses we advanced here. They have also provided empirical results that on first glance seem to be similar our results reported below. Due to space limitation, we leave a more extensive survey of the literature to Appendix E in supplementary material. In this section, we critically focus on two recent studies, because they look the most similar to what we report here. We first show that a study by Morelli and Rohner (2015) is theoretically unsound and statistically un-robust (our replications of their results are in Appendix D in supplementary material). While explicitly noting the useful and valid factors, mechanisms, and theses from it, we also show that a study by Asal et al. (2016) too suffers from several weaknesses.

A common weakness of Morelli and Rohner (hereafter, M&R 2015) and Asal et al. (2016) is their reliance on the PETRODATA data set for coding oil location or production without acknowledging that data on oil discovery and production within

⁷ We lay out our qualitative hypotheses for case studies in Tang et al. (2017).

the PETRODATA data set have many missing observations. Thus, M&R's (2015, pp. 37) claim that their data set on oil discovery and production is complete which simply cannot be substantiated. Indeed, within the PETRODATA set, only 540 entries (out of 891) or 60.6% have data on discovery time; but only 309 entries (out of 891) or 34.8% have data on production time. In contrast, not only we use two different data sources (USGS and PETRODATA) to code our key independent variable, we obtained almost identical results with the two different data sources (see Sect. 4 below for details).

3.1 M&R's (2015) "Oil Gini" and the Onset of Civil War

M&R (2015) first built a simple bargaining model that addresses both group concentration and resource concentration. Their simple model points to the possibility that when group concentration and resource concentration are both high, groups then cannot credibly commit themselves to peaceful bargains under a variety of conditions, and civil war may result. Their model suggests a key empirical prediction "that conflicts are fuelled by non-governing ethnic minority groups living in very oil-rich regions...civil war is likely when resource discoveries happen in regions that are significantly populated by groups that do not belong to the governing coalition in the country."

Then, by merging Lujala et al. (2007) PETRODATA data set and the Geo-referencing of Ethnic Groups (GRED) data set by Weidmann et al. (2010), M&R (2015, pp. 33) created an indicator of "Oil Gini" to capture "the unevenness of oil field distribution across ethnic group for a given country and year" at the country level. At the group level, M&R invented an indicator of " $R1/R1 + R2$ " to capture a group's relative concentration of oil within a country. Within " $R1/R1 + R2$ ", $R1$ denotes resources (i.e., oil and gas) within the core territory of a minority group (group 1), whereas $R2$ denotes resources within the core territory of all other groups within a country. Hence, " $R1/R1 + R2$ " is "the surface of an ethnic group's territory covered with oil and gas as a percentage of the country's total surface covered with oil and gas" at the (ethnic) group level (ibid, 33). They report quantitative results that corroborate their model and hypotheses.

Unfortunately, M&R's theoretical and empirical exercises suffer from three critical shortcomings. Most critically, their key explanatory variables at both the country level (i.e., Oil Gini) and the group level (i.e., $R1/R1 + R2$) are problematic, because different situations may have the same "Oil Gini" value at the country level or the same $R1/R1 + R2$ at the group level, and yet, these situations may hold very different, even diametrically opposite, implications for ethnic war. Simply put, not all situations of (un-)equal distribution with the same Oil Gini score or the same $R1/R1 + R2$ value are the same.

First and at the country level, "Oil Gini" cannot differentiate two situations of extreme uneven distribution of oil and gas: both a situation in which all the oil is located within the core territory of a majority group and a situation in which all the oil is located within a minority region will have an "Oil Gini" value of one. Yet, these two situations will have diametrically opposite impact on the onset of ethnic war. In the former case, the probability that a minority group will rebel for the oil in

the territory of the majority group is almost zero and any ethnic war experienced by this country will have little to do with oil.⁸ In contrast, in the latter case, the probability that the minority group will demand to control the oil revenue will be extremely high and hence the chance of an ethnic war is going to be high, as suggested by M&R's own logic (see also Sect. 2 above). This suggests that the effect of "Oil Gini" on the onset of ethnic war may be non-monotonic, yet "Oil Gini" cannot cope with this possibility. For M&R (2015, pp. 32), the relationship is monotonically (and linearly) positive: "war [will] be more likely when resource concentration and group concentration are high."

Second, $R1/R1 + R2$, M&R (2015)'s measurement of uneven distribution of oil at the group level, is also problematic. According to them, $R2$ within $R1/R1 + R2$ contains all other groups except the first group. This neglects the fact that many multiethnic countries have more than one (minority) groups and each of the subordinate minority groups may stake a separate (and separatist) claim to the oil within its core territory. As such, what really matters is not one group versus the rest but one group versus the majority or dominant group. The case of Indonesia in which both Aceh and Irian Jaya/West Paupa held substantial oil and gas is one such example.

More critically, like their logic at the country level, M&R's logic at the group level suggests that $R1/R1 + R2$ must be high enough to trigger the onset of civil war (non-ethnic and ethnic war), although they fall short of specifying a threshold level for $R1/R1 + R2$. Their logic thus ignores the possibility that a subordinate minority group may possess only a small fraction of the total oil and gas reserve in a country and yet may still rebel, because the group deems the small fraction now being controlled by the dominant majority group as a sufficient *casus belli*. Aceh in Indonesia has been such a tragic example. According to M&R's own measurement, Aceh's $R1/R1 + R2$ value is only 0.027, a very small value. Yet, Hasan di Tiro, the former leader of Free Aceh Movement still called for Acehnese's rebellion against "the neo-colonialist Javanese state", because gas in Aceh belongs to Acehnese (di Tiro 1984, for more detailed discussion of the Aceh-Indonesia conflict, see Tang et al. 2017 and the references cited there).

Third, although M&R's two major independent variables are strictly related to ethnic groups—their claim that their model does not depend on ethnicity notwithstanding (M&R 2015, pp. 34, fn. 12), their dependent variables at both the country level and the group level are civil war onsets and incidences, thus including both ethnic wars and non-ethnic civil conflicts. This is not only theoretically inconsistent, but can seriously bias estimations. Moreover, in most of their regressions, they use OLS rather than logit or probit models to estimate the impact of oil concentration upon the onset of civil war, even though it is widely acknowledged that OLS estimation tends to generate smaller standard error thus exaggerates the significance of explanatory variables when the dependent variable is binary, multinomial, or ordinal.

⁸ The Nagorno Karabakh conflict between Armenians and Azerbaijanis in Azerbaijan is a perfect example of such a case (Kaldor 2007). For a more detailed discussion, see our accompanying qualitative paper (Tang et al. 2017). Here, suffice to say that the case supports our theory while contradicting M&R's theory.

Unsurprisingly, M&R's results that supposedly show that oil concentration has a robust and significant effect upon the onset of civil war are at both the country level and the group level are extremely fragile, contrary to their claims of robustness, as our replications of their results clearly show (see Appendix D in supplementary material for detail).

3.2 Asal et al. (2016)

On surface, our study is also quite similar to Asal et al. (2016). Upon closer look, however, it is evident that our study betters Asal et al. (2016) both theoretically and empirically.

First, our theorization is both more integrated and more fine-grained than theirs. Contrary Asal et al. (2016), we do not hold that a group has to be excluded from the center to rebel. This is so because a minority group, by definition, is almost always under-presented in the center and ex ante grievance can amplify the degree of a minority group's feeling of being dominated or excluded. We also explicitly differentiate not only ethnic civil war from non-ethnic civil war but also secessionist ethnic war from ethnic infighting. We then theorize that oil located with the core territory of minority group impacts only secessionist ethnic war, but not non-ethnic civil war or ethnic infighting (see Sect. 2).

Second and consistent with our theory, we provide more fine-grained results, showing that oil located with the core territory of minority group impacts only secessionist ethnic war, but not non-ethnic civil war or ethnic infighting (see Sect. 4).

Third, although Asal et al. (2016) hypothesized that the interactive term between political exclusion and oil location drives the onset of ethnic war, they initially failed to obtain such clear-cut results initially (10–11). As a result, they had to use a less straightforward technique to obtain results that can support their hypotheses. In contrast, our empirical results have consistently been supporting our hypotheses (see Sect. 4). In addition, for some reason, Asal et al. (2016) did not report country-level results.

Finally, econometric analyses by Asal et al. (2016) leave much room for improvement, to say the least. For instance, they did not perform regressions for rare events. In contrast, we have performed a variety of robustness checks and obtained positive results consistently.

3.3 Summary

Despite their various merits, the above-mentioned two studies have some shortcomings. In contrast, our empirical exercises are consistent with our theoretical exercises, because the ethno-geographical location of oil—our key explanatory variable—is firmly underpinned by our new theory. Moreover, our key explanatory variable is almost entirely exogenous, clear-cut, non-arbitrary, which have few missing data points. As becomes clear in Sect. 4, our key explanatory variable, oil location, is significant across all specifications with an array of control variables, at both the country level and the group level.

4 Quantitative Data and Results

Humphreys (2005, pp. 518–522) complained that earlier indicators of oil (and natural resources) are simply too aggregated for differentiating the different mechanisms through which natural resources impact civil war. Accordingly, Humphreys suggests that one solution is to construct more fine-grained indicators of oil (and civil war) that allows for finer resolution of mechanisms. We concur with Humphreys (2005) but go further. We believe that a more valid solution is to construct clear-cut and non-arbitrary measurement of oil (and natural resources), informed by rigorous theorizing. Existing indicators of oil, including oil rent, oil versus export (Collier and Hoeffler 2004), oil export dummy (Fearon and Laitin 2003), (local) oil production (Humphreys 2005; Hunziker and Cederman 2012; Sorens 2011), 100 US\$ per capita oil income as the cutoff point (Ross 2006, 2012), or relative distribution/concentration (M&R 2015) are endogenous to conflict, arbitrary, or invalid. Moreover, these indicators are not underpinned by a sophisticated theory that links oil with the onset of ethnic war.

Our theory explicitly posts that it is whether a significant amount of oil is located within the core territory of a (subordinate) minority group that really connects oil with the onset of ethnic war. We construct two new explanatory variables to capture this ethno-geographical location of oil. Our two explanatory variables are not only clear-cut dichotomous variables with few missing data points, but also avoid most, if not all, of the pitfalls associated with earlier indicators of oil (e.g., endogeneity, arbitrariness).

4.1 Sample and Key Explanatory Variables

Following our theory, we restrict our sample to multiethnic countries at the country level and minority groups that are concentrated with a core territory at the group level.

At the country level, we rely on the Geo-Ethnic Power Relations (GeoEPR-ETH) data set by Wucherpfennig et al. (2011), which is the GIS-informed version of Ethnic Power Relations (EPR) data set (Cederman et al. 2009; Wimmer et al. 2009). Countries with a homogenous population (e.g., the two Koreas) are excluded from the GeoEPR-ETH data set, because they, by definition, cannot experience ethnic war. The GeoEPR-ETH data set covers 125 multiethnic countries with a population more than half million, including 110 onsets of ethnic war from 1946 to 2005. Our final data sets cover the same 125 countries from 1946 to 2005 (group level) and from 1946 to 2010 (country level).

At the (ethnic) group level, following our theory and Wucherpfennig et al. (2011, pp. 428–429), we restrict our sample to minority ethnic groups with a territorial base, that is, minority groups that are coded as 1 (regionally based), 3 (regional and urban), and 6 (aggregate) in the GeoEPR-ETH data set (Wucherpfennig et al. 2011; Bormann 2011). Our final data set covers 491 minority groups within the 125 multiethnic countries.

Within these samples, following our theory, we construct two new explanatory variables, one at the group level and one at the country level.

At the group level (variable name: oil location, group), for each minority ethnic group with a territorial base, we code it 1 when a group's core territory has oil and 0 when it does not.

At the country level (variable name: oil location, country), we code it 1 when a multiethnic country has oil located within core territories of minority group(s) and 0 when (1) a country does not have any oil, (2) although it has oil, its oil is located within the territory of the dominant majority group, and (3) the country, despite being multiethnic, has no minority group concentrated within a particular region (i.e., the country is evenly occupied by a mixture of ethnic groups). At the country level, we also restrict our sample to only countries with oil as a robustness check. Results with this smaller sample are essentially identical to results with the sample of all countries (see Table 4).

We construct the two key explanatory variables by merging the information on the geographical location of ethnic groups in the GeoEPR-ETH with either the information on the geographical location of *oil basins* from U.S. Geological Survey (hereafter, USGS) or the information on the geographical location of *actual oil fields* in the PETRODATA data set by Lujala et al. (2007). By so doing, we are able to locate most major oil basins or major oil fields into specific geographical locations and then determine whether oil basins cover the core territory of a minority group (i.e., oil basin only covers the core territory of a majority group or there is no concentration of minority groups), or whether oil fields are located within the core territory of a minority group.

The reason why we have two sets of key explanatory variables, one underpinned by data on oil basins whereas the other by data on actual oil fields, requires a bit explanation.

USGS provides the most complete coverage on all known oil basins in the world. *Yet, oil basin does not mean actual oil deposit, and certainly does not mean actual oil discovery and production.* Thus, using the USGS data on oil basin also allows us to avoid at least some endogeneity in oil discovery and production (Melando 2014; cf. M&R 2015; Asal et al. 2016). Moreover, the USGS data should overestimate the number of (minority) ethnic groups with actual oil field(s). Logically, this means that the explanatory variable constructed from the USGS data (oil location, USGS) should point to a diminished impact on the onset of ethnic war by the ethno-geographical location of oil. Consequently, if our hypotheses are supported even with USGS data, we shall have great confidence that the ethno-geographical location of oil is positively associated with the onset of ethnic war in the real world.

The PETRODATA data set by Lujala et al. (2007) represents a major advancement: for the first time, geographical information of oil–gas fields based on GIS data has been brought into the study of civil conflict. The basic entry unit in the PETRODATA data set is oil (and gas) field either with a discovery date or a production date.

The PETRODATA data set, however, suffers from a severe problem of missing data. Within the data set, of the 891 entries of oil fields, only 540 entries (out of 891) or 60.6% have data on discovery time; but only 309 entries (out of 891) or 34.8%

have data on production time.⁹ For our purpose here, this severe problem of missing data means that we cannot possibly know whether the PETRODATA data set underestimates or overestimates the number of (minority) ethnic groups with actual oil fields located in their core territories. Yet, if our quantitative hypotheses are also supported by the PETRODATA data set, in addition to being supported by the USGS data, we shall have great confidence that our hypotheses hold in the real world. We thus use the results with the PETRODATA data set as additional robustness checks (reported in Appendix C in supplementary material). Here, suffice to say that results with the PETRODATA data set are almost identical to the results with the USGS data set (see immediately below): both sets of results strongly support our quantitative hypotheses.

To improve our coding of the two key explanatory variables based on the PETRODATA data set, we also employ information within the large map collections from the University of Texas Perry-Castaneda Library Map Collections. Some of the high resolution maps from this collection provide not only information on geographical information of ethnic groups but also geographical information of major oil and gas fields. Whenever possible, we also use other information (e.g., Li 2011; *Petroleum Economists* by World Energy Atlas, *Oil and Gas Journal*, and internet) to further improve the PETRODATA data set.¹⁰

We admit one key drawback of our key explanatory variable (i.e., oil location): it is a dichotomous variable that is often time-invariant. As both a compensating measure and a robustness check, we multiply it with oil price from 1946 to 2005 (for group level) or 2010 (for country level), in nominal dollars per million British Thermal Units of natural gas priced at the Henry Hub in Louisiana compiled by Ross and Mahdavi (2015) from *British Petroleum (BP) Statistical Review* and the *Economist Intelligence Unit (EIU)*. We reason that oil price is mostly exogenous to any specific ethnic or non-ethnic civil war within a country. Rather, oil price has mostly been driven by consumption and overall production in the world and key interstate wars and their aftermath (e.g., the Fourth Israel–Arab War, Iraq’s invasion of Kuwait, and U.S. invasion of Iraq in 2003) rather than by any particular ethnic or non-ethnic civil war. We thus use the product of oil location and oil price as another key explanatory variable for another set of robustness check (see Appendix B in supplementary material). Here, suffice to say that using the product oil location and oil price as the key explanatory variable yields almost identical results to models with oil location as the key explanatory variable.

We, however, refrain from artificially making our key explanatory variables more time-variant by multiplying them with a country’s oil production (per capita), rent, or value, because these instruments of oil suffer from serious endogeneity problem with ethnic and non-ethnic civil war (Humphreys 2005; Ross 2006, 2012; for earlier empirical evidences, Brunnschweiler and Bulet 2009; Mitchell and Thies 2012). Indeed, our own results reported below also indicate that at least at the national

⁹ Our calculation is based on the latest version (version 1.2, 2009) of the PETRODATA dataset. Thus, our numbers are slightly different from what Lujala et al. (2007, pp. 245) reported.

¹⁰ We are now preparing to release our updating of the PETRODATA dataset.

level, oil production per capita is not significantly associated with the onset of ethnic war after controlling for oil location.

We believe that our key explanatory variables, short of more accurate information on the date of discovery and production or the amount of oil extracted from the core territory of minority groups (which is difficult, if not entirely impossible, to obtain), represent the most accurate and comprehensive indicators on the ethno-geographical location of oil so far.¹¹

4.2 Dependent Variables and Control Variables

Our key dependent variable is the onset of ethnic war. Our dependent variables also have two levels: group and country. To facilitate interpretation and comparison with results reported by earlier studies, in addition to the coding of the onset of civil war in the UPCD/PRIO Armed Conflicts Data set (ACD data set, version 4-2009, for details on the ACD data set, see Gleditsch et al. 2002; Harbom and Wallensteen 2009), we also adopt the coding of the EPR data set. The EPR data set not only singles out the onset of ethnic war (variable name: Ethonset), but also refines the coding of ACD by further differentiating ethnic war into infighting ethnic war between groups within the central government from rebellion ethnic war by group not within the central government (Wimmer et al. 2009; Cederman et al. 2010). According to our theory, we shall expect the ethno-geographical location of oil to be significantly associated with only ethnic or secessionist civil war but not with governmental civil war at the group level (H1A–G and H1B–G). At the country level, we shall expect the ethno-geographical location of oil to be significantly associated with only rebellion ethnic wars but not with infighting ethnic wars (H2A–C and H2B–C). Description of control variables at both the group level and the country level is found in Appendix A in supplementary material, due to space limitation.

Following the recommendations of Beck et al. (1998) and Carter and Signorion (2010), cubic splines for peace years are introduced to reduce possible biases in working with binary or multinomial panel data at both the group level and the country level, although they are not shown in the tables for the sake of space.

All descriptive statistics and explanations of the key independent variables, dependent variables, and control variables are in Appendix A in supplementary material.

4.3 Model Specifications and Results

At the group level, our main specification function is:

$$\text{Onset}_{i,t} = \alpha \text{Oil_Loc}_{i,t} + \beta \text{Group}_i + \gamma \text{Year}_t + \varepsilon_{i,t}, \quad (1)$$

where $\text{Onset}_{i,t}$ is the onset of ethnic war with a group i in year t , $\text{Oil_Loc}_{i,t}$ is the location of oil at the group level or the country level (our key independent variable),

¹¹ Because our new indicators integrate the ethnic and geographical location of oil and gas, we name our new dataset the EGLOG dataset. A more detailed description of the dataset will be presented elsewhere. In phase II, we aim to integrate data on discovery and production dates, improving upon Lujala et al. (2007).

Group_{*i*} is the group fixed effect, Year_{*t*} is the year fixed effect, and $\varepsilon_{i,t}$ is the error term. Standard errors are allowed to cluster according to group.

At the country level, our main specification function is:

$$\text{Onset}_{i,t} = \alpha \text{Oil_Loc}_{i,t} + \beta \text{Country}_i + \gamma \text{Year}_t + \varepsilon_{i,t}, \quad (2)$$

where Onset_{*i,t*} is the onset of ethnic war within a country *i* in year *t*, Oil_Loc_{*i,t*} is the location of oil at the country level (the key independent variable), Country_{*i*} is the country fixed effect, Year_{*t*} is the year fixed effect, and $\varepsilon_{i,t}$ is the error term. Standard errors are allowed to cluster according to country.

Because our dependent variables are either dichotomous or trichotomous, we employ binary or multinomial logistic regression throughout. Because results at the group level drive results at the country level, we present results at the group level first followed by results at the country level below. In the main text, we report only regular logit results with a minimal number of tables. More robustness checks, including those with penalized maximum likelihood logistic regression that check rare events biases (Firth 1993),¹² are reported in the online appendixes. Here, suffice to note that our results are extremely robust.

Table 1 shows results at the group level with the key explanatory variable being oil location constructed from USGS data [denoted as *oil location (USGS, group)*]. In model 1 (the baseline model), we first test the explanatory power of *oil location (USGS, group)* when the dependent variable is onset of ethnic war as recorded by the EPR data set, with a minimal number of control variables to facilitate interpretation (Ray 2003). As expected, *oil location (USGS, group)* is positively and significantly associated with the onset of ethnic war (model 1). In terms of odds ratio, model 1 suggests that the probability that a minority group with oil located in its core territory will rebel against a state is 1.92 times of the probability that a minority group without oil will rebel. This result holds in model 2 even after we control for a battery of control variables as conventionally controlled in other group-level studies (e.g., Cederman et al. 2010; Weidmann et al. 2010; Wucherpfennig et al. 2011).

The more interesting results appear in models 3 and 4, which are multinomial logit models. After dividing civil war within the ACD data set into two types: territorial (or ethnical secessionist) and governmental (i.e., aiming for the control of central government, thus non-secessionist), striking results emerge: whereas *oil location (USGS, group)* remains positively and significantly associated with the onset of ethnic civil war, it is insignificant with the onset of governmental (i.e., non-ethnic) civil war. In terms of relative risk ratio, model 3 suggests that in a given group year, groups with oil location coded as 1 are 3.15 times more likely to experience an onset of territorial (ethnic) war than groups with oil location coded as 0 but are only 1.14 times more likely to experience an onset of governmental (non-ethnic) war than groups with oil location coded as 0.

These results strongly support the core argument of our theory that the ethnogeographical location of oil impacts only the onset of secessionist/ethnic war but not non-ethnic or non-secessionist (i.e., governmental) civil wars. As far as we can tell,

¹² Firth's penalized maximum likelihood logistic regression actually came before the more known King-Zeng (2001) solution in political science, and it is more readily implemented in STATA.

Table 1 Ethno-geographical location of oil and the onset of ethnic conflict, group-level, with USGS data set (dependent variable is onset of ethnic conflict experienced by a group, 1945–2005)

	Model (1)		Model (2)		Model (3)		Model (4)	
	Ethonset (EPR)		Ethonset (EPR)		Ethonset (ACD)		Ethonset (ACD)	
					Territorial		Territorial	
					Governmental		Governmental	
Oil location (USGS, group)	0.66** (0.29)		0.48** (0.24)		1.15*** (0.39)		0.93** (0.40)	
Distance to capital (ln)			0.15** (0.069)				0.25*** (0.093)	
Distance to border (ln)			−0.15** (0.065)				−0.24** (0.099)	
Excluded			1.28*** (0.28)				1.07*** (0.41)	
Downgraded2			1.75*** (0.38)				1.20** (0.55)	
Size of core territory			0.27*** (0.087)				0.10 (0.11)	
War history			0.71*** (0.16)				0.51** (0.26)	
GDP per capital (ln)			−0.37*** (0.13)				−0.33 (0.22)	
Population size (ln)			−0.11 (0.12)				−0.031 (0.15)	
Peace years	−0.22*** (0.074)		−0.10 (0.075)		−0.37*** (0.097)		−0.17* (0.087)	
Constant	−4.27*** (0.26)		−1.75 (1.32)		−4.74*** (0.35)		−4.11** (1.91)	
Pseudo R^2	0.029		0.11		0.038		0.13	
N	24,791		19,749		24,791		19,749	
Wald χ^2	35.2***		147.9***		54.1***		442.0***	
Group FE	Yes		Yes		Yes		Yes	

Models 1 and 2 are logit model and model 3 and 4 are multinomial logit model

Robust standard errors clustered according to group in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Splines of peace years are controlled but not shown

we are the first to report such strikingly differentiating results when using oil as a key explanatory variable in the study of ethnic civil war at the global level. These results also strongly support the notion that ethnic civil wars have some fundamental differences from non-ethnic civil wars (e.g., Horowitz 1985; Wimmer 2013), and it may not be always wise to study them together as they are fundamentally similar (e.g., Fearon and Laitin 1996; Walter 2001).

In Table 2, we focus on the onset of ethnic war at the country level, as recoded by the EPR data set. Again, the results strongly support our theory and hypotheses. Model 1 is our baseline model, and our key explanatory variable *oil location (USGS, country)* is positively and significantly associated with onset of ethnic war. In terms of odds ratio, model 1 of Table 2 suggests that the probability that a country with oil location coded 1 will experience an onset of ethnic war is 4.19 times of the probability that a country with oil location coded as 0. Again, this result holds as we add more and more control variables progressively (models 2 and 3). In models 4 and 5, we drop ongoing wars, and *oil location (USGS, country)* remains positively and significantly associated with the onset of ethnic war. Our hypothesis H2A-C is thus strongly supported. Note, however, that oil production per capita is insignificant.

In Table 3, we move to more fine-grained analyses of onset of ethnic war, again at the country level. Here, ethnic conflict is divided into two categories: ethnic infighting among power-holders (infighting) and rebellion (i.e., an excluded ethnic group rebels against the state), according to the EPR data set. Again, the results strongly support our theory and hypotheses. In model 1 (the baseline model), *oil location (USGS, country)* remains positively and significantly associated with the onset of ethnic rebellion but not with the onset of infighting. In terms of relative risk ratio, model 1 suggests that in a given country year, countries with oil location coded as 1 are six times more likely to experience an onset of rebellion ethnic war than countries with oil location coded as 0 but are only 1.6 times more likely to experience an onset of infighting ethnic war than countries with oil location coded as 0. Again, the overall result holds as we add more and more control variables progressively in model 2 and model 3.

Thus, our key explanatory variable (i.e., oil location) captures the different impact of oil location upon two different types of ethnic war (i.e., infighting vs. rebellion). Again, as far as we can tell, we are the first to report such strikingly differentiating results when using oil as a key explanatory variable in the study of ethnic war. Our hypothesis H2B-C is strongly supported. Again, note that oil production per capita is insignificant.

In Table 4, we restrict the sample to country-years with actual oil production (i.e., country-years without actual oil production are dropped from the sample) and replicate the models in Tables 2 and 3. Results with a reduced sample are mostly consistent with results in Tables 2 and 3. When the dependent variable is the onset of ethnic war at the country level, oil location retains its significance although the level of significance is reduced after adding more control variables (model 1 and model 2). When ethnic war is differentiated into infighting and rebellion/secessionist, oil location is only significantly associated with rebellion ethnic wars

Table 2 Ethno-geographical location of oil and the onset of ethnic conflict, country-level (dependent variable is onset of ethnic conflict experienced by a country, 1946–2010)

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Oil location (USGS, country)	1.43*** (0.32)	0.78*** (0.29)	0.75** (0.32)	0.79** (0.34)	0.73** (0.36)
Ratio of excluded population		0.36*** (0.084)	0.25*** (0.080)	0.38*** (0.097)	0.30*** (0.099)
Segmentation		0.15*** (0.032)	0.098** (0.040)	0.17*** (0.043)	0.13* (0.067)
Imperial past		0.37 (0.40)	0.68* (0.40)	0.90* (0.53)	1.10** (0.52)
GDP per capita (ln)		−0.12*** (0.035)	−0.10** (0.040)	−0.084*** (0.031)	−0.062* (0.035)
Size of population (ln)		0.15** (0.066)	0.13* (0.068)	0.098 (0.077)	0.095 (0.086)
Ongoing war		0.21 (0.57)	0.18 (0.59)		
Year	0.0067 (0.0057)	0.010 (0.0062)	0.0079 (0.0063)	0.017*** (0.0061)	0.014** (0.0067)
ELF			1.30*** (0.48)		1.06 (0.67)
Area of mountainous terrain (ln)			0.19** (0.092)		0.096 (0.10)
Regime change			0.12 (0.25)		0.29 (0.26)
Anocracy			0.30 (0.22)		0.39 (0.28)
Oil production per capita (ln)			0.0049 (0.037)		−0.000095 (0.038)
No. of peace years (country)	0.0018 (0.11)	0.18 (0.21)	0.18 (0.22)	0.23* (0.14)	0.23* (0.13)
Constant	−17.9 (11.3)	−27.22** (12.3)	−23.70* (12.5)	−40.47*** (12.1)	−36.29 (13.2)
Pseudo R^2	0.073	0.13	0.14	0.12	0.14
N	7843	7661	7566	6863	6773
Wald χ^2	54.8***	243.6***	190.6***	115.4***	145.7***
Country FE	Yes	Yes	Yes	Yes	Yes

Logit models; model 4 and model 5 drop ongoing wars; Robust standard errors clustered according to country; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; Splines of peace years are controlled but not shown

Table 3 Ethno-geographical location of oil and the onset of *ethnic conflict*, country-level (dependent variable is onset of ethnic conflict experienced by a country, 1946–2010, differentiated into ethnic infighting and ethnic rebellion, according to EPR)

	Model (1)		Model (2)		Model (3)	
	Infighting	Rebellion	Infighting	Rebellion	Infighting	Rebellion
Oil location (USGS, country)	0.48 (0.47)	1.80*** (0.36)	0.050 (0.50)	0.90*** (0.34)	0.18 (0.58)	0.94*** (0.37)
Ratio of excluded population (ln)			−0.036 (0.15)	0.49*** (0.089)	−0.21 (0.16)	0.46*** (0.089)
Segmentation in the center			0.31*** (0.062)	0.040 (0.041)	0.30*** (0.063)	0.049 (0.047)
Imperial past			2.57** (1.04)	0.043 (0.40)	3.05*** (1.17)	0.13 (0.44)
ELF			−0.18 (1.18)	1.38*** (0.53)	1.08 (1.02)	1.33*** (0.50)
GDP per capita (ln)			−0.16** (0.081)	−0.090** (0.035)	−0.15** (0.076)	−0.092** (0.041)
Population size (ln)			−0.28** (0.12)	0.32*** (0.076)	−0.41*** (0.12)	0.30*** (0.086)
Ongoing war			1.09 (1.05)	0.065 (0.67)	0.93 (0.99)	0.065 (0.68)
Year	0.024** (0.011)	0.0024 (0.0058)	0.029** (0.013)	0.0034 (0.0069)	0.024* (0.013)	0.0037 (0.0072)
Area of mountainous terrain (ln)					0.47*** (0.18)	0.093 (0.11)
Regime change					0.039 (0.57)	0.14 (0.29)
Anocracy					0.36 (0.40)	0.36 (0.23)
Oil production per capita (ln)					0.014 (0.026)	0.0070 (0.047)
No. peace years (country)	0.21 (0.24)	−0.032 (0.12)	0.49 (0.36)	0.14 (0.26)	0.36 (0.33)	0.15 (0.26)
Constant	−53.7** (22.6)	−9.81 (11.4)	−63.89** (26.7)	−16.55 (13.5)	−54.22** (26.2)	−17.38 (14.1)
Pseudo R^2	0.075		0.16		0.17	
N	7843		7661		7566	
Wald χ^2	54.8***		320.3***		330.8***	
Country FE	Yes		Yes		Yes	

Multinomial logit models; splines of peace years are controlled but not shown

Robust standard errors clustered according to country in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 Replication of country-level results within only oil producing countries (dependent variable is onset of ethnic conflict experienced by a country, 1946–2010, differentiated into ethnic infighting and ethnic rebellion, according to EPR)

	Model (1)		Model(2)		Model(3)		Model(4)	
	Ethonsset (EPR)		Ethonsset (EPR)		Infighting	Rebellion	Infighting	Rebellion
Oil location (USGS)	2.30*** (0.77)		1.95* (1.00)		1.51 (0.92)	2.54*** (0.78)	3.12** (1.25)	2.13** (1.07)
Ratio of excluded population			0.056 (0.14)				−0.48* (0.27)	0.19 (0.14)
Segmentation			0.061 (0.040)				0.54*** (0.18)	−0.0014 (0.038)
Imperial past			1.78*** (0.67)				4.50 (3.86)	1.48** (0.66)
GDP per capita (ln)			−0.049 (0.038)				−0.13* (0.073)	−0.039 (0.039)
Population size (ln)			0.042 (0.12)				−1.34*** (0.29)	0.19 (0.14)
Ongoing war			0.31 (0.89)				3.93*** (1.05)	0.029 (0.91)
Year	−0.0065 (0.0083)		−0.0064 (0.0097)		0.021 (0.022)	−0.012 (0.0074)	0.021 (0.042)	−0.010 (0.0095)
ELF			2.14*** (0.67)				0.021 (1.81)	2.45*** (0.71)
Area of mountainous terrain (ln)			0.30** (0.13)				0.92** (0.37)	0.25* (0.14)
Regime change			−0.17 (0.40)				−13.8*** (0.52)	−0.018 (0.41)
Anocracy			0.43 (0.40)				0.57 (0.89)	0.32 (0.42)
Oil production per capita (ln)			−0.012 (0.048)				−0.16 (0.11)	−0.0014 (0.049)
No. of peace years (country)	−0.15 (0.15)		0.059 (0.34)		−0.040 (0.45)	−0.16 (0.17)	0.82 (0.59)	0.028 (0.37)
Constant	7.50 (16.4)		3.25 (19.4)		−49.0 (44.7)	17.8 (14.6)	−45.5 (83.6)	9.26 (18.5)
Pseudo R^2	0.085		0.18		0.082		0.21	
N	3794		3735		3794		3735	
Wald χ^2	18.7**		233.8***		73.7***		12.843.2***	
Country FE	Yes		Yes		Yes		Yes	

Models 1 and 2 are logit model and model 3 and 4 are multinomial logit model

Robust standard errors clustered according to country in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Splines of peace years are controlled but not shown

in the baseline model (model 3), although it becomes significant with both infighting and rebellion after more control variables are added (model 4).

At the group level and the country level, signs in front of control variables are highly consistent with the results reported by Cederman et al. (2010), Wucherpfennig et al. (2011) and Wimmer et al. (2009). Variables that are significant in their regressions remain significant in our regressions, and their directions remain steady. For instance, at the group level (Table 1), both *war history* (i.e., previous conflict) and distance between the core territory of a minority group and a state's capital (*cap distance*) are positively and significantly associated with onset of ethnic war, whereas the nearest distance between the core territory of a minority group and a foreign country (*border distance*) is negatively and significantly associated with onset of ethnic war. Two political variables from the EPR data set that have been consistently found to be and positively associated with onset of ethnic war (Cederman et al. 2010; Weidmann et al. 2010; Wucherpfennig et al. 2011), “excluded” and “downgraded2”, are also significantly and positively associated with onset of ethnic war at the group level in our regressions (see also Asal et al. 2016). Meanwhile, at both the group level and the country level, GDP per capita is significantly and negatively associated with the onset of ethnic war. Interestingly, at the country level, after controlling for the ethno-geographical location of oil, oil production per capital is no longer significant (Tables 2, 3, 4). This result further strengthens our core argument that it is the ethno-geographical location of oil rather than oil production per se that truly connects oil and the onset of ethnic war.

We perform several sets of robustness check and these results are reported in Appendixes B and C in supplementary material, respectively. In Appendix B of supplementary material, we check our results with USGS by employing the product of oil location and oil price as the key explanatory variables. We also check our results using penalized maximum likelihood logistic regression that check rare events biases (Firth 1993). In Appendix C of supplementary material, we replicate all the regressions with the PETRRODATA data set and obtain almost identical results with the USGS data. As noted above, this fact should give us great confidence that our empirical hypotheses hold in the real world.

4.4 Summary

To summarize, quantitative evidences based on our new indicators strongly support our quantitative hypotheses. At the group level, the ethno-geographical location of oil is strongly and positively associated with the onset of secessionist ethnic war, but not with the onset of non-ethnic (i.e., governmental) ones. At the country level, the ethno-geographical location of oil is strongly and positively associated with the onset of rebellion ethnic war, but not with the onset of ethnic infighting. To our knowledge, we are the first to report such fine-grained and conclusive evidences regarding oil and the onset of ethnic war with a global data set. These results are highly robust across a wide variety of robustness checks (see Appendixes B and C in supplementary material for details). By comparison, earlier results reported by M&R's (2015) study are quite fragile (see Appendix D in supplementary material for details), mostly likely due to the questionable logic of their key explanatory

variable (i.e., Oil Gini and $R1/R1 + R2$). In addition, results reported by Asal et al. (2016) are not as fine-grained as what we report here (see also Paine 2016).

5 Discussion and Conclusion

(Formerly united) Sudan and Nigeria are similar on least one front: although both have plenty of oil, most of the oil in both countries is located in the core territories of subordinate minority groups. Both countries had experienced bloody secessionist ethnic wars, and one of them ended in the formal breakup of the country (Sudan). Indonesia is another oil-rich country, and most of its oil is located within the territory controlled by its two dominant majority groups (Javanese and Sundanese). Even though only a small proportion of its gas reserve is located within the province of Aceh populated by a subordinate minority group (the Acehnese), Indonesia too had experienced a savage ethnic war. Yet, Gabon, another Africa country with plenty of oil, significant ethnic diversity, little democracy, but an even distribution of ethnic groups, ethnic peace has prevailed so far. Thus, although all these four multiethnic countries are major oil producers, their encounters with the specter of ethnic war have differed greatly. *The impact of oil on ethnic war is thus conditional, rather than an inescapable “(oil) curse”.*

We have advanced a more integrated and fine-grained theory regarding the conditional association between oil and the onset of ethnic war, singling out the ethno-geographical location of oil as the key in linking oil with the onset of ethnic war. We then present quantitative evidences to support our theory. Along the way, we correct key shortcomings of several recent studies. To our knowledge, we are the first to report such fine-grained and conclusive evidences regarding the ethno-geographical location of oil and the onset of ethnic war. Together with evidences from comparative case studies with process-tracing that demonstrate the mechanisms singled out by our theory really did operate in driving ethnic wars in an accompanying paper (Tang et al. 2017), we provide a more complete and close-to-definitive answer to the question whether and how oil is associated with the onset of ethnic war. Our theory and empirical results hold important implications for understanding the “oil curse” (and the broader “natural resource curse”) and ethnic war in general.¹³

First and foremost, our theory and evidences point to a broader theory regarding mineral resources and ethnic war. Whenever a significant chunk of commodity-type mineral resources (oil, gas, copper, gold, and diamond) is located in the core territory of a minority group, that group is more likely to rebel, especially if the group has been marginalized or dominated by the central government, all else being equal. As such, a state with significant commodity-type mineral resources located within the core territories of minority groups is more likely to experience ethnic war, all else being equal. We are now in the process to extend our theory in this direction with more systematic evidences.

¹³ Obviously, our exercises also point to some key policy implications for preventing ethnic war. For details, see our accompanying qualitative paper.

Second, although we do not engage with the other famous thesis linking oil (and other commodities) with civil war, that is, the “weak state (capacity) thesis” by Fearon and Laitin (2003), our studies cast some doubt on the “weak state (capacity) thesis”. Our studies suggest that “weak state capacity” may not be the primary channel or mechanism through which oil impacts ethnic war (see also Smith 2014). Our studies, however, do point to some possible directions for investigating the interaction between state capacity and the ethno-geographical location of oil and other commodity-like mineral sources and how this interaction impacts ethnic conflict. For instance, admitting that state capacity is endogenous to both ethnic civil war and non-ethnic civil war (Thies 2010), it will be interesting to examine that whether states may intentionally devote less resource to necessary infrastructure in restless ethnic minority regions with plenty of oil and how such different choices by states impact ethnic conflict.

Third, our theory and empirical results from case studies (Tang et al. 2017) reinforce the argument that understanding ethnic war requires not only more disaggregated and fine-grained analyses with geographical information but also careful case studies with process-tracing (Sambanis 2004; Ross 2004a; Smith 2014). Even with GIS data, purely quantitative exercises cannot really differentiate genuinely positive cases from false positive cases. For instance, in most quantitative exercises, whether according to oil income or rent at the national level or according to the “Oil Gini”, and even our own oil location, the two Chechnya wars would have been identified as positive cases that suggest a link between oil and ethnic war. Yet, the two Chechnya wars had little to do with oil located within Chechnya. The Chechens rebelled not because of oil but because of their desire to become independent again (Tang et al. 2017).

Acknowledgements This research is supported by a “985 project” 3rd phase bulk grant (2010–2013) and an “Academic Excellence” bulk grant (2014–2016) from Fudan University to Shiping Tang. We thank Konstantin Ash, Michael Ross, Jack Paine, Nils Weidmann, and other members of our “Five-Corners-Field School” for critical comments. Kai Wang and Ke Wu provided excellent research assistance. Special thanks go to Guy Michaels, Massimo Morelli, and Dominic Rohner for sharing their data sets and do files. The usual disclaimer applies. Data and do files for replication are available from CPSR’s online appendix Website.

References

- Asal, Victor, Michael Findley, James A. Piazza, and James Igoe Walsh. 2016. Political exclusion, oil, and ethnic armed conflict. *Journal of Conflict Resolution* 60 (8): 1343–1367.
- Beck, Nathaniel, Jonathan Katz, and Richard Tucker. 1998. Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science* 42 (4): 1260–1288.
- Blattman, Christopher, and Edward Miguel. 2010. Civil war. *Journal of Economic Literature* 48 (1): 3–57.
- Bormann, Nils-Christian. 2011. Codebook for GeoEPR-ETH. http://www.icr.ethz.ch/data/growup/geoepr-eth/geoepr-eth_v.2.0_codebook.pdf. Accessed 11 Nov 2013.
- Brunnschweiler, Christa N., and Ervin H. Bulet. 2009. Natural resources and violent conflict: Resource abundance, dependence, and the onset of civil wars. *Oxford Economic Papers* 81: 651–674.
- Buhaug, Halvard, and Scott Gates. 2002. The geography of civil war. *Journal of Peace Studies* 53 (4): 544–569.

- Carter, David B., and Curtis S. Signorion. 2010. Back to the future: Modeling time dependence in binary data. *Political Analysis* 18: 281–292.
- Cederman, Lars-Erik, Halvard Buhaug, and Jan Ketil Rod. 2009. Ethno-nationalist dyads and civil war: A GIS-based analysis. *Journal of Conflict Resolution* 53 (4): 496–525.
- Cederman, Lars-Erik, Andreas Wimmer, and Brian Min. 2010. Why do ethnic groups rebel? New data and analysis. *World Politics* 62 (1): 87–119.
- Cederman, Lars-Erik, Kristian Skrede Gleditsch, and Halvard Buhaug. 2013. *Inequality, grievance, and civil war*. Cambridge: Cambridge University Press.
- Collier, Paul, and Anke Hoeffler. 1998. On the economic causes of civil war. *Oxford Economic Papers* 50 (4): 563–573.
- Collier, Paul, and Anke Hoeffler. 2004. Greed and grievance in civil war. *Oxford Economic Papers* 56: 563–595.
- Tiro, Di, and Hasan Muhammed. 1984. *The price of freedom: The unfinished diary*. Stockholm: Ministry of Education. ASNLF/Aceh Sumatra National Liberalization Front.
- Fearon, James D., and David D. Laitin. 1996. Explaining interethnic cooperation. *American Political Science Review* 90 (4): 715–735.
- Fearon, James D., and David D. Laitin. 2003. Ethnicity, insurgency, and civil war. *American Political Science Review* 97 (1): 75–90.
- Fearon, James D., and David D. Laitin. 2011. Sons of the soil, migrants, and civil war. *World Development* 39 (2): 199–211.
- Firth, David. 1993. Bias reduction of maximum likelihood estimates. *Biometrika* 80 (1): 27–38.
- Gleditsch, Nils Petter, et al. 2002. Armed conflict 1946–2001: A new dataset. *Journal of Peace Research* 39 (5): 615–637.
- Harbom, Lotta, and Peter Wallensteen. 2009. Armed conflict, 1946–2008. *Journal of Peace Research* 46 (4): 577–587.
- Hegre, Håvard, and Nicholas Sambanis. 2006. Sensitivity analysis of empirical results on civil war onset. *Journal of Conflict Resolution* 50 (4): 508–535.
- Hechter, Michael. 1970. *Internal colonialism: The celtic fringe in British National Development 1536–1966*. Berkeley: University of California Press.
- Horowitz, Donald. 1985. *Ethnic groups in conflict*. Berkeley: University of California Press.
- Humphreys, Macartan. 2005. Natural resources, conflict, and conflict resolution: Uncovering the mechanisms. *Journal of Conflict Resolution* 49: 508–537.
- Hunziker, Philipp, and Lars-Erik Cederman. 2012. No extraction without representation: Petroleum production and ethnonationalist conflict. Prepared for the APSA Annual Meeting, August 30 to September 2, 2012 in New Orleans.
- Kaldor, Mary. 2007. Oil and conflict: The case of Nagorno Karabakh. In *Oil wars*, ed. Mary Kaldor, Terry Lynn Karl and Yahia Said, 157–182. London: Pluto Press.
- King, Gary, and Langche Zeng. 2001. Explaining Rare Events in International Relations. *International Organization* 55(3): 693–715.
- Koubi, Vally, Gabriele Spilker, Tobias Böhmelt, and Thomas Bernauer. 2014. Do natural resources matter for interstate and intrastate armed conflict? *Journal of Peace Research* 51 (2): 227–243.
- Le Billon, Philippe. 2001. The political ecology of war: Natural resources and armed conflicts. *Political Geography* 20: 561–584.
- Li, Guoyu. 2011. *World Atlas of oil and gas basins*. Hoboken: Wiley-Blackwell.
- Lujala, Paivi, Jan Ketil Rod, and Najda Thieme. 2007. Fighting over oil: Introducing a new dataset. *Conflict Management and Peace Science* 24 (1): 239–256.
- Melando, Victor. 2014. Not manna from heaven after all: The endogeneity of oil. Seattle: mimeo, Department of Political Science, University of Washington.
- Mitchell, Sara McLaughlin, and Cameron Thies. 2012. Resource curse in reverse: How civil wars influence natural resource production. *International Interactions* 38 (2): 218–242.
- Morelli, Massimo, and Dominic Rohrer. 2015. Resource concentration and civil wars. *Journal of Development Economics* 117: 32–47.
- Paine, Jack. 2016. *Greedy or aggrieved? A unified theory of oil and separatist civil war*. Rochester: Department of Political Science, University of Rochester.
- Ray, James Lee. 2003. Explaining interstate conflict and war: What should be controlled for? *Conflict Management and Peace Science* 20 (1): 1–29.
- Ross, Michael L. 2004a. How do natural resources influence civil war? Evidence from 13 cases. *International Organization* 58 (1): 35–67.

- Ross, Michael L. 2004b. What do we know about natural resources and civil war?". *Journal of Peace Research* 41 (3): 337–356.
- Ross, Michael L. 2006. A closer look at oil, diamonds, and civil war. *Annual Review of Political Science* 9: 265–300.
- Ross, Michael L. 2012. *The oil curse: How petroleum wealth shapes the development of nations*. Princeton: Princeton University Press.
- Ross, Michael L. 2014. What have we learned about the resource curse? Unpublished manuscript, Department of Political Sciences, UCLA. Accessed 10 July 2014.
- Ross, Michael, and Paasha Mahdavi. 2015. Oil and gas data, 1932–2014. doi:10.7910/DVN/ZTPW0Y. Harvard Dataverse, V2. Accessed 15 Nov 2015.
- Sambanis, Nicholas. 2004. Using case studies to expand economic models of civil war. *Perspectives on Politics* 2 (2): 259–279.
- Smith, Benjamin. 2014. Exploring the resource-civil war nexus. http://www.benjaminbsmith.net/uploads/9/0/0/6/9006393/chapter_13-smith_resources_civil_war.web.pdf. Accessed 20 Jun 2015.
- Sorens, Jason. 2011. Mineral production, territory, and ethnic rebellion: the role of rebel constituencies. *Journal of Peace Research* 48 (5): 571–585.
- Tang, Shiping. 2015. The onset of ethnic war: A general theory. *Sociological Theory* 33 (3): 256–279.
- Tang, Shiping, Yihan Xiong, and Hui Li. 2017. Does oil cause ethnic war? Comparing evidences from process-tracing with quantitative results. *Security Studies*, forthcoming.
- Thies, Cameron G. 2010. Of rulers, rebels and revenue: State capacity, civil war onset and primary commodities. *Journal of Peace Research* 47 (3): 321–332.
- Van der Pleog, Frederick. 2011. Natural resources: Curses or blessing? *Journal of Economic Literature* 49 (2): 366–420.
- Walter, Barbara F. 2001. *Committing to peace: The successful settlement of civil wars*. Princeton: Princeton University Press.
- Weidmann, Nils B., Jan Ketil Rød, and Lars-Erik Cederman. 2010. Representing ethnic groups in space: A new dataset. *Journal of Peace Research* 47(4): 491–499.
- Weiner, Myron. 1978. *Sons of the soil: Migration and ethnic war in India*. Princeton: Princeton University Press.
- Wimmer, Andreas. 2013. *Waves of war*. Cambridge: Cambridge University Press.
- Wimmer, Andreas, Lars-Erik Cederman, and Brian Min. 2009. Ethnic politics and armed conflict: A configurational analysis of a new global dataset. *American Sociological Review* 74: 316–337.
- Wucherpfennig, Julian, et al. 2011. Political relevant ethnic groups across space and time: Introducing the Geo-EPR dataset. *Conflict Management and Peace Science* 28(5): 423–437.

Hui Li is associate professor of at the School of International Relations and Public Affairs (SIRPA), Fudan University, Shanghai, China. His research interests covers corruption, ethnic conflict, and state capacity. His recent research has appeared in *Democratization*, *Social Indicators Research*, and *Policy and Society*. Email: lhui@fudan.edu.cn.

Shiping Tang is Fudan Distinguished Professor and Dr. Seaker Chan Chair Professor at the School of International Relations and Public Affairs (SIRPA), Fudan University, Shanghai, China. His most recent book, *The Social Evolution of International Politics* (Oxford University Press, 2013), receives the “Annual Best Book Award” from the International Studies Association (ISA) for the year of 2015. He is also the author of *A General Theory of Institutional Change* (Routledge, 2011), *A Theory of Security Strategy for Our Time: Defensive Realism* (Palgrave Macmillan, 2010), and many articles in international relations, comparative politics, and philosophy of the social sciences. Email: twukong@yahoo.com; twukong@fudan.edu.cn.